Data-Driven Approaches to Narrative Personalisation through Psychologically Motivated Models

Jussi Waltteri Nybom

A thesis submitted in partial fulfilment of the requirements for the degree of PhD in Creative Computing at the University of the Arts London

February 2024

Contents

Abstract

AI-driven personalisation offers a clear opportunity for creative industries to engage audiences more effectively. This project seeks to understand how such personalisation can be effectively and ethically exploited in story experiences to generate greater audience engagement. More specifically, this thesis addresses personalisation of narratives to accommodate the user's preferences, with an aim to understand and accommodate them better. For this, three studies are conducted.

In the first study, an interactive narrative is created with the purpose of incorporating the user's choices to create a user profile featuring the Five-Factor Model (FFM) and the Need for Affect (NFA), with the questions designed to understand the user's preferences within the narrative and therefore possibly indicating their personality in general. Next, a narrative is personalised to fit with the user's estimated personality. It is hypothesised that the choices the user makes within the narrative would have at least some correlation with the choices they would make in real life, meaning the narrative could be used as a personality test. Even if little or no correlation with real-life preferences can be found, the choices could indicate preferences within fictional narratives, such as how complex, imaginative or painful they prefer narratives to be, and what the protagonist should be like. Nevertheless, it did turn out the interactive narrative could be used to measure at least Extraversion and Emotional Stability. The effectiveness of personalising the story is then tested, seeing how effective it is to change the style of language according to Extraversion levels, the protagonist personality according to other FFM factors, and the ending according to the NFA. The results were strikingly strong, with personalisation appearing to improve the experience across all traits and with both personality test and interactive narrative results.

In the second study, we attempt to use Natural Language Processing (NLP) for modifying the language, so that the personalisation could be done automatically and not by hand. For this, a number of language models were trained and used to create different version of a short story. Different versions of the ending were also created. The results were then tested on participants, and their opinions on the story and its language were compared with their FFM personality scores, their reading skills and their age and gender. The results presented a complicated picture featuring some surprises.

In the third study, the Myers–Briggs Type Indicator (MBTI) was used instead, with several machine learning algorithms tried for classifying users by their MBTI type based on text they

have written on social media. It was suggested this approach could be used for an MBTI-based recommender system that identifies novels with authors, characters or narrators similar to the reader. The results suggested that the approach could indeed work, with results in the extraversion dimension particularly promising, but better data would be needed to gain strong enough results for a good recommender system.

There are several potential impacts relating to this work. It could lead to the creation of new measures for testing people’s preferences in art. User profiles using the measures could then be used to personalise narratives to fit with the user’s preferences in various ways. By integrating personality frameworks, recommender algorithms can suggest novels, films and other works that not only align with users' personality traits but also cater to their broader preferences, offering a more tailored and enriching experience, with little usage data needed. Similar approaches with NLP can also be used to alter pre-existing works.

An extensive literature review is also conducted, giving a wide introduction to the topic and related fields. This is then used as a background for suggesting more possible future pathways for personalising narratives.

Chapter I: Introduction

In today's rapidly evolving technological landscape, artificial intelligence has ignited a transformative wave across numerous industries, with particular resonance in the creative realm. This transformative wave is not merely a trend but a paradigm shift that is reshaping the way we approach creativity, innovation, and problem-solving.

Artificial intelligence encompasses a spectrum of cutting-edge technologies, including machine learning, natural language processing, computer vision, and deep learning, among others. These technologies empower AI systems to analyse vast datasets, detect intricate patterns, and make predictions or recommendations. Consequently, this revolution has far-reaching implications for the creative domain, transcending traditional boundaries and driving innovation in previously unimaginable ways.

One of the most compelling aspects of AI's integration into creativity is its potential to augment human capabilities. For instance, AI-powered tools have demonstrated the ability to generate creative content, such as art, music, and literature. Artists, musicians, and writers can collaborate with AI to explore new horizons and push the boundaries of their creative expressions. This synergistic partnership between humans and machines is fostering an era of enhanced creativity, opening doors to uncharted territories of artistic exploration.

Moreover, AI-driven content creation is not limited to emulation but extends to personalisation, automatically tailoring content to individual preferences. AI algorithms can craft tailored content, taking into account individual preferences, such as personalised book recommendations based on reading habits or music playlists that resonate with the listener's emotional state. This personalisation creates immersive and engaging experiences for users, transforming the way they interact with creative works. The convergence of AI-driven personalisation and the creative arts offers a profound opportunity to engage audiences in novel and compelling ways. This synergy not only enhances the consumption of creative content but also blurs the boundaries between creator and consumer, ushering in an era of immersive, tailored experiences.

Imagine a world where a novel would be different for every reader. Perhaps people who prefer sadder stories would be given sadder endings; people preferring easier reading would be given more simple language; people preferring poetic language would be given just that, and so on. The concept of tailoring novels to suit the preferences of individual readers is a captivating vision for the future of literature. In this world, each reader gets a literary journey that is uniquely their own, an experience that resonates deeply with their tastes, emotions, and cognitive preferences. How could this be done?

Personalising narratives can happen as changing the storyline, or the tone, or the presentation – would the user prefer brighter colours, or darker humour, or more ambiguous endings? This could be figured out in various ways, for example, by using common methods from recommender systems, collaborative filtering can discover similar users and thus predict what the user would be likely to like; content-based filtering could offer more of the things the user has liked; and knowledge-based filtering could just ask what the user likes. Other possible methods could also be thought of, and of course hybrid methods are a strong option, as well.

However, recommendations have not been typically used with narratives, which are an environment of their own. Recommendations are often offered in isolation, just offering the best possible individual items, often much of the same. There has not been much work on recommending sequences or bundles; it is sequences that make a narrative, and much more than just the plot is involved: there are many factors in the presentation.

Moreover, recommendations can often just keep repeating themselves and offering more of the same, predictable predictions, but narratives should often be about surprises. Different circumstances and different individuals need different emphases on whether familiarity or novelty is better, and surprises can be either good or bad, or even both at the same time. Replayability can also be important, especially with games, and the experience being different each time could also encourage repeatedly watching a film, for example. Perhaps the best recommendations can do is to find a good way of surprising people, to create varied, mixed experiences, rather than to offer predictable patterns. It can be expected that when people engage with an interactive narrative, they are not looking for something conventional, something they are used to, something that will help them fall asleep, but rather something new, exciting, surprising, even strange.

It is easy to imagine someone describing to their friend this fancy new film where the AI created the weirdest story for them, and being all excited about it, perhaps giving it another go shortly afterwards. On the other hand, a conventional narrative would be described as no different from anything else, and therefore not really worth the bother. Nevertheless, such novelty value could get AI-generated narratives started with popularity, but it might not last without deeper quality. A good system will be able to keep surprising again and again, but also understand the value of familiarity and all other aspects of conventional storytelling.

The AI must understand story arcs: a story must have a beginning, a middle, and an end, and they should have varying tones. A story that is equally happy or miserable from the first page to the last is unlikely to hold much appeal; there must be moves from happiness to misery or vice versa, and perhaps a move back. Researching the element of surprise can also help develop other recommendations: filter bubbles and repetition are a notorious feature of recommender systems. Offering something new should open up new worlds to users in large services such as YouTube, which includes all manner of videos in the world, but typically displays just a narrow corner of this to each individual user, with little option to browse everything it holds. Giving the user more power over the recommendations should be a good option, giving more room for customisation rather than automatic personalisation.

Giving the user more options could also help create a more representative representation of the user, to reduce the distance between the user and how they are represented through data, understanding more of their preferences to cater more of their needs. How can data understand the user better? One way of doing this would be to work on the personalisation methods, perhaps innovating ways of combining different methods.

Improving computers' understanding of users and bridging the distance between human experiences and their data representations pose critical challenges. Addressing these challenges is essential for enhancing personalisation, particularly in narratives. Virtual environments offer unique opportunities to assess personality, surpassing the limitations of traditional questionnaires, and warrant the development of innovative approaches in this realm. Efforts should focus on devising methods to collect more user data and exploring diverse options for adaptation to user preferences. The adaptation of systems to user preferences can be facilitated through a range of alternative options, providing valuable guidance for automated personalisation.

To measure and bridge these gaps effectively, it is imperative to develop ways of collecting user data and study methodologies for achieving this. Systems should learn to dynamically adjust to users' preferences, employing a variety of alternative options. Enhancing user customisation, which can serve as a source of guidance for automated personalisation, would be helpful. User experience testing, involving interactive machine learning and psychophysical measurements, can be employed to compare users' descriptions of their experiences with computational interpretations. Generalising these experiences to users with similar characteristics is vital, emphasising the importance of understanding personality and emotion in personalisation. The synergy of artificial intelligence and human feedback should be a central focus, combining computational insights with users' subjective assessments. Lastly, the convergence of personalisation and narratives should be explored within the context of understanding narrative structures and complexities, presenting a novel avenue for the application of personalisation methodologies.

The foundation of personalised novels lies in understanding the reader. Advanced AI algorithms could profile readers by analysing their reading history, genres they favour, emotional responses to various narrative elements, and even their cognitive style. Personality frameworks such as the Five-Factor Model (FFM) or the Myers-Briggs Type Indicator (MBTI) could serve as valuable tools for understanding readers at a deeper level, being able to predict much of individual preferences in art and entertainment.

The plot, a central pillar of any novel, could be crafted in real-time based on the reader's preferences. While some elements of the story may remain constant, pivotal moments, character arcs, and even the story's outcome could be fluid. Readers favouring suspense might find themselves in a thriller-like climax, while those who prefer heartwarming tales could experience a heartening conclusion. Authors or AI systems could craft alternate story branches that seamlessly integrate with the central narrative, with the reader's preferences, choices or reactions influencing the plot's direction.

Language is a powerful tool in storytelling, and it could be tailored to align with readers' linguistic preferences. For those who prefer simpler language, the prose could be more straightforward and concise. Readers who appreciate poetic and eloquent language might be immersed in beautifully crafted descriptions and metaphors. With the assistance of natural language processing (NLP), natural language generation (NLG) and generative AI, authors could customise the linguistic style of the novel to cater to individual tastes. There could be a few alternative paths that they author could write for the story, and then the most suitable one could get picked based on the user profile. A blend of manual and automated approaches may be used in creating personalised novels. While AI systems can handle certain aspects like plot adaptation and linguistic style, authors or literary experts could play an active role in shaping key elements of the story. The creative input of humans would ensure that the novels maintain the depth, artistic integrity, and thematic coherence required for a compelling narrative.

Potentially, hyper-personalised novels could even incorporate real-time feedback mechanisms. Readers might have the option to provide feedback on their emotional responses and reading experiences. This feedback loop could allow the novel to adapt even further, fine-tuning the narrative, language, and emotional tone as the story unfolds. This could even be done through physiological measurements. This, however, raises the question of how the collection and use of personal data to create hyper-personalised novels would need to be conducted with strict adherence to privacy and ethical standards. Users should have control over their data and be informed about how it will be utilised. Additionally, safeguards should be in place to prevent biased or discriminatory personalisation.

These are some examples of research avenues and possibilities that this thesis has sought to explore. The aim is to explore the intricacies of personalising narratives to accommodate individual preferences, thereby creating a deeper connection between audiences and the stories they consume. Through a multi-faceted examination of interactive narratives, Natural Language Processing (NLP) techniques, and the use of personality frameworks like the Five-Factor Model (FFM) and the Myers–Briggs Type Indicator (MBTI), we explore the potential for personalisation to redefine the way we create, consume, and interact with creative content across a spectrum of mediums, including literature, film, and music, but with specific focus on narrative fiction.

What this thesis presents is a combination of approaches for personalising story experiences that can be used together or separately. A lot of the focus in the studies is on short stories, and the discussion will often speak of novels, but the approach is just as relevant to other story experiences such as interactive narratives, films and games.

The focus is on using approaches from personality psychology to adapt stories to suit people with different personalities, and finding ways to identify their personalities in ways that do not require sitting through traditional personality tests. For this, the literary preferences associated with different personalities are also explored, and different personalisation approaches and writing styles are tried out in user studies.

What is envisioned is a personalisation approach or system that can adapt just about every aspect of a story experience to suit a user's personality, be that the personality of the protagonist or the narrator, the style of language, the direction of the plot, or even its themes. If the user is expected to like the story to be happy, it can be taken into that direction; if they would likely prefer less formal language, that's what they'll get. This could mean using AI to edit the stories, or the author preparing multiple variations. The approach could also be used in a recommender system to discover works that are already well suited for the reader in the first place.

The literature review (Chapter II) explores personalisation in games, interactive storytelling, and recommender systems. It provides an introduction to the concept of personalisation and its historical usage in gaming. Significant focus is given for the potential of recommender systems and their role in tailoring content to user preferences. The discussion extends to player modelling and profiling, which also find relevance in narrative experiences. While player typology models are limited in scope and applicability, the Five-Factor Model of personality (FFM) emerges as a promising method for understanding user preferences, particularly when data is scarce. The Need for Affect (NFA) is considered as a viable measure for assessing emotional and thematic preferences in media. Alternatively, given the MBTI's wide popularity in the general population, using the MBTI instead of FFM could broaden the applicability and acceptance of the recommender system among a wider audience, and help with finding data that can be used research and practical purposes. The chapter also explores the automatic generation of stories, both interactive and non-interactive, with and without player modelling. These generated narratives can serve as a foundation for recommender systems, enhancing user engagement and diversifying recommendations. Text-based personality recognition is noted as a potential way to create a user profile, and different factors and aspects of reading preferences are discussed. The discussion also covers general factors in reading preferences: what makes a narrative interesting or appealing, and why people care about fictional characters and may seem to enjoy painful narratives. Finally, the potential of emotion detection is noted.

Method is discussed in Chapter III, briefly describing the mixed methods approach used, noting the use of user studies, interactive and non-interactive narrative creation, personality frameworks, natural language processing, machine learning, data collection and analysis, and discussing ethical considerations.

In the first study (Chapter IV), a user study is presented, for which an interactive narrative was created for the purpose of creating a user profile based on the Five-Factor Model (FFM) and the Need for Affect (NFA). The questions were designed to reveal user preferences within the narrative and potentially offer insights into their general personality traits. The hypothesis was that the user's choices within the narrative could exhibit correlations with their real-life preferences, potentially serving as a form of personality assessment. Even if such correlations were minimal or absent, these choices could still indicate preferences within fictional narratives, such as language style, complexity, emotional intensity, and protagonist characteristics. Nevertheless, the interactive narrative turned out to be quite good at capturing some personality traits. Subsequently, a short story was personalised to align with the user's estimated personality, adapting language according to Extraversion, protagonist personality with other FFM traits, and ending according to the NFA. This turned out to be a striking success. Both the interactive narrative and the personality test worked well for the personalisation, indicating it is possible the interactive narrative could have had its advantages in capturing personality over the traditional personality test. The study also uncovers other intriguing findings, such as the language preferences of Extraverted and Introverted individuals. Overall, the findings represent multiple ways to personalise narratives according to personality, along with identifying personality in an entertaining manner.

The second study (Chapter V) focuses on utilising psychological models in conjunction with text style transfer. It aims to leverage Natural Language Processing (NLP) for automated personalisation, reducing the need for manual intervention. Various language models were trained and tested on participants, with their preferences compared to their FFM personality scores. The aim was to increase depth to the previous findings about different personalities and their preferences for more or less formal or creative use of language, and to see how well NLP could adapt language style. The previous study highlighted the effectiveness of the Five-Factor Model (FFM) in assessing literary preferences and character adaptation. However, this study focuses on language style adaptation, yielding less definitive outcomes. The previously identified link between Extraversion and language formality seemed to hold but requires further investigation, considering variations in language versions. Personalising narratives through NLP remains an underexplored domain with immense potential, particularly seeing how suddenly LLMs have improved after the study was conducted.

The third study (Chapter VI) explores the Myers–Briggs Type Indicator (MBTI) as a means of identifying personality traits from text data, namely social media posts the user has written. The study draws from the Personality Café dataset by Keh and Cheng (2019) and the MBTI9K dataset by Gjurković and Šnajder (2018), both including social media posts and the MBTI type of the users. It employs various machine learning algorithms to classify users based on their MBTI type using their textual data. The ultimate goal is to create a recommender system that incorporates MBTI-based recommendations, using multiple methods based on the reader's MBTI. As the need for balanced predictions for different personality types is a key consideration for an effective recommender system, the efforts focused on minority classes in an imbalanced dataset. The investigation into the MBTI as a tool for recognising personality from text data reveals its potential applicability, especially considering its widespread popularity, as data are more readily available than with FFM, and many interested in the MBTI could also have interest and trust in a recommender based on it. However, it is noted that the MBTI, unlike the FFM, is not based in linguistic enquiry, and is generally less empirically based. Furthermore, many previous studies on textual MBTI recognition turn out to be excessively optimistic, often reporting unrealistic results, or reporting the results without necessary details. It was found that expectations should be curtailed, as recognising MBTI traits from text is indeed a realistic task, but not precise to a very high degree, and the data might not always be sufficient or ideal. While the consistency of personality judgments from novels was not found quite sufficient for an effective recommender system, the study highlights the potential of such a system with the need for more comprehensive data.

The outcomes of this thesis could have several significant impacts. It may lead to the development of novel measures for assessing people's artistic preferences. User profiles based on these measures could then be applied to personalise narratives in various ways, including plot direction, language usage, and character/narrator personalities. Additionally, recommendations could be made for works that align with these preferences, addressing the common "cold start" challenge faced by recommender systems with new users, when there is no data on which to base recommendations.

The work explores how we understand and harness personalisation in creative content. The potential impact extends across multiple domains. By tailoring narratives, content, and recommendations to individual preferences, creators and content providers can enhance user engagement. Users are more likely to connect with content that resonates with their personality traits and emotional preferences.

The concept of interactive narratives that adapt to user preferences opens up the possibility of entirely new narrative structures. Creators can experiment with storytelling formats that empower users to co-author their experiences, blurring the lines between creator and consumer.

Recommender systems, armed with user profiles based on personality can overcome the common challenge of recommending more of the same. Instead, they can diversify recommendations, exposing users to a wider range of creative content that aligns with their unique tastes. The consideration of the Myers–Briggs Type Indicator (MBTI) as a personality framework for personalisation can potentially broaden the acceptance and usability of recommender systems. The MBTI's popularity and accessibility may make it easier to collect user data and encourage users to engage with personalised recommendations.

Additionally, insights gained from personalisation in narrative fiction can extend beyond literature and storytelling. The methodologies explored here can be applied in various domains, including music, film, education, and healthcare, offering tailored experiences to users. Finally, as AI-driven personalisation becomes more prevalent, ethical considerations become paramount. This work underscores the importance of responsible data handling and transparency in personalisation efforts.

In terms of future directions, there is a vast landscape to explore. Ongoing research can refine personalisation models to capture even subtler nuances of user preferences, further enhancing the quality of recommendations and narrative personalisation. Newer language models are vastly better than the old ones, and significant improvements could be seen in text style transfer. More studies on interactive narratives working as personality tests would be helpful. Different personality models could be explored and developed, perhaps specifically for these purposes, but perhaps even having wider applicability. There is a prospect of adapting the methodologies explored here to different domains, such as education and healthcare, which can open up new frontiers for personalised experiences in these areas. Moreover, allowing the users more freedom to adapt and customise their own user models could also be an interesting prospect, helping with user-centric design, prioritising user feedback; the use of user studies and using them for iterative design processes should likely become more common to ensure that personalisation efforts align with users' evolving preferences and expectations. Exploring ways to involve users actively in shaping their personalised experiences can empower them and lead to more satisfying interactions. Multi-modal integration could also be an intriguing approach, combining, for example, user behaviour, sentiment analysis, and biometric data with personality frameworks can create a holistic understanding of users, enabling more precise personalisation.

Contributions

The research yielded several noteworthy discoveries. Firstly, it became evident that interactive narratives possess a remarkable ability to capture certain aspects of personality traits. Specifically, the study underscored the effectiveness of interactive narratives in discerning traits such as Extraversion and Emotional Stability. Personalising the protagonist in the short story based on users' Five-Factor Model (FFM) personality scores yielded highly promising results, regardless of whether these scores were derived from interactive narrative interactions or traditional personality assessments. This suggests that interactive narratives may offer even a more nuanced and insightful portrayal of users' personalities compared to conventional testing methods. These findings underscore the potential of narrative-driven approaches for personality assessment, prompting reflection on the adaptability of traditional tests in capturing subtle preferences within fictional contexts.

Conventional personality assessments typically rely on abstract and sometimes ambiguous questions that prompt individuals to reflect on their behaviours, preferences, and tendencies. However, this approach has inherent limitations, as respondents may interpret questions differently, leading to varied responses. Moreover, the abstract nature of these inquiries can obscure the translation of answers into personality trait assessments. Conversely, interactive narratives plunge users into distinct, palpable scenarios, directing them through situations designed to elicit genuine reactions, behaviours, and choices. As users traverse these interactive encounters, their responses are inclined to reflect their authentic inclinations and tendencies more accurately. By furnishing a setting that mirrors real-life decision-making, interactive narratives present the prospect of offering a more genuine insight into an individual's personality.

The collective evidence strongly supports the proposition that interactive narratives represent a highly promising avenue for assessing personality traits. This method harbours substantial promise across diverse domains, especially in personalising narratives to suit individual preferences. The interactivity inherent in these narratives renders them notably captivating, providing a more enthralling substitute to conventional personality evaluations. This holds particular relevance for individuals with a penchant for narrative experiences. Furthermore, the incorporation of gamification elements amplifies engagement, rendering the personality assessment process not only enjoyable but also in harmony with the inclinations of those drawn to interactive and game-like content.

Independently from the interactive narrative, the personalised short story emerged as a distinctive and immersive feature. Tailored to individual user profiles, the short story proved remarkably adept at reflecting users' personalities and emotional tendencies, reflected in higher reader satisfaction than in the control group. The personalised short story offered a bespoke literary experience tailored to each user's unique personality traits as delineated by the FFM and the NFA. This approach not only heightened user immersion within the narrative but also yielded a wealth of data for comprehensive analysis.

The personalised narrative also reveals additional intriguing discoveries, including the language inclinations observed between Extraverted and Introverted individuals. While previous research has indicated that Extraverts typically employ a less formal writing style compared to Introverts, their reading preferences in this regard have not been thoroughly examined. Interestingly, it was discovered that these preferences seem to extend to the reading habits of both groups when it comes to short stories.

The subsequent inquiry into AI-facilitated text style transformation and reader predilections, contingent upon personality traits and additional variables, broadened comprehension of reading inclinations and unveiled fresh avenues for investigation. Noteworthy among the findings was the identification of associations between particular personality characteristics and predilections towards distinct writing formats. Additionally, the age of participants emerged as a significant factor, with older readers exhibiting diminished contentment concerning language coherence and overall narrative enjoyment.

The final study noted that numerous preceding studies on textual MBTI identification exhibited an overly sanguine outlook, frequently presenting outcomes that were either unrealistic or lacking essential particulars. Through a comparative examination of machine learning models, insights into their performance were gleaned, with models specifically tailored to address imbalanced data, notably the Easy Ensemble approach, showcasing effectiveness in managing minority classes. However, it became evident that expectations needed tempering, as discerning the MBTI attributes from text proved feasible but not to an exceedingly precise extent, with data adequacy and quality occasionally posing challenges. Although the reliability of personality assessments derived from novels was deemed insufficient for facilitating a robust recommender system, the study underscored the system's potential efficacy, contingent upon the availability of more extensive and comprehensive datasets. The results were overall decent, with particularly good results for the Extraversion dimension, which could end up being useful even on its own.

This research has made significant strides in elucidating how personalisation techniques rooted in personality traits can augment user experiences within interactive narratives. However, it also underscores the vast potential for continued exploration and innovation within this domain. The efficacy of certain personalisation methods and the inherent flexibility of interactive narratives herald exciting prospects for further research and development in personalisation techniques. One promising avenue for future investigation lies in exploring personalisation based on alternative personality frameworks or traits. While this study primarily focused on techniques derived from established models like the Five-Factor Model or the MBTI, there exists a plethora of other personality dimensions that could enrich user engagement and immersion.

Moreover, personalisation could extend beyond conventional personality models to encompass other individual differences that influence reading experiences. One proposed approach to devise a literature-specific framework involves compiling a comprehensive list of emotions readers seek from literature and creating reader profiles based on the importance of each emotion to them. Exploring alternative frameworks offers avenues for a more nuanced comprehension of reader preferences. Researchers can explore the complex dynamics of individual interactions with fictional content by incorporating various psychological models, nurturing reader communities organised by personality types, and adjusting player type models for literature. Continued research and experimentation are essential to refine these frameworks and tailor them to the distinct landscape of literature. The integration of emotional frameworks represents a particularly promising direction, enabling a personalised approach that transcends traditional personality classifications.

Chapter II: Literature Review

1. Introduction

This chapter explores the intersections of personalisation, gaming, interactive storytelling, and recommender systems, examining their significant influence on creative and digital realms. The ultimate aim is to lay the groundwork for the approach developed in later chapters and to explain the concepts discussed there, considering work done previously and how they can be built upon.

The chapter starts with an introduction to personalisation, exploring both its historical roots and its contemporary manifestations. Personalisation, in the context of digital entertainment, is the art of tailoring content to individual preferences, and it has taken centre stage in the realms of gaming and interactive storytelling. These technologies have harnessed the power of personalisation to immerse users in experiences uniquely tailored to their tastes and desires. A general introduction to the topic is made, considering both its benefits and problems, noting that while it can improve the enjoyability and personal relevance of content, it can also be executed badly, or be problematic regarding privacy or create experiences that are difficult to review when they are different to different people. These are important considerations that could potentially determine whether the field is worth the risks involved in the first place.

The concept of player modelling and profiling, fundamental to personalisation in gaming experiences is discussed in the following section. These techniques, traditionally applied in the gaming domain, are found to be powerful tools for enhancing narrative engagement as well. Profiling approaches are noted to be able to recognise playstyles during gameplay.

The fourth section explores psychological approaches to personalisation and player profiling. The Five-Factor Model of personality (FFM) emerges as a robust method for uncovering the core preferences of users, especially when faced with scant data. Moreover, the chapter shines a spotlight on the potential of the concept of the Need for Affect (NFA), a measure of emotional and thematic preferences in media. The NFA emerges as a promising option for understanding how users engage with narratives, particularly in terms of emotional depth and intensity. The potential of the widely popular but less empirical Myers–Briggs Type Indicator (MBTI) is also discussed.

The fifth section progresses to player type models, investigating their applications in assessing player distributions and personalisation. Bartle's model categorises players into Killers, Socialisers, Achievers, and Explorers, providing a framework for understanding player preferences. Demographic Game Design models, Brainhex, and critiques of player typologies further contribute to the discourse, underlining the need for a broader trait theory of playing preferences.

The sixth section discusses Procedural Content Generation (PCG), examining its algorithmic approach to creating game content efficiently. PCG is presented as a cost-effective solution for generating content beyond a designer's imagination. The discussion includes Experience-Driven Procedural Content Generation (EDPCG), viewing content as a crucial component of player experience and suggesting adjustments to optimise playing experiences.

This then leads to the topic of interactive storytelling, especially in relation to Natural Language Processing (NLP), and how NLP can be used to edit and personalise narratives, but also to how interactive narratives and gamification can be useful in creating personality profiles. The section explores challenges related to supporting meaningful user choices while maintaining plot coherence. Approaches such as plot-based manipulation, character-based interaction, and techniques like branching story graphs and Drama Managers are detailed. Player modelling in narrative design is acknowledged as an underexplored area, with references to specific applications like PaSSAGE, which adapts stories based on a player's style of play.

Narrative generation is discussed in the context of semantic representation and Natural Language Generation (NLG). Text style transfer, involving supervised and unsupervised methods, disentanglement approaches, and paraphrase generation, is discussed. Evaluation metrics, including accuracy, fluency, and content preservation, are acknowledged as facing challenges, with a recognition that current metrics might not fully capture the essence of style transfer. An additional dimension is introduced in the form of gamification and personality testing. Gamification, incorporating game elements into tasks, is discussed for its potential to enhance engagement and response in personality assessment. The application of gamified survey-type tasks in various fields, such as marketing and education, is acknowledged. Interactive narratives are posited as potential personality tests, referencing specific studies where participants' narrative choices correlated with certain traits.

The subsequent section focuses on Recommender Systems (RS) and their pivotal role in predicting user preferences and offering personalised suggestions. Typically reliant on data-rich user profiles, recommender systems can evolve to incorporate advanced techniques like collaborative filtering, opening new horizons in personalised content recommendations. The chapter underscores the importance of recommender systems, not only as facilitators of content discovery but as potent instruments in shaping our digital experiences. These systems, discussed in detail, are positioned as the bridge between users and a vast array of creative content, from novels to films to music. The chapter's exploration illuminates the role of personality in overcoming the perennial "cold start" problem faced by new users in recommender systems, ensuring that recommendations are not mere repetitions of the familiar but diverse and enriching.

Emotions are recognised as significant in recommender systems, with some systems categorising emotions for personalised recommendations. Personality-aware recommender systems, often utilising the Five-Factor Model or Myers-Briggs Type Indicator, show correlations between personality traits and user preferences. Trust is noted as a critical aspect, impacting user preference for machine or human recommendations. Social networks offer avenues for building trust, with community-based recommenders incorporating friends' preferences. The potential downside includes the creation of filter bubbles, limiting exposure to diverse information.

The exploration of recommender systems concludes with a reflection on the balance between personalisation and diversity. Balancing personalisation with diversity is deemed crucial to avoid filter bubbles and provide novel recommendations. Diversification strategies, such as injecting randomness, utilising genetic algorithms, and periodically switching algorithms, are discussed, with evaluation metrics aiding in assessing trade-offs associated with introducing novelty.

The chapter also introduces the concept of personality detection through writing style, an alternative avenue for understanding users' preferences and personalities. With this multifaceted exploration of personalisation, gaming, interactive storytelling, and recommender systems, the chapter paves the way for a deeper understanding of how technology can be harnessed to create richer, more engaging digital experiences, forever blurring the lines between creator and consumer.

The subsequent sections study text-based personality recognition, reading preferences, and emotion detection. Text-based personality recognition is acknowledged as a challenging task in computational stylometry, involving predicting personality traits from writing style. The Five-Factor Model and Myers–Briggs Type Indicator serve as commonly used taxonomies in this context. Despite challenges related to limited labelled data and disparities between the FFM and the MBTI models, text-based personality recognition is considered promising for personalisation.

To be able to personalise narratives, it is also important to understand the appeal of narratives in the first place, and so concepts such as interest, empathy and pain in the realm of narratives are discussed and introduced, and individual differences in them is considered. Reading preferences are explored in three dimensions: interest, empathy for characters, and painful responses. Factors influencing textual interest, including coherence, ease of comprehension, and themes, are analysed. The reader's ability to empathise with characters is discussed in terms of perceived similarity, optimal distance, and transparency. Emotional distress in readers when characters undergo suffering is acknowledged, influenced by factors such as relationship, perceived similarity, and likability.

The final section focuses on emotion detection. Affective computing, recognising, interpreting, and simulating human emotions using various physiological and behavioural signals, is discussed. Despite challenges related to invasiveness and impracticality, using commonly available devices for emotion detection is seen as a promising avenue for enhancing personalisation.

In conclusion, this literature review consolidates diverse perspectives on personalisation in digital experiences. From its historical roots to contemporary applications in gaming, interactive storytelling, and recommender systems, each section contributes to a nuanced understanding of personalisation's multifaceted dimensions. Interactive narratives are positioned as tools for generating personalised content through creating a deeper understanding the user through their choices, as well as being very much capable of being personalised themselves. Personality-based recommender systems are underscored for their role in addressing the persistent "cold start" problem. It is noted psychological frameworks should be able to create a better overall profile of the user than player typologies, which are more suited for competitive environments.

Approaches for generating and adapting narrative content are discussed, with the aim of doing this based on the user’s preferences. The other side of the relationship between writing and personality and preferences is also noted, and the potential of text-based personality recognition is considered as one way of laying the foundation of personality-based personalisation. And of course, in seeking to create appealing content, we must consider what makes interesting and appealing text in the first place, and especially in the context of personal differences. All this leads us closer to sketching an approach for personality-based personalisation, to be explored in subsequent chapters.

2. Uses and Problems of Personalisation

While the terms customisation and personalisation are at times used interchangeably, personalisation typically denotes a system-driven adaptation, while customisation is driven by user preferences (Sundar & Marathe, 2010). In personalisation, a model of each individual user is created automatically, and the relevant experience is tailored to them accordingly. Customisation, on the other hand, is done by the user, not the computer. A system may enable users to make changes to the experience to meet their preferences.

Personalisation often features recommender systems, information filtering systems for predicting how a user would rate an item, and providing suggestions related to these interests, based on information about the user and their interests that is gathered either explicitly through ratings the user has given, or implicitly by interpreting their actions. There are two main types of recommender systems: ones based on collaborative filtering (CF), where the basic approach is to recommend items that similar users have liked; and content-based filtering (CB), which recommends items similar to the ones the user has liked. There are also other types of systems that can be considered recommender systems, such as knowledge-based systems.

The benefit of personalisation is an improved user experience which requires little or no effort from the user. However, it leaves the user no opportunity to improve on the system's guesses. Some users might also find the experience unnerving or creepy, particularly if information they would rather keep private has been used, meaning personalisation can also be an ethically loaded issue, which must be kept in mind.

The reliance of personalisation on extensive user data introduces a delicate trade-off for the user: a balance between the convenience derived from tailored experiences and the preservation of privacy. The more a system personalises content based on user data, the greater the intrusion into the user's privacy. On the other hand, customisation, though potentially requiring more active involvement from the user, offers a heightened sense of autonomy. Users engaging in customisation are actively involved in deciding what elements they want to tailor, fostering a personalised experience while maintaining a semblance of control over the information shared.

Interactive machine learning is an advanced approach to customisation. It can offer users without machine learning or programming expertise an effective way to customise systems. When systems make mistakes, users can often fix them providing corrective demonstrations to the system rather than changing the machine learning algorithm or program code. For instance, a user can provide additional examples of an action that was misclassified by a trained model, along with the correct label for this action, and reasonably hope that the next model trained on the augmented training set will improve its performance on that type of action (Bernardo et al., 2017).

Despite the potential inconvenience associated with customisation, its intrinsic value lies in the empowerment it offers to users. The ability to dictate one's preferences and curate the experience can contribute significantly to the overall enjoyment and satisfaction of the user. This empowerment extends to privacy concerns with recommender systems. Sundar and Marathe (2010) suggest that customisation can override privacy concerns, as users willingly participate in the tailoring process, feeling more in control of their shared information.

As a strategic approach, customising privacy settings at the outset emerges as a proactive measure. Zhang and Sundar (2019: 87) argue that users who take the initiative to customise their privacy settings are likely to receive more positively embraced recommendations. This proactive engagement not only aligns the recommendations more closely with user preferences but also signals to recommender systems the user's willingness to share specific information for a more tailored experience. In essence, the act of customising privacy settings becomes a pivotal step in shaping the dynamics of the user-system interaction, striking a balance between personalised experiences and privacy considerations.

Personalisation has become increasingly prevalent within the gaming landscape, eliciting strong expectations from players (Zad, Angelides & Agius, 2012). This trend not only meets the anticipated desires of gamers but also serves as a catalyst for heightened player loyalty and enjoyment (Teng, 2010; Turkay & Adinolf, 2010). In a study appearing to demonstrate the power of personalisation in games, Fischer, Kastenmuller and Greitemeyer (2010) conducted a seminal study that delved into the impact of personalised in-game characters. The findings revealed a notable increase in arousal and self-activation, coupled with a decrease in aggression, when compared to default game characters. This trend was corroborated in a subsequent study by Hollingdale and Greitemeyer (2013). The infusion of personalisation into gaming experiences not only enhances player engagement but also shapes emotional responses and behavioural tendencies. Adaptive AI should be well placed to improve enjoyability of games; games that are less predictable have been found more enjoyable. Snowdon and Oikonomou (2011) studied games using item randomisation to make the game a slightly different every time you played, finding that not knowing what would happen next improved both replayability and initial enjoyment of the game.

Moreover, marketing studies have consistently demonstrated that consumers tend to associate brands with specific personality types, expressing a preference for brands embodying traits aligning with their preferences (e.g. Fennis & Pruyn, 2007). This consumer inclination provides companies with a compelling reason to tailor the personality of their dialogue systems in accordance with the characteristics deemed appealing to their target market (Mairesse & Walker, 2010: 3). In the gaming industry, this alignment between brand personality and player preferences underscores the potential for personalised dialogue systems to enhance player engagement and immersion. By integrating personalisation strategies that resonate with the desired brand image, gaming companies can establish a more profound and lasting connection with their player base.

In the realm of educational, persuasive, and other serious games, the efficacy of personalisation has been firmly established (e.g. Peirce & Wade, 2010). A wealth of studies attests to the transformative impact of personalisation on persuasive systems, particularly in fostering the desired change in behaviour (Orji, 2014; Kaptein et al., 2012; Dijkstra, 2014; Busch et al., 2016). The nuanced nature of human motivation is evident in research indicating that a persuasive approach effective for one group may yield contrasting outcomes for another (Kaptein et al., 2012; Orji et al., 2013). This underscores the imperative of tailoring persuasive strategies to align with the diverse motivations and preferences of distinct user groups.

Beyond traditional gaming contexts, ambient games, designed for sensory engagement and relaxation, distinctly benefit from the analysis of player behaviour to dynamically adapt the game context (Schouten et al., 2011). The soothing nature of ambient games relies on a careful calibration of game elements in response to the player's behavioural cues, providing an immersive and calming experience. This application of personalisation strategies extends beyond conventional gaming paradigms, demonstrating the versatility and effectiveness of tailored experiences in diverse gaming genres.

A notable challenge associated with personalisation is its inherent tendency to yield unique experiences for each user, rendering direct performance comparisons virtually impossible. This individualised nature of personalisation introduces complexities in evaluating and testing adaptive products, especially within the context of gaming. The mechanisms employed in games utilising online learning are marked by unpredictability, presenting a formidable obstacle to comprehensive testing. Furthermore, personalisation can also appear unfair in the context of competition. Dynamic difficulty adjustments, while aiming for a balanced gaming experience, can elicit strong negative reactions from players. A classic example is observed in the Mario Kart series, where the leading player misses out on power-ups, potentially causing them to fall behind towards the end. Despite the frustration this may evoke, it serves to level the playing field and maintain a competitive dynamic (Vicencio-Moreira, Mandryk & Gutwin, 2014). Players who grasp this mechanic may strategically choose to remain just behind the leader, introducing a dynamic shift in the gameplay. Whilst this might be a major problem with personalisation in games, the issue is less relevant in other fields. Nevertheless, it should be noted that giving reviews to personalised story experiences can be difficult, as the experiences can end up vastly different for different users.

In any narrative featuring choices it is important to consider the concept of flow state, commonly discussed in game design, where the user becomes fully immersed in the experience. In creating and maintaining it, it is important to consider interruptions and interface confusion. When a user stops to think outside the experience, the flow is broken, and they become aware of their surroundings. Having to make many choices can be overwhelming and disruptive (Chen, 2007). The interface needs to be well designed not to distract. Banding the users depending on their skill levels can be helpful in game design. User experiments can discover potential flow exit points where anxiety or boredom levels become high. Such boundary cases should then be examined for options that could maintain the flow state (Neal, 2012).

Hallifax et al. (2019) note that applicability of findings on personalisation and gamification faces considerable challenges. Firstly, diverse studies are often conducted in disparate domains, limiting the generalisation of results. Secondly, the reliance on distinct user typologies or personality models across studies introduces variability. Lastly, the lack of consistency in the game elements considered and the examination of different levels of abstraction in motivational strategies further complicates the synthesis and practical application of these findings. Consequently, the diverse contexts and methodologies of these studies hinder the seamless integration of their results into a cohesive framework.

Similarly, personalised narratives in novels have the potential to enhance reader engagement. In the same way that personalised in-game characters captivate players, tailoring character traits, plot developments, and even writing styles to align with individual reader preferences can make stories more immersive and compelling. When readers encounter elements in a story that resonate closely with their personal tastes and experiences, they are likely to feel a stronger connection to the narrative. This connection can lead to a more profound and enjoyable reading experience, as the story feels more relevant and engaging to the individual reader.

Personalisation in novels can also lead to increased reader loyalty. Just as personalised games have been shown to increase player loyalty, offering personalised reading experiences can encourage readers to return to an author or a series. Knowing that a novel will cater to their specific interests and preferences can make readers more likely to seek out future works by the same author or publisher. This loyalty can be further reinforced through adaptive storytelling techniques, where digital and interactive novels use reader choices and feedback to dynamically alter the storyline, creating a unique reading experience for each individual. This mirrors the adaptive AI and randomisation strategies used in games to enhance replayability and enjoyment.

In addition, personalisation enables more effective marketing strategies and recommendations. By understanding reader preferences, publishers can suggest books that are more likely to appeal to individual readers, increasing the chances of a purchase. This approach is similar to how personalised dialogue systems in games align with player preferences to enhance engagement and immersion. Tailoring marketing efforts to individual tastes can make the process of discovering new books more efficient and enjoyable for readers, leading to higher satisfaction and engagement with the publisher's offerings.

Personalisation also has significant potential in educational and persuasive contexts. Just as personalisation in educational games has proven effective in fostering desired behavioural changes, personalised novels can be used for educational and persuasive purposes. Tailoring the narrative to address specific learning objectives or motivational factors can enhance the impact of educational content, making it more relevant and engaging for readers. This approach can be particularly effective in creating narratives that cater to the diverse preferences and motivations of different reader groups, thereby increasing the overall effectiveness of the educational or persuasive message.

3. Player Modelling and Profiling

Player modelling and profiling represent pivotal aspects of contemporary game development, profoundly influencing the design, personalisation, and optimisation of gaming experiences. Player modelling is based on what is recorded during gameplay interaction; player profiling categorises players on static information such as personality, gender and age (Yannakakis et al., 2013: 46). Player profiling models can be used anywhere conventional personality models can be used; the profile covers traits that can be of interest for many applications (Bakkes, Spronck & van Lankveld, 2012: 77). At its core, player modelling involves the construction and application of models to predict and understand player behaviour within gaming environments. These models, ranging from statistical algorithms to computational frameworks, serve as pivotal tools for analysing player-game interactions and preferences. Conversely, player profiling involves categorising players into distinct groups based on shared characteristics or behaviours (Yannakakis & Hallam, 2011). The principles underlying player modelling and profiling can also be applied to written narratives, such as novels, to enhance reader engagement and satisfaction. These methodologies can be adapted to personalising written narratives by using reader data, such as reading habits, preferences, and feedback, to tailor the story to individual readers.

Practically, player modelling and profiling find applications across multiple facets of game development. Personalised gaming experiences are facilitated through adaptive difficulty levels and customised content delivery, ensuring engagement and immersion tailored to individual players. Moreover, predictive analysis and dynamic quest generation contribute to player retention and sustained engagement, optimising the gaming experience for prolonged user involvement. In-game advertising and monetisation strategies benefit from player profiling, enabling targeted advertisements and personalised microtransactions that enhance player engagement and revenue generation.

The methodological landscape of player modelling and profiling encompasses various approaches, each offering unique insights into player behaviour. Data-driven techniques leverage telemetry data and machine learning algorithms to extract meaningful patterns from player interactions. In contrast, rule-based systems rely on expert knowledge and predefined rules to model player behaviour, often guided by game designers' expertise. Hybrid approaches amalgamate data-driven insights with rule-based systems, fostering a holistic understanding of player preferences and tendencies.

Approaches to player modelling can also be broadly categorised into model-based (top-down) and model-free (bottom-up) methodologies, each presenting distinct characteristics. In the model-based approach, a theoretical framework serves as the foundation for constructing a player model. This methodology mirrors approaches employed in humanities and social sciences, where theoretical models are posited to elucidate phenomena, often followed by empirical experiments to validate these models. Conversely, the model-free approach involves the development of an unknown mapping model that correlates user input with a user state representation. This methodology mirrors the empirical processes observed in exact sciences, where observations are systematically analysed to derive models without relying on strong initial assumptions. Interestingly, the majority of existing works in player modelling exhibit a hybrid nature, encompassing elements of both top-down and bottom-up approaches, demonstrating a nuanced integration of theoretical frameworks and empirical observations (Yannakakis et al., 2013: 47). This hybridity underscores the adaptability and versatility required in the dynamic landscape of player modelling research.

Games can offer lots of behavioural indicators to infer a player’s personality. The extensive array of activities available in massively multiplayer online games (MMOGs) has positioned them as virtual laboratories for social science research due to their popularity and the diverse interactions they facilitate (Ducheneaut, 2010). In leveraging MMOGs as research environments, Shen et al. (2012) utilised data from World of Warcraft (Blizzard Entertainment, 2004), employing automated text analysis techniques to build upon Yee et al.'s (2011) work, incorporating textual and social networking data for a comprehensive analysis. Additionally, Canossa, Martinez, and Togelius (2013) delved into the realms of Minecraft (Mojang, 2011), employing life motivation questionnaires, revealing that players' self-reported life motives were reflected in the virtual worlds they constructed within Minecraft. These studies demonstrate the potential of using player data to infer personality traits and preferences, which can also be applied to personalising written narratives, should suitable data be available.

The exploration of game-related data extends beyond in-game activities, encompassing external sources such as game review sites. Sacco, Liapis, and Yannakakis (2016) highlighted the potential of scores and sentiment-analysed textual reviews from platforms like Metacritic or GameRankings as valuable inputs for models. These models can then be instrumental in crafting game content tailored to specific demographics or interests, derived from players' in-game achievements or preferred games. This nuanced approach underscores the depth of insights that can be gleaned from diverse sources within and outside the gaming environment, providing researchers with multifaceted perspectives for understanding player behaviour and preferences.

Demographic factors such as gender, age (Yee, 2006a) and nationality (Bialas, Tekofsky & Spronck, 2014) are taken as uncontested in player profiling, but this is quite not the case with the role of personality (Yannakakis & Togelius, 2018b: 218-219). Some factors seem well established though – for example, the link between personality and being drawn to violent video games (Markey & Markey, 2010), and the differences between Introverts and Extraverts. Denden et al. (2017), for example, found differing preferences in game elements for Introverts and Extraverts, such as Extraverts preferring elements such leaderboards and progress bars. In-game behaviour might not match with real-life behaviour (e.g. van Lankveld et al., 2010; 2011), but many studies have found strong correlations between them (e.g. Yee et al., 2011; Tekofsky et al., 2013). It can perhaps be concluded that some mapping between in-game behaviour and personality can be done, but it cannot be expected to be one-to-one. These insights can be applied to personalising written narratives by tailoring story elements to align with the personality traits and preferences of individual readers.

For personalisation based on playstyle, static approaches calculate playstyles based on self-reports made before gameplay, with the player unaware of the process (Birk et al., 2015; Magerko et al., 2008). Bontchev and Georgieva (2018), however, argue that automatic style recognition during the play is much more promising, being done by analysing individual player interactions and achieved results implicitly at runtime, and enabling dynamic style-based adaptation of various features. This approach could be mirrored in written narratives as well, especially interactive narratives, by dynamically adapting the storyline based on reader interactions and feedback, creating a personalised and engaging reading experience.

The stereotype approach (Kobsa, 1993) is a common method for using player profile information, assigning a player to a subgroup of the population, represented by a stereotype, the key characteristics of which have been previously defined. Appropriate responses to each stereotype can then be decided, and when the gameplay is taking place, the model can be adjusted to fit the individual and not just the stereotype anymore. Examples of the approach include Yannakakis & Hallam (2007) and Thue et al. (2007), who relate the stereotypes to the players' gaming profiles, rather than their characteristics in real life.

One way of doing profiling is to use a factorial model, manually partitioning data space to attach different meanings to various aspects of the data. There would be several factors, each informed by certain game variables that tell us something about the player. Charles et al. (2005: 14) offer the example of a player with a numerical profile of (0.8, 0.2, 0.2, 0.5) who tends to avoid direct, close-up conflict. However, figuring out which variables to use and how might not be straightforward. A skilled game designer might do this intuitively or by trial and error, but data mining tools or unsupervised statistical techniques such as factor analysis could also be useful, identifying correlations between the variables, helping the designer attach a meaning to them.

Despite their utility, player modelling and profiling present several challenges and ethical considerations. Concerns regarding data privacy and security necessitate robust safeguards to protect players' personal information. Algorithmic bias poses a significant risk, requiring measures to ensure fairness and transparency in decision-making processes. Additionally, the overreliance on models may overlook the nuanced and evolving nature of player interactions, highlighting the importance of human oversight and intervention.

Looking towards the future, advancements in machine learning techniques hold promise for enhancing player modelling and profiling capabilities. Deep learning and reinforcement learning offer opportunities for developing more sophisticated systems capable of real-time adaptation to evolving player behaviour. Cross-platform profiling emerges as a potential trajectory, enabling unified player profiles across different gaming platforms for seamless gaming experiences. Furthermore, the integration of biometric data for emotion recognition presents avenues for creating more immersive and emotionally resonant gaming experiences.

The various studies on player modelling and profiling suggest that these methodologies offer valuable insights into player behaviour and preferences, which can be leveraged to create personalised gaming experiences. Many of these approaches can also be applied to personalising written narratives by tailoring the story to individual reader preferences and characteristics. This approach has the potential to enhance reader engagement, satisfaction, and loyalty, creating a more immersive and compelling reading experience.

4. Psychological Models

Psychological models can offer an empirical, top-down approach to personalisation. The most commonly used psychological models of personality used in personalisation, player modelling and related topics have been the Five-Factor Model (FFM) (John, Donahue &

Kentle, 1991; John, Naumann & Soto, 2008) and the Myers–Briggs Type Indicator (MBTI) (Myers, 1962). Studies utilising these models have demonstrated significant strengths. One major strength is the comprehensive understanding they provide into user behaviour, allowing for detailed predictions and personalisation strategies. This understanding is particularly useful in domains such as gaming and interactive storytelling, where the alignment of content with user traits can greatly enhance engagement and satisfaction. The versatility of these models, especially the FFM, also stands out, as they are applicable across a wide range of fields including literature, film, and music, beyond their traditional use in psychology.

However, these studies also reveal certain weaknesses. A notable limitation is the potential over-reliance on static personality traits, which might not fully account for the dynamic nature of user behaviour. This can lead to simplistic or stereotypical personalisation strategies that may not adapt well to changes in user preferences over time. Additionally, the cultural and contextual limitations of these models can affect the accuracy of personalisation efforts, as personality traits may manifest differently across different cultural settings. Furthermore, the use of psychological profiling raises ethical concerns, particularly regarding privacy and data security. There is a need for rigorous safeguards to protect user data and ensure that personalisation does not become intrusive or manipulative.

The relevance of psychological models to the broader thesis on personalisation is substantial. These models provide a structured approach to understanding and categorising user preferences and behaviours, which is essential for developing personalised experiences in digital content and interactive environments. By leveraging well-established models like the FFM and MBTI, developers and researchers can design content that is more engaging and closely aligned with individual user characteristics.

Other models from psychology have also occasionally been used in personalisation: Gómez-Gauchía & Peinado (2006), for example, had users take a questionnaire based on the temperament theory of David Keirsey (1998), widely used for selecting job candidates, to personalise non-playable characters, NPCs. They found Keirsey's model more relevant than the FFM for video games since they consider acting in them usually more important than thinking, and Keirsey's theory is focused on what people do rather than think. Nevertheless, in player modelling, the FFM has been by far the most widely used model. Furthermore, we also raise the prospect of using the Need for Affect (Maio & Esses, 2001), which considers individuals' tendencies toward seeking or avoiding emotional experiences, adding further depth to the personalisation strategies. This model can enrich the personalisation framework by catering to the emotional needs of users, potentially leading to more resonant and immersive experiences.

4.1. The Five-Factor Model

In the realm of personality psychology, the Five-Factor Model (FFM), often referred to as the Big Five, stands as the cornerstone framework. It was developed by many independent sets of researchers who analysed words describing people's behaviour (Digman, 1990). This model encapsulates five fundamental traits that are widely acknowledged to encompass the breadth of human personality variation. These traits include Extraversion, which reflects the extent to which individuals are outgoing, sociable, and assertive; Agreeableness, which pertains to the degree of kindness, cooperativeness, and empathy individuals exhibit towards others; Conscientiousness, which denotes the level of organisation, responsibility, and self-discipline individuals demonstrate in their actions; Emotional Stability, also known as Neuroticism, which gauges the degree of emotional resilience, calmness, and stability individuals possess in the face of stress or adversity; and Openness to Experience, which captures individuals' receptivity to new ideas, curiosity, and imagination.

The Five-Factor Model serves as a robust framework for comprehensively understanding and categorising various aspects of personality. Each of these five traits represents a continuum, with individuals falling somewhere along the spectrum for each trait. This model provides a nuanced and comprehensive lens through which psychologists can explore and analyse the complexities of human personality, offering insights into individual differences and behavioural tendencies across diverse populations. It typically uses surveys based on sets of questions, such as the International Personality Item Pool (Goldberg, 1999). It is also possible to identify relevant behavioural cues by observing, for example, the user’s interaction (Dunn et al., 2009) or speech and language (e.g. Argamon et al., 2005). Personality questionnaires tend to have a high predictive value, but they are not objective like observer reports and require effort from the user (Mairesse & Walker, 2010: 2).

These traits would appear to influence preferences for different features in narratives and media. Weaver (1991) found that people with high scores in Neuroticism, the opposite of Emotional Stability, had a strong preference for sad music and avoided light-hearted film genres such as comedy and action or adventure. Psychoticism, the opposite of Agreeableness, indicated a preference against comedy but one strongly in favour of graphically violent horror movies. Media preference profiles could also successfully discriminate between different levels of Neuroticism and Psychoticism. Similarly, Gunter (1985) reported Neuroticism to indicate less preference for violent film clips. Zuckerman and Litle (1986) found evidence that Sensation Seeking (represented by Openness and Extraversion) involved a preference for erotic, violent and frightening films, concluding that they prefer novel and arousing media across genres. Various correlations between Big Five traits and types of media have also been found in several later studies, such as Rawlings and Ciancarelli (1997), Rentfrow, Goldberg and Zilca (2011) and Rentfrow and Gosling (2003).

The study by Soto-Sanfiel, Aymerich-Franch, and Romero (2014) examines earlier research on how FFM personality traits correlate with media consumption habits. People with high levels of Extraversion typically engage less with traditional media forms such as television, radio, or leisure reading. They prefer media that satisfy their need for social interaction, favour in-person activities over mediated communication, and show greater interest in reality shows. Individuals high in Openness, who are intellectually curious and open to new experiences, gravitate toward unconventional activities and media, like artistic, informative, or erotic content. They tend to dislike soap operas but appreciate violent media for its aesthetic qualities, showing a preference for new and innovative content over traditional media. Those high in Agreeableness favour direct social contact over media use, react negatively to sensational or distressing content, and find interactive narratives fitting for television. On the other hand, Conscientious individuals, known for their discipline and dependability, do not display a clear pattern of traditional media use, with some researchers suggesting this trait may not significantly affect television consumption and is sometimes overlooked in communication studies.

Their 2014 study on interactive narratives also examines how personality traits influence the enjoyment of fiction, particularly in the context of story ending choices. Consistent with previous findings, people high in Extraversion tend to prefer stories with happy endings, while highly Agreeable individuals show stronger emotional reactions to tragic stories. Those scoring high in Openness, however, derive greater enjoyment when choosing tragic endings, which reflects their attraction to complex and challenging situations. This is in line with their interest in interactivity, commitment, and violent content with aesthetic appeal. The study also explores how individuals identify with characters, revealing that those high in Extraversion, Openness, and Agreeableness have stronger identification. Additionally, high Openness and Agreeableness are associated with greater cognitive-emotional empathy, enhancing their connection with the characters. People with high Extraversion and

Agreeableness report higher levels of enjoyment, while those high in Conscientiousness experience less. The study acknowledges the limitation of its small sample size, suggesting that a larger sample could improve the reliability of its findings.

Annalyn et al. (2018) investigated the association between personality traits and book preferences based on user-generated tags. The results indicate that Extraverts exhibit a preference for books with social themes, such as relationships and chick lit, while Introverts tend to gravitate towards fantasy, science fiction, and supernatural forces, indicating a proclivity for imaginative content. Agreeable individuals show a preference for books with family and religious themes, whereas less Agreeable individuals are attracted to dark-themed content and cult classics with controversial narratives. Open individuals lean towards intellectually challenging and classic literature, contrasting with individuals scoring lower on Openness, who prefer mainstream and less cognitively taxing content. Neurotic individuals are drawn to narratives reflecting emotional states and alternative realities, with an added preference for books with aesthetically pleasing covers. Lastly, Conscientious individuals favour informative content contributing to professional development, while those with low Conscientiousness scores prefer light-hearted and youth-oriented books. The study emphasises the role of book preferences in predicting personality traits, shedding light on cultural differences in reading preferences and supporting the use of user-generated data for comprehensive audience profiling.

Michelson (2014) discusses how when investigating the artistic preferences of Extraverts, researchers observe a tendency for them to choose more novel or exciting options in paintings, sculptures, or poetry. However, when it comes to actual engagement with the arts, Extraversion is found to have a negative or neutral correlation with most aesthetic activities. Introverts, in contrast, show a preference for fiction reading. Neuroticism predicts a preference for emotionally positive fiction. Agreeableness is associated with people-centred genres like romance novels and soap operas. Conscientiousness correlates with a preference for predictable formats and structures in reading, while Openness to Experience is linked to a greater interest in complex literature and aesthetic experiences.

Teng (2009) found the FFM traits of Openness, Conscientiousness and Extraversion as relating to playing online, but notes that if such personality measures obtained in a game context differ from those obtained in a real-world context, the results could be different. Zammitto (2010), however, had different results for Openness and Conscientiousness, and noted that personality factors in her study explained only 2.6-7.5% of game preferences. Hirsh, Kang and Bodenhausen (2012) found that a message can be more persuasive if it is aligned with the recipient's personality profile, using the FFM personality dimensions. Desire for excitement and social rewards was a stronger factor Extraverted people, connection with family and community for Agreeable people, efficiency and goal pursuit for Conscientious people, safety and security for Neurotic people, and creativity and intellectual stimulation for people with Openness.

The Five Domains of Play theory (5D) (Vandenberghe, 2012) translates the Big Five into five aspects of gaming motivation: people with a high score in Openness to Experience seek novelty; Conscientiousness matches with challenge; Extraversion with stimulation; Agreeableness with harmony; and Neuroticism with threat. De Vette et al. (2016) tested the model using an online questionnaire with the 10-item Big Five Inventory (Goldberg, 1999), and five questions on the game preferences. For participants younger than 60, four out of five personality traits correlated significantly but weakly with their corresponding game preference domains ($r=0.13\text{-}0.30$, $p<0.05$). Agreeableness showed no significant correlations, and older participants lacked gaming experience.

Nagle, Wolf & Riener (2010) applied the Big Five for adjusting difficulty in a first-person shooter game, with a linear regression model aiming to optimise enjoyment and gameplay duration. Van Lankveld et al. (2011) modified Neverwinter Nights, a third-person role-playing video game, to observe whether they could find a correlation between players' gameplay metrics in the game and their FFM profile, finding they could indeed model all the five factors but noting that behaviour outside the game could still be different. De Lima, Feijó & Furtado (2018) also created a method for creating a FFM profile during gameplay for the players, which are then used to define their quests.

Various other online activities, such as personal websites (Vazire & Gosling, 2004), Facebook profiles (Back et al., 2010), emails (Gladis, 1993) and even email addresses (Back, Schmukle & Egloff, 2008), can also reveal a person's personality to human observers. In fact, computer-based personality judgments are more accurate than those made by humans, according to Youyou, Kosinski and Stillwell (2015), who created an algorithm they found to accurately predict personalities simply based on likes on Facebook, as found in the MyPersonality dataset. Computer-based judgments had higher correlation ($r = 0.56$) with subjects' self-ratings than human judgments did ($r = 0.49$). The likes are predictive of preferences aligned with the Big Five theory: for example, people with high Openness to Experience often like Salvador Dalí, meditation, and TED talks, and high Extraversion is linked with liking partying, dancing, and the reality show star Snookie (Youyou, Kosinski & Stillwell, 2015: 1037).

These studies examining various online activities highlight the potential for inferring personality traits from digital behaviour, demonstrating that these digital footprints can reveal significant insights into an individual's personality. This is highly relevant to the central thesis on personalisation, providing empirical evidence supporting the feasibility of using digital data to assess personality traits. This capability is crucial for developing sophisticated personalisation systems that tailor content and interactions to individual users based on inferred personality traits. The demonstration of computer-based personality judgments outperforming human assessments underscores the potential for automated systems to enhance personalisation efforts, making them more precise and scalable. Thus, these findings are integral to understanding how digital footprints can be leveraged in personalisation strategies across various domains.

4.2. The Myers–Briggs Type Indicator

A model that enjoys widespread popularity within business and the public around the world is the Myers–Briggs Type Indicator, or the MBTI. It is based on Jung’s personality type theory and was first published by Isabel Briggs Myers and Katharine Cook Briggs in 1962 (Myers, 1962). The origins of the MBTI trace back to the early 20th century when Katharine Briggs became interested in the theories of Swiss psychiatrist Carl Jung. Jung's work on psychological types, outlined in his book *Psychological Types* (1921), laid the groundwork for the development of the MBTI. Briggs, along with her daughter Isabel, sought to create a practical tool that could help people understand themselves and others better. The first version of the MBTI was officially published in 1962, and it gained popularity as a tool for personal development, career counselling, and team-building exercises. The underlying philosophy of the MBTI is rooted in the belief that understanding one's personality type can lead to improved communication, decision-making, and overall well-being.

It consists of four dimensions, the first one being Extravert (E) vs. Introvert (I), similarly as in the FFM. The second is Sensation (S) vs. Intuition (N). It is about whether a person prefers to process facts with their senses or find deeper meanings. Sensing types rely on concrete and factual information, focusing on the present reality. Intuitive types are more inclined to interpret and add meaning to information, often considering future possibilities. The third dimension is Thinking (T) vs. Feeling (F), referring to how a person makes decisions. Thinking types prioritise logic and objective analysis, aiming for fairness. Feeling types, on the other hand, consider the impact of decisions on people and focus on harmony and empathy. The final dimension is Judgement (J) vs. Perception (P), which is about whether the person tends to deal with the world using their second dimension, or the third. Judging types prefer order, planning, and decisiveness, while Perceiving types are more adaptable, spontaneous, and open to new information. The MBTI is dichotomous; each person is judged to be either one type or the other in each of the four dimensions. Therefore, there are sixteen possible MBTI personality types.

Despite its widespread popularity, the MBTI has faced scrutiny over test validity and reliability. For example, research by Pittenger in 1993 revealed substantial variability in test results among individuals tested at different times. This controversy has sparked debates within the scientific community, challenging the instrument's consistency. One major criticism is the dichotomous nature of the categories, which may oversimplify the complexity of human personality. Critics argue that people fall on a spectrum for each dichotomy, and forcing them into rigid categories may limit the accuracy of the assessment. Cultural variations in the interpretation of MBTI results add another layer of complexity. While the tool is widely used in the United States and Europe, its applicability and relevance in other cultural contexts may be limited. The concepts of individualism and collectivism, for example, may influence how traits like Extraversion or Introversion are perceived.

In studies on and related to personalisation systems, the FFM is often preferred to the MBTI due to its empirical foundation and ability to provide more precise personality insights. However, the MBTI's enduring popularity and accessibility, particularly in non-academic settings, make it a valuable resource for generating broad datasets on individual preferences and behaviours. For instance, it has been noted that algorithms trained on MBTI data can sometimes outperform those trained on FFM data in specific applications, such as text-based personality prediction (Celli & Lepri, 2018). This suggests that while the MBTI may have limitations in terms of scientific rigor, its widespread use and the familiarity of its categories can still provide practical benefits in certain domains.

The relevance of these models to the thesis lies in their application to personalisation techniques, where understanding user personality can significantly enhance the tailoring of experiences and content. The choice between MBTI and FFM should be guided by the specific requirements of the application, including the need for precision, the nature of the data available, and the cultural context. As personalisation technologies continue to evolve, integrating insights from both frameworks could offer a more holistic approach to understanding and catering to individual differences.

4.3. The Need for Affect

Feeling sad while watching a sad film has been found to correlate with enjoying the film, according to Oliver (1993), in whose later study (Oliver, Weaver & Sargent, 2000) the participants preferring sad films enjoyed the sadness, but for those who did not like them, sadness was inversely related to enjoyment. Similarly, a meta-analysis of 35 articles (Hoffner & Levine, 2005) found that the people taking the most pleasure in horror films are also the ones who report the most negative affect during watching them. This indicates that people who enjoy painful art enjoy it precisely because of the pain.

There have been various studies and versatile results on the topic of what sort of personal characteristics explain enjoyment of negative emotions in art. For example, De Wied, Zillmann, and Ordman (1994) found that people with high empathy enjoyed tragic films more. Garrido and Schubert (2011) studied how individual differences in empathy, absorption, fantasy proneness, rumination, and dissociation, affected enjoying negative emotion in music, reported by half of the 59 participants. Absorption and music empathy were the best predictors of enjoyment. Vuoskoski et al. (2012) discovered a preference for sad music indicated high Openness to Experience and Empathy. In Rentfrow and Gosling (2003), a preference for intense and rebellious music such as heavy metal indicated high Openness to Experience and Agreeableness but low Neuroticism. Thompson, Geeves and Olsen (2018) found that fans of violent Death Metal music had lower Conscientiousness and Agreeableness levels.

One promising factor could be the Need for Affect (NFA), which refers to how motivated people are to seek emotion-inducing situations and activities (Maio & Esses, 2001). Media use or preference is not a part of its definition or operationalisation, but the study did have participants rate their willingness to see specific films after having read descriptions of how interesting, happy, and sad each of them was supposed to be. The willingness to see happy and sad films rather than less emotional films was higher for individuals with a strong NFA. They note that while sensation seeking is conceptually similar to the NFA, they found them empirically distinct.

Appel (2008) focused on preference for media with affectively negative content, using film synopses using emotional adjectives from the positive and negative affect schedule

(Watson, Clark & Tellegen, 1988), or no emotional adjectives. High NFA scores predicted willingness to watch films with affectively negative rather than positive or neutral content, but only in females. However, the study was limited by its reliance on intended exposure based on hypothetic descriptions of films, rather than seeing actual films. Bartsch, Appel & Storch (2010), on the other hand, had participants watch drama or horror in the cinema, finding that people with high NFA had higher levels of negative and ambivalent emotions, but evaluated these emotions more positively on a metaemotional level, meaning they enjoyed them more. They view the NFA as the first personality trait that has been found to be a consistent predictor of individuals' engagement with negative and ambivalent emotion experiences regardless of gender or genre. The significance of the NFA remained even when the Big Five was statistically controlled for.

5. Player Type Models

Understanding player behaviour and preferences is fundamental to creating immersive and engaging gaming experiences. Player type models have emerged as valuable tools in game design, enabling developers to categorise players based on their motivations, preferences, and playstyles. By segmenting the player base into distinct types, player type models facilitate personalised game experiences, adaptive content delivery, and targeted design strategies.

Player type models, also known as player taxonomies or player archetypes, aim to capture the diverse range of player motivations and behaviours observed within gaming communities. These models draw inspiration from psychological theories, personality frameworks, and behavioural analytics to categorise players into meaningful groups (Tondello et al., 2016, 2019). By identifying common player characteristics and playstyles, player type models offer insights into player preferences, challenges, and engagement drivers.

Player type models can be used in at least two situations: assessing the distribution of player types in the target audience, helping to make game design decisions; and personalisation. Player type models, just like personality trait models, seek to capture more or less stable individual differences, but are more specific, seeking to explain differences in behaviour in limited circumstances.

The significance of player type models in game design lies in their ability to inform tailored design decisions, content personalisation, and player engagement strategies. By recognising the diverse motivations and preferences of players, developers can create adaptive gaming experiences that cater to individual needs and preferences. Player type models enable the implementation of personalised game mechanics, dynamic difficulty adjustment, and targeted content delivery, enhancing player satisfaction and retention. Moreover, player type models facilitate player segmentation and community management, enabling developers to design inclusive and diverse gaming environments that accommodate different playstyles and preferences. By understanding the unique needs and motivations of various player types, developers can foster positive social interactions, community engagement, and collaborative gameplay experiences.

The exploration of player type models within the context of this thesis serves as an important avenue for evaluating how these models might be leveraged to tailor narrative experiences in various creative domains, particularly in gaming. Player type models, which categorise individuals based on their motivations, preferences, and playstyles, offer a structured approach to understanding how different types of players engage with interactive narratives. This opens up the possibility of customising narrative elements to enhance engagement, immersion, and overall satisfaction.

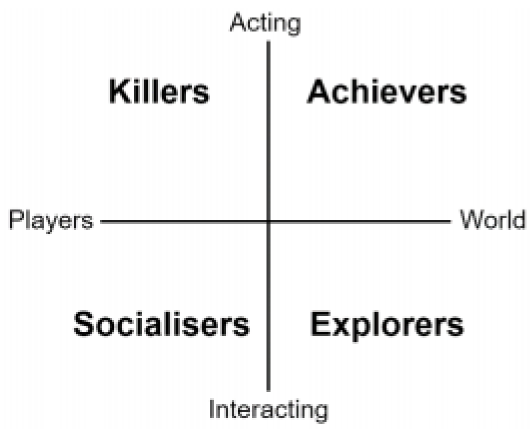
This adaptability is not limited to gaming. In fact, the boundaries between games and interactive narratives are increasingly blurred. Story-driven games like *The Witcher 3* (CD Projekt Red, 2015) and *Red Dead Redemption 2* (Rockstar Games, 2018) serve as prime examples of narratives that evolve based on player choices, offering different experiences based on how players interact with the game world. These games function as interactive narratives where story elements respond to the player's decision-making process. The more responsive the game is to the player's personality or type, the deeper the sense of immersion and personal connection becomes.

This concept of integrating player type models to enhance personalised narrative experiences is especially relevant given the increasing overlap between gaming, literature, and interactive media. As games become more narrative-rich and interactive storytelling becomes more prevalent across media platforms, player type models can serve as a blueprint for how personalisation can enhance user experience across domains. The flexibility of these models makes them particularly valuable in designing not only games but also virtual reality experiences, interactive films, and multimedia art installations, where personalisation can directly impact how narratives are perceived and engaged with.

5.1. Bartle Types

Possibly the most commonly used player personality categorisation is the Hearts, Clubs, Diamonds and Spades model by Bartle (1996), dividing players into Killers, Socialisers, Achievers and Explorers according to two dimensions of playing style: action versus interaction, and world-oriented versus player-oriented. Achievers seek rewards with little or no gameplay benefit, simply for the prestige. Explorers prefer discovering areas, Easter eggs and glitches, as well as creating maps. Socialisers might use the game simply for meeting other players. Killers, on the other hand, are competitive, particularly against human opponents. Bartle did not test the model empirically on independency of the types, or on psychometric quality criteria, but the model has later been used in many studies, and as a starting point for creating new classifications. Some have considered the types too restrictive to be used in many types of games (Kyatric, 2013); it was indeed built considering just players of Multi-User Dungeons (MUDs), a rather old-fashioned, text-based type of games. Bartle (1996) does not share the concerns, but considers the model very generalisable. Yee (2006b), however, notes that when Bartle's questionnaire asks respondents to choose between Achiever and Explorer patterns, what results is a dichotomy which might not exist in reality.

Figure 1: Bartle types (Bartle, 1996).



Later, Bartle (2004) extended the model by adding a third dimension, Implicit/Explicit, which indicates whether the player behaves in an unconscious or considered manner, splitting the original types into two. The roles can change depending on the situation in the game. The implicit types are Opportunists, Hackers, Friends and Griefers, and the explicit types are

Planners, Scientists, Networkers and Politicians.

Yee (2006b) used a factor-analytic approach of questions based on Bartle’s player types, identifying three main components and ten subcomponents of player motivation with some, though low statistical validity. The resulting Gamer Motivation Profile introduced a multidimensional model comprising eight motivational factors, including Achievement, Social, Immersion, and Competition, providing a more nuanced understanding of player behaviour. Indeed, Bartle’s approach has inspired many followers; one of many examples could be Stewart (2011), who combined Bartle's four player types, the four Keirsey Temperaments (Keirsey & Bates, 1984), and the demographic game design model by Bateman, Lowenhaupt and Nacke (2011). The four categories of players in the Stewart model are Artisan/Killer/Experientialist, Guardian/Achiever/Gamist, Rational/Explorer/Simulationist and Idealist/Socialiser/Narrativist.

5.2. Demographic Game Design Models

The first Demographic Game Design model, known as DGD1 (Bateman & Boon, 2005) is an adaptation of Myers-Briggs typology to games. It uses types drawn from the MBTI: Conquerors, who enjoy control and competition, are linked to the MBTI preferences Thinking and Judging; Managers, who enjoy strategic planning and seek to master the game, are linked to Thinking and Perceiving; Wanderers, who explore and play for fun, to Feeling and Perceiving; and Participants, who play games for the company, to Feeling and Judging. The second Demographic Game Design model (DGD2) (Bateman, Lowenhaupt & Nacke, 2011) added a Hardcore/Casual dimension as well consideration for different skill sets and preference for single or multiplayer.

In a study conducted by Dias and Martinho (2011), participants engaged in a multifaceted exploration involving the MBTI and gameplay experiences. The participants initially completed a survey to ascertain their Myers-Briggs personality type. Following this, they immersed themselves in the gaming environment of a game inferring their DGD1 player type from their behaviour and then selecting how content is managed and presented to the player based on the inferred player type, finding this improved enjoyment.

The DGD1 model categorises players into specific types, such as Wanderers and Participants, each characterised by unique gameplay features, such as less severe penalties for deaths. The study involved dividing participants into two groups: one group played the game tailored to their identified player type, while the other group experienced a version adapted for a different player type. The results indicated that the group playing the game aligned with their identified player type reported significantly higher levels of enjoyment and immersion.

McMahon, Wyeth, and Johnson (2012) further explored the connection between the DGD1 model and the Five-Factor Model, positioning it as a more robust alternative to the MBTI. Drawing on the work of McCrae and Costa (1989), who found the FFM to be more reliable, the study identified correlations between certain DGD1 player types and FFM traits. For instance, Conquerors, Managers, and Participants exhibited high levels of Conscientiousness, while Participants and Managers displayed a proclivity for Openness to Experience. It is important to note that the study encountered a limitation, namely the absence of a validated measure for the DGD1 player types. Participants had to self-select their player type based on simple descriptions, introducing a potential source of subjectivity. This highlights the need for further research and refinement of frameworks like DGD1 to ensure more precise and reliable categorisation of player types for future studies and applications.

5.3. Brainhex

Brainhex, introduced by Nacke, Bateman, and Mandryk in 2011, presents a distinctive approach by shifting from subjective measures of emotions to incorporating measures of neurobiological responses. Drawing inspiration from existing player typologies and the extensive literature on game emotions, Brainhex identifies seven distinct archetypes. These archetypes encapsulate various player preferences and inclinations: Achievers are driven by goals, Conquerors relish challenging opponents, Daredevils seek excitement and risks, Masterminds enjoy puzzles and strategic thinking, Seekers have a penchant for exploration, Socialisers thrive on social interactions, and Survivors find enjoyment in frightening experiences.

Despite the intriguing framework proposed by Brainhex, Rogers, Kamm and Weber (2016) encountered challenges when implementing activities tailored for specific player types. Their study failed to establish a clear correlation between the Brainhex player types and the corresponding interest levels in the designed activities. The complexities involved in predicting player experiences based on these archetypes surfaced as a notable hurdle.

Another study by Busch et al. (2016) focused on a personalised location-based game, specifically targeting two player types from the Brainhex model: Mastermind and Seeker. The findings indicated that these player types were not robust predictors of the player experience for personalised missions. The conclusion drawn from this research suggests that player type models, including Brainhex, require both conceptual and empirical refinement to ensure their validity. This includes comprehensive coverage of player personalities, enhancement in relating to traits, and addressing concerns related to exhaustiveness and non-redundancy. The quest for a more accurate and reliable player typology remains an ongoing challenge in the field of game research.

5.4. HEXAD

Building upon established psychological theories, including Pink’s four drives theory (Pink, 2009) and self-determination theory (Ryan & Deci, 2000), the HEXAD framework posits six core player types: Achievers, Socializers, Philanthropists, Free Spirits, Players, and Disruptors (Tondello et al., 2016). The HEXAD framework departs from traditional player typologies by emphasising the dynamic and evolving nature of player motivation. Rather than categorising players into fixed archetypes, the framework acknowledges the fluidity and complexity of player preferences, recognising that individuals may exhibit multiple motivations across different contexts and experiences. One of the key strengths of the HEXAD framework lies in its integration of psychological principles and empirical research.

The HEXAD scale initially comprised 24 items, but concerns regarding survey fatigue and dropout rates prompted the researchers to develop a shorter, more concise version of the scale resulting in a 12-item version of the HEXAD scale, known as HEXAD-12. It was created through exploratory factor analysis and validated through confirmatory factor analysis (Krath et al., 2023).

5.5. Criticisms of Player Typologies

While all these prior player preference models provide useful insights towards understanding different player motivations, most of them are limited by the lack of empirical validation, unavailability of a standard assessment tool, or are only suitable for a specific game or genre. In addition, they fail to consider different elements of play that have surfaced more recently, such as body movement-controlled games, and different styles of play, such as electronic sports or casual games.

Hamari and Tuunanen (2014) noted that most player typologies have the same limitations. Most of them have taken cue from Bartle's (1996) original work, which was only ever intended for its context of MUDs and was never empirically validated. The motivational factors considered tend to be at least similar to Bartle's. Most studies consider just online games, typically massively multiplayer online games (MMOs). They also tend to classify players in a handful of distinct types, which is an imprecise way of representing complex creatures with various traits and interests.

One of the primary criticisms of player typologies is their tendency to oversimplify the complex and multifaceted nature of human motivation. Critics argue that reducing individuals to predefined categories fails to capture the richness and variability of their psychological states and behavioural tendencies. By pigeonholing players into discrete typologies, researchers risk overlooking the nuanced interplay of personal, situational, and contextual factors that influence player behaviour. Critics also point to the inherent subjectivity involved in categorising players based on their motivations. Player typologies do not tend to consider personality or motivation until after the typology has been developed (Tondello et al., 2016).

Typological frameworks often rely on researchers' interpretations of player behaviour and motivations, which may be influenced by their own biases, assumptions, and theoretical perspectives. This subjectivity introduces potential sources of error and ambiguity into typological analyses, undermining the validity and reliability of the resulting classifications. Furthermore, the construction of player typologies often relies on self-report measures, such as surveys or questionnaires, which may introduce biases and inaccuracies into the data. Self-reported motivations may not always align with actual behaviour, leading to discrepancies between stated preferences and observed actions. Additionally, the use of fixed-choice response formats in surveys may limit participants' ability to express the full complexity of their motivations, potentially resulting in oversimplified or misleading results. Bateman, Lowenhaupt and Nacke (2011) found that type theories have usually proven inadequate, and future player typology will need foundations in the form of a new trait theory of playing preferences, rather than deploying existing psychological models. Yannakakis et al. (2013), however, suggest determining a fundamental personality model for game behaviour that would have some correspondence with the Five-Factor Model, but also cover different characteristics.

Another criticism of player typologies pertains to their generalisability across different cultural, demographic, and contextual settings. Typologies developed in one cultural context may not fully capture the motivational dynamics present in other cultures, leading to challenges in applying typological frameworks across diverse populations. Moreover, the evolving nature of gaming culture and technology complicates efforts to develop universal typologies that remain relevant and accurate over time. Additionally, some argue that player typologies risk perpetuating stereotypes and stigmatising certain player groups. By associating specific motivations with particular player types, typologies may inadvertently reinforce preconceived notions about gamers and their preferences. This can lead to unfair or inaccurate characterisations of individuals based on their gaming habits, potentially contributing to social stigma and discrimination within gaming communities.

While player type models offer insights into player preferences, they may not be as effective for personalising literary narratives as psychological models. One reason is that player type models often simplify the complexities of human preferences and behaviours into broad categories, which may not adequately capture the nuanced and multifaceted nature of individual narrative engagement. Furthermore, player type models are generally more suited to game design contexts where player interaction patterns are well-established and can be more easily segmented.

In contrast, psychological models, particularly those based on established theories like the Five-Factor Model (FFM) of personality, provide a more granular and empirically validated framework for understanding individual differences. These models are not only more comprehensive but also offer deeper insights into intrinsic motivations and cognitive styles, which are crucial for creating more nuanced and personalised narrative experiences. Thus, while player type models contribute to the discourse on personalisation, psychological models are likely to be more valuable for the specific goal of tailoring literary narratives to individual players.

6. Procedural Content Generation

Procedural Content Generation (PCG) is an innovative and rapidly evolving field within computer science and game development, transforming the traditional approach to content creation through the use of algorithms and computational methods. The fundamental idea behind PCG is to leverage these algorithms to dynamically generate various elements of a game or virtual environment. Instead of relying on manual, static creation, PCG enables the automatic generation of content, providing developers with new tools and possibilities for creating immersive and diverse experiences. PCG stands out as a cost-efficient method for dynamically creating game content, offering a means to deliver vast amounts of algorithmically generated content without excessively consuming memory resources or development time. It is increasingly being done with the assistance of machine learning. By training models on existing content, developers can create algorithms that learn patterns and generate new content aligned with learned aesthetics or structures. This approach adds an element of adaptability and learning to the procedural generation process (Khalifa et al., 2019).

PCG is not limited to games, however; its applications extend to various fields, including content creation for virtual reality simulations, architectural design, and music composition. In virtual reality, procedural techniques can generate realistic landscapes or cityscapes, providing a more immersive experience for users (Shaker, Togelius & Nelson, 2016).

Narrative PCG (NPCG) specifically refers to the automatic creation of story elements, including plotlines, character arcs, and dialogue, within an interactive medium. NPCG's primary advantage is its potential to tailor narrative experiences dynamically based on user interactions, choices, and preferences. Narrative PCG offers a method to personalise literary narratives in real-time, adjusting the storyline to better suit individual player's psychological profiles and preferences. Thus, PCG has potential to revolutionise how interactive narratives are crafted and experienced, moving beyond static, pre-written content to a more dynamic and personalised storytelling approach.

One of the primary applications of PCG is in the generation of game levels, where algorithms can dynamically create levels, maps, characters, items, and narratives on-the-fly or in real-time. This departure from manual design introduces an element of unpredictability and randomness, enhancing replayability and creating novel experiences for players (Shaker, Togelius & Nelson 2016). PCG has a longstanding history in games, dating back to Rogue (Toy, Wichman & Arnold, 1980), and modern examples, such as Minecraft (Mojang, 2011) and Love (Steenberg, 2010), showcase almost entirely procedurally generated environments (Yannakakis et al., 2013: 54). PCG addresses the challenge of keeping content fresh and engaging for players, mitigating the issue of player fatigue in games with extensive playtime. Procedural generation introduces variability, ensuring that players encounter new challenges and environments even after extended gameplay (Green, 2016).

The success of PCG lies in its adaptability and versatility. Developers can tailor procedural algorithms to suit the specific needs and goals of a project, whether aiming for a particular aesthetic, level of difficulty, or narrative structure. This flexibility is particularly advantageous when dealing with limited resources or tight development schedules (Smith et al., 2011). The key advantage of PCG is its ability to efficiently create vast and diverse game worlds. By using algorithms, developers can generate expansive landscapes or intricate mazes without the need for manual design, particularly beneficial for open-world games where vast, explorable environments are crucial for player experience (Yannakakis & Togelius, 2018a). This technique can surpass the designer's imagination, either as a standalone process or through mixed-initiative design approaches, as demonstrated in Smith, Whitehead & Mateas (2011).

The fusion of PCG with search algorithms has the potential to unearth novel and enjoyable game content (e.g. Loiacono, Cardamone & Lanzi, 2011; Togelius et al., 2010). When integrated with player modelling, PCG can take a step further, leading to the automatic generation of personalised game content and interactive narratives. Notable instances include affect-driven narrative systems in Façade (Mateas & Stern, 2003), FearNot! (Aylett et al., 2005), Storybricks (Namaste Entertainment, 2012), and affect-centred game narratives, as seen in Final Fantasy VII (Square Product, 1997).

The Experience-Driven Procedural Content Generation (EDPCG) framework, proposed by Yannakakis & Togelius (2011), views content as a fundamental component shaping player experience. This framework advocates for content adjustments to optimise the overall playing experience. Various forms of content, such as racing tracks (Togelius, De Nardi & Lucas, 2007; Loiacono, Cardamone & Lanzi, 2011), strategy maps (Togelius et al., 2010, August), game rule sets (Browne, 2008), buildings (Martin et al., 2010), weapons (Hastings, Guha & Stanley, 2009), and spaceships (Liapis, Yannakakis & Togelius, 2011; Liapis, Yannakakis & Togelius, 2012), have been generated using models that consider player experience.

Moreover, the concept of Adaptive Content Generation tailors content based on user preferences, exemplified in games like Galactic Arms Race (Hastings, Guha & Stanley, 2009). In this game, weapons evolve based on the player's previous use and preferences, aligning content creation with individual player likes and experiences.

Despite its advantages, PCG poses challenges. Achieving a balance between randomness and structured design requires careful consideration, and developers must ensure that generated content aligns with the overall vision of the game. Fine-tuning algorithms to avoid undesirable outcomes or ensure diversity in generated content can be complex tasks (Shaker, Togelius & Nelson, 2016)

7. Interactive Storytelling and NLP

Generating stories automatically or semi-automatically falls within the domain of Narrative Procedural Content Generation (Narrative PCG or NPCG), a subset of the broader field of Procedural Content Generation (PCG). Unlike general PCG, which encompasses the automatic creation of various game elements such as levels, environments, or characters, Narrative PCG specifically focuses on the generation of narrative elements, including plotlines, character arcs, and dialogue. While Narrative PCG can overlap with interactive storytelling by incorporating user choices and adaptive storylines, the two can also exist independently. Narrative PCG primarily aims to create coherent and engaging stories, which may or may not involve interactivity.

Maintaining control over the narrative is a critical consideration for writers in Narrative PCG, particularly when AI or Natural Language Processing (NLP) techniques are employed to generate story content. Traditional narrative design methods, such as predefined story arcs or plot points, are often used to ensure the story remains cohesive and compelling. However, integrating player modelling can enhance personalisation within these narratives. Player modelling involves analysing and understanding the player's preferences, behaviours, and personality traits to tailor the narrative experience more closely to the individual.

Narrative PCG is not limited to entertainment but is also applicable in educational settings and personality testing, where personalised storytelling can enhance engagement and learning outcomes. The use of gamification techniques, such as incorporating game-like elements (rewards, challenges), can further increase user interest and participation, making the experience both educational and enjoyable.

7.1. Narrative Design and Player Modelling

Two types of approaches are common in Narrative PCG: plot-based and character-based. A plot-based approach manipulates the narrative structure of the story and has been used in systems such as Façade (Mateas & Stern, 2003), Mimesis (Young et al., 2004), and IDA (Magerko, 2005). The Russian formalist Vladimir Propp’s (1928) idea of narrative functions such as absence, interdiction and transgression being the basic units of folktales, with a fixed chronological order, has been considered in many approaches, such as Grasbon and Braun (2001). Character-based generation, on the other hand, emphasises the development and interaction of characters within the story. In this approach, narratives evolve based on the actions and relationships of the characters, as seen in projects like FearNot! (Aylett et al., 2005). While this can offer more dynamic and personalised experiences, it also poses challenges in maintaining plot coherence and ensuring that the narrative remains engaging and meaningful.

The integration of player modelling in Narrative PCG serves as a bridge between these approaches and the project's goals. By leveraging data on player preferences and behaviours, narrative elements can be tailored to resonate more deeply with individual users, enhancing the personalisation of the storytelling experience. This application of player modelling is particularly relevant to the thesis, as it investigates how psychological models can inform and enhance the generation of personalised narratives.

Indeed, an interactive narrative can be constructed in various ways. A common technique not involving artificial intelligence is to build a branching story graph (e.g. Riedl & Young, 2006) with alternative actions the user can choose. Another option is a Drama Manager (DM), an omniscient background agent determining what will happen next. Typically, a human game designer creates targets that the DM should aim for to create a good experience.

Some examples of interactive narratives selecting story paths include Peinado and Gervas (2004), who introduced a case-based reasoning system mimicking pen-and-paper role-playing games. It dynamically selects story events to suit the preferences of a player type. Mateas and Stern’s Façade (2003) has an event selection mechanism founded on the dramatic writing concept of dramatic beats, the smallest unit of dramatic action. The beats are selected based on natural language input from the user, with the guidance of a drama manager which directs the story along an Aristotelian tension arc. Riedl and Stern (2006) took use of the same technology to develop the Automated Story Director, which can choose events according to a partially ordered plan not just for dramatic purposes, but also for educational ones.

The realm of quest generation and interactive storytelling within video games has witnessed significant exploration, with various frameworks and algorithms contributing to the enhancement of dynamic narrative experiences. Sullivan, Mateas and Wardrip-Fruin (2010) introduced a comprehensive framework where a game manager dynamically adjusts the structure of quests based on the player's history and the current state of the game world. Their rule-based system operates on a library of quests, allowing for the dynamic recombination of quest elements. Another notable framework by Breault, Ouellet and Davies (2021) relies on automated planning, using a deterministic planning algorithm to generate quests based on a world description represented as a set of facts. De Lima, Feijó and Furtado (2014) propose a dynamic solution based on hierarchical task decomposition and planning under non-determinism, addressing the challenges of handling non-deterministic events and supporting quests with multiple endings, thereby influencing the game's narrative.

While genetic algorithms have not been extensively explored for quest generation, related works in general narrative generation provide valuable insights. McIntyre and Lapata (2010) describe a story generator system employing an evolutionary search strategy. Their algorithm operates directly on text sentences, addressing syntax and semantics through genetic operations like mutation. The fitness function in their approach evaluates coherence as a key criterion. Previous research, such as that by Ong and Leggett (2004) and Giannatos et al. (2011), has explored the use of story templates and graphs in conjunction with genetic algorithms. Ong and Leggett's system recombines story components using a genetic algorithm, with story fitness determined by pre-rated events. Giannatos et al. (2011) employ an evolutionary algorithm to suggest new story events for an existing story graph, rating the stories based on spatial locality, thought flow, and motivation. The final fitness is derived from the average of these ratings.

Nairat, Dahlstedt and Nordahl (2011; 2013) present a unique approach integrating evolutionary methods into a character-based system. Their method utilises an interactive genetic algorithm to create characters, focusing on internal states and action rules defining personality and behaviour. The integration of genetic and planning algorithms is further explored by Giannatos et al. (2012), where genetic algorithms generate plan operators as narrative units for constructing new story plots. However, the approach faces challenges, such as the lack of generated meaning for operators.

Player modelling has been widely used to adapt computer games, but relatively little in determining storylines. None of the above approaches involve player modelling, unlike PaSSAGE (PlayerSpecific Stories via Automatically Generated Events) (Thue et al., 2007), an interactive storytelling system which bases its storytelling decision on an automatically learned vectorial model of each player's style of play. A user study with the system found it better than two fixed, pre-authored stories for certain types of players, and particularly for females in terms of fun and agency.

Vectorial models of user types have been used in some interactive narratives. Barber and Kudenko's (2007) system records the personality of its users by their decisions leading to predefined increments or decrements to a vector of personality traits. Seif El-Nasr's (2007) system Mirage also creates a vector from character traits (heroism, violence, self-interestedness, and cowardice) for the sake of making more engaging drama. Sharma et al.'s (2010) system uses past captured game traces and player survey data to create player models, used to dynamically determine the best next plot point to each user.

Player modelling could also be treated as a content recommendation problem (Medler 2009) or a collaborative filtering problem, as in Yu & Riedl (2013), who introduce a solution to the sequential recommendation problem called Prefix-Based Collaborative Filtering (PBCF), which learns a user's preferences from ratings on story fragments and then chooses successive plot points.

There is also work on automatic generation of distinct linguistic pragmatics for narration. The Personage system (Mairesse & Walker, 2010) maps the Five-Factor Model to a wide range of linguistic parameters. Rishes et al. (2013) took use of Personage for creating variations of stories generated from a semantic representation composed of events and character intentions, with the results being the same regarding content, but slightly different stylistically, such as in the use of swear words, exclamation marks and stuttering. The style of the language can be personalised to the user, as in Ritschel, Baur and André (2017) who adapted a chatbot's linguistic style to the user's FFM personality type. Some studies had before Personage already shown the use of the FFM in affecting the conversational agents: for example, André et al. (2000) introduced a system which allowed modifying the agents' utterances by selecting different values for Extraversion, Agreeableness and Openness to Experience. Cassell and Bickmore (2003) noted that when they had their conversational agent do small talk, Extraverted users felt that they know her better if she produced social language, resulting in a more satisfying interaction, but Introverted users rated that version of lower.

Narrative-based games such as Heavy Rain (Quantic Dream, 2010) can feature thousands of lines of dialogue manually authored by several writers. Data-driven approaches can also be used, such as through crowdsourcing. Data on players interacting with each other or with non-playable characters (NPCs) can be used to train the NPCs to respond in similar manners using n-grams (contiguous sequences in language), as in Orkin and Roy (2007). Designers could also collaborate with a computer by taking turns on adding sentences in a narrative, and the computer could then provide meaningful sentences by matching the current narrative with similar ones on the cloud (Swanson & Gordon, 2012). Relying more on natural language generation (NLG) is also increasingly becoming an option: for example, the early versions of AI Dungeon 2 (Walton, 2019) used OpenAI's GPT-2 language model, trained using texts in the style of Choose Your Own Adventure books, to create an interactive textual narrative where the user selects a character and the AI generates a few lines of narrative before letting the user write in what happens next, which then might or might not happen, as the AI generates more narrative. The narratives showed some coherence, but managing to create one worth reading for dramatic pleasure seemed like an unlikely prospect, making the appeal of the experience the curiosity in what the AI can produce. However, as the GPT models have improved, the game has been updated, and the opportunities are ever increasing.

7.2. Text Style Transfer

A central aspect of narrative personalisation in this thesis is changing the style of language to suit the reader's preferences. This brings to focus the concept of text style transfer, a transformative process aimed at altering the style of a given text while preserving its original meaning. The objectives of style transfer encompass a variety of dimensions, including changing sentiment, formality, genre, or even the political slant and gender of the writer. Despite sentiment, gender, and political transfer being common tasks within text style transfer, it is pertinent to note that these alterations may not strictly constitute changes in style but rather in the underlying meaning of a sentence. For instance, sentiment transfer might transform a positive statement like "I love this restaurant" into its negative counterpart, "I hate this restaurant," a task that might lack meaningful utility.

In the context of our exploration, authentic style transfer should focus on transitioning from the writing style of one author to another. Ideally, this process would leverage parallel data, wherein identical texts exist in different versions, composed or translated into distinct styles.

However, the scarcity of such parallel data has prompted the development of various workarounds. Notably, methods such as creating pseudo-parallel data, as demonstrated in works like Jin et al. (2019) or Nikolov and Hahnloser (2018), have been employed. These approaches primarily utilise unsupervised methods, which means they don't require large amounts of labelled data for training. This allows for the exploration of style transfer even when there isn't an abundance of parallel data available for comparison. Essentially, unsupervised methods enable the model to learn patterns and relationships within the data without explicit guidance or supervision from labelled examples.

However, within the domain of text style transfer, there are also supervised methodologies. These methods often repurpose sequence-to-sequence models originally designed for machine translation, as exemplified by works such as Carlson, Riddell, and Rockmore (2018) and Wang et al. (2020, 2019). For instance, Jhamtani et al. (2017) employ a word mapping strategy to "translate" between Shakespearean and modern English. Additionally, semi-supervised approaches, incorporating pseudo-parallel data alongside actual parallel data, have been explored (Shang et al., 2019; Zhang, Ge & Sun, 2020; Liu, Wang & Okazaki, 2022).

Many unsupervised approaches in this domain can be characterised as disentanglement methods. Their primary goal is to separate style from content, manipulating latent representations to combine content with a target style. Autoencoding frameworks, which involve neural network architectures designed to learn efficient representations of input data, are frequently employed in such approaches, as evidenced by the work of Hu et al. (2017), Shen et al. (2017), Fu et al. (2018), and John et al. (2019).

Given the inherent difficulty in disentanglement, these methods may achieve satisfactory style accuracy at the expense of content preservation. This means that although the style may be accurately modified, the essential meaning or information within the text may be lost or distorted. To address this issue, researchers often turn to translation-based processes, which involve translating the text from one language to another and then back to the original language. This approach, known as back-translation, helps to ensure that the content of the text remains intact while modifying its style. Several studies have explored and applied back-translation techniques to improve the quality and fidelity of style transfer in natural language processing tasks (Logeswaran et al., 2018; Zhang, Ding & Soricut, 2018; Prabhumoye et al., 2018; Prabhumoye, Tsvetkov & Salakhutdinov, 2018; Lample et al., 2019). However, this approach may entail limited control over the target style during the generation process.

Paraphrase generation is to express the same information in different ways. Paraphrasing has a lot in common with text style transfer, making it a potential approach. Krishna et al. (2020) use unsupervised paraphrase generation which creates pseudo-parallel data by feeding sentences from different styles through a diverse paraphrase model. This normalises the input sentence by stripping away information that is predictive of its original style. This then enables training an inverse paraphrase model specific to the original style, which attempts to regenerate the original sentence. This process should enable the model to reproduce the original style without undue changes to the input semantics.

In the realm of style transfer, the absence of robust automatic evaluation metrics poses a considerable challenge. The commonly used metrics, including accuracy, fluency, and content preservation or similarity, often fall short in capturing the nuanced aspects of style transformation. Automatic metrics, requiring only system outputs and source texts, are frequently employed as rewards during training.

Accuracy evaluation often involves training a classifier to discern whether a generated sentence aligns with the target style. A prevalent technique is employing a single-layer convolutional neural network (Kim, 2014) for this purpose. Fluency, indicating the correctness and naturalness of a sentence, is typically measured by the perplexity of a language model. Despite its popularity, BLEU (bilingual evaluation understudy) (Papineni et al., 2002) is deemed problematic in the context of style transfer. Originally designed for language translation, BLEU proves less suited for evaluating style transfer models, primarily focused on altering style by necessitating changes in words. BLEU predominantly rewards n-grams with exact matches in the reference system, disregarding semantic meaning. Studies, such as that by Sulem, Abend, and Rappoport (2018), highlight the inadequacy of BLEU scores in reflecting grammaticality and meaning preservation. Additionally, style transfer inherently involves using different words than the source sentence, as observed by Chen and Dolan (2011).

Alternative methods, like the average Levenshtein distance, which correlates well with meaning preservation and grammaticality (Sulem, Abend & Rappoport, 2018), tend to favour small, precise edits. However, these may be more suitable for attribute transfer rather than the broader goals of style transfer. Chen and Dolan (2011) combine BLEU with their new PINC (Paraphrase In N-gram Changes) scoring, which rewards sentences for their divergence from the original – an approach contrary to BLEU's principles. Interestingly, researchers often aggregate different metrics for accuracy, fluency, and content preservation to obtain an overall score. This practice, however, presents a paradox, as these metrics tend to be inversely correlated. For instance, a model transforming text into Shakespearean style might yield a high score in accuracy and fluency by generating sentences like "Wherefore art thou Romeo?" regardless of the input. On the other hand, a model merely copying its input could consistently score well in terms of similarity and fluency (Pang, 2019).

In the study of text style transfer, particularly within the context of Narrative Procedural Content Generation, several key insights emerge that are crucial to understanding and applying this technology. First, the scope of style transfer extends beyond mere linguistic alterations; it involves intricate processes that balance style modification with the preservation of the original content's meaning. This balance is particularly challenging in the absence of parallel data, which has led researchers to explore alternative methods such as unsupervised learning and back-translation. These approaches enable stylistic changes while maintaining the integrity of the content, although they may offer limited control over the precise outcome.

Moreover, the methodologies employed in text style transfer, such as the distinction between plot-based and character-based narrative generation, provide valuable parallels to the challenges faced in Narrative PCG. These methodologies highlight the importance of maintaining a coherent narrative while allowing for personalisation and adaptability based on user input. This consideration is essential in creating engaging and immersive narrative experiences that resonate with individual users.

A significant challenge in this field is the evaluation of style transfer outcomes. The inadequacy of traditional metrics like BLEU, which are more suited to tasks like language translation, underscores the complexity of assessing stylistic transformations. The search for more appropriate evaluation tools reflects the need for metrics that can better capture the subtleties of style change, ensuring that both accuracy and content preservation are adequately measured. Indeed, it is indeed challenging to maintain the quality and coherence of text in text style transfer (Hovy & Spruit, 2016).

In relation to the broader study, these insights underscore the relevance of sophisticated style manipulation techniques to the personalisation of literary narratives. By integrating findings from text style transfer, the study aims to enhance Narrative PCG, allowing for more personalised and nuanced storytelling. This integration is informed by an understanding of user preferences and characteristics, which can significantly enrich the narrative experience. Thus, the exploration of text style transfer not only contributes to the technical advancements in Narrative PCG but also offers a framework for creating more individualised and engaging literary works.

7.3. Gamification and Personality Testing

Current personality tests, while widely used and valuable for understanding individual differences, face several challenges that limit their effectiveness. One significant issue is the repetitive and predictable nature of these assessments, which can lead to decreased participant engagement and attentiveness. This monotony can result in superficial or careless responding, where participants may not fully engage with the questions, potentially compromising the accuracy of the results. Additionally, traditional personality tests often fail to sustain participants' interest over time, which can be particularly problematic in long or repeated assessments.

Gamification offers a potential solution to these challenges by incorporating game elements into the testing process, thereby enhancing participant engagement and response quality. By framing questions within a narrative or game-like context, gamification can make the assessment process more enjoyable and immersive, which helps to maintain attention and reduce careless responding. This approach leverages elements such as storytelling, challenges, and rewards to create a more engaging and dynamic experience.

Gamification involves incorporating game elements into tasks or activities to enhance participant engagement and response (Landers et al., 2018). Gamification has gained prominence in various fields, such as marketing, employee training, and education. Incorporating game elements into tasks increases intrinsic interest and enjoyment. In fact, just presenting an activity as a game can profoundly influence engagement. Consequently, gamified survey-type tasks and assessments are gaining significance as valuable tools for motivating and framing activities related to consumer behaviour, organisational procedures, and participation in scientific research (Lieberoth, 2015). Gamification often monitors the users and motivates them with points, badges and leaderboards, but this does not tend to maintain interest in the long term. To avoid this, high levels of personalisation would be needed to maintain suitable motivation and provide a better end-user experience (Gadiyar, 2014).

Successful integration of game elements has been shown in various studies, but there's a need for a detailed account of how specific elements impact outcomes. For example, Landers and Callan (2012) added game elements to an educational task, improving motivation and learning outcomes, but they couldn't identify the specific elements responsible for the change.

In the context of personality assessment, gamification can enhance engagement and reduce careless responding. A common issue with personality tests is that they can become repetitive and fail to maintain participant attention. Framing questions within a narrative helps sustain attention and motivation, reducing careless responding. Narrative-driven personality measures also allow for stealth assessment, embedding assessment within a game-like context, maintaining the participant's flow state.

This means that interactive narratives could potentially be used as personality tests. This is an approach used by McCord, Harman and Purl (2019). They proposed that the participants will seek to complete the narrative effectively rather than focus on projecting their personality in a flattering way. In their user study, they found statistically significant correlations for some but not all FFM traits. In two narratives, in which scores represent willingness to choose a course of action associated with one trait over another, Openness and Neuroticism failed to get significance; in one, where scores represent willingness to choose a course of action associated with high levels of a trait over lower levels, it was Conscientiousness that was non-significant. Later, Harman and Purl (2022) confirmed that the results were more repeatable than with personality tests. In Harman and Brown (2022), they also added illustrations to the interactive narrative, but found that this appeared to make no difference.

8. Recommender Systems

8.1. Introduction to Recommender Systems

Recommender systems (RS), also known as recommendation engines or simply recommenders, represent sophisticated information filtering systems designed to predict and cater to a user's potential interest in items by providing tailored suggestions. These systems operate by gathering and analysing user-specific information, obtained either explicitly through user-provided ratings or implicitly through the interpretation of their actions, such as item clicks or views. This dual approach enables recommenders to develop a comprehensive understanding of user preferences, enhancing the accuracy and relevance of their suggestions.

At the heart of recommender systems lies the objective of personalising user experiences, ensuring that the content or items recommended align closely with individual tastes and preferences. The explicit collection of user ratings provides direct insights into their preferences, allowing recommenders to make informed predictions about their potential interest in similar items. Implicit data, on the other hand, focuses on user behaviour, extracting valuable information from their interactions with the platform to infer preferences and provide more detailed recommendations.

The pervasive influence of recommender systems extends to some of the world's most popular websites, where they play a crucial role in shaping user engagement. Platforms like Amazon, YouTube, and Netflix leverage recommender systems to curate personalised content suggestions. By facilitating the discovery of new items or content aligned with individual tastes, these systems significantly impact user satisfaction and the overall success of digital platforms.

Recommender systems distinguish themselves conceptually by presenting personalised sets of items, setting them apart from comparable processes like internet filtering. Unlike generic filtering, recommenders utilise user-specific justifications to generate individualised recommendations (Beliakov, Calvo & James, 2011: 706). Notably, recommendations are often proactively displayed, even in the absence of specific user searches, appearing on the homepage of a website or at the bottom of a product-related page.

Two primary types of recommenders exist: collaborative filtering (CF) and content-based filtering (CB). Collaborative filtering relies on recommending items liked by similar users, whereas content-based filtering suggests items similar to those the user has previously enjoyed. Knowledge-based systems, on the other hand, solicit explicit user input about preferences and offer recommendations accordingly; however, whether they fall under the recommender system category can vary. Typically, the term RS specifically denotes personalised recommendations, excluding systems that recommend popular products or solely rely on location-based suggestions.

Within the realm of recommender systems, diverse approaches include demographic methods (e.g. Wang, Chan & Ngai, 2012), utility-based methods (e.g. Huang, 2011), and community-based methods (e.g. Grasso & Bergholz, 2007), though these are less frequently discussed. Recent research has seen a shift towards hybrid systems that combine multiple recommendation approaches, addressing challenges like data sparsity and knowledge acquisition. For instance, collaborative filtering struggles with "cold start" issues, where it cannot recommend items that have no ratings, or make recommendations for new users. However, content-based filtering mitigates this limitation by recommending similar items based on their features, not solely on ratings. Noteworthy examples of hybrid systems include Netflix, which blends collaborative and content-based filtering techniques to enhance recommendation accuracy and relevance.

Recommenders are typically described as presenting items to users for one reason: so that the user would see items they are interested in. However, there is also another reason: to learn more about their preferences. Augmenting recommenders with active learning (AL) personalises the process, as it provides new information to the system, which can then use it for new recommendations. This can also help the user become more aware of their preferences. A common way to get the process started with new users is to ask them to rate some items, known as training points, which are then used for creating a model approximating the user's preferences. The items they then explore can also help them and the system understand what they like (Rubens, Kaplan & Sugiyama, 2011: 735-6).

In an article considered seminal in the field, Herlocker et al. (2004: 9-11) list ten popular tasks for recommenders to help with:
• annotation in context, highlighting the items the user might like out of the ones shown on a TV schedule, for example
• find good items, ranked and with predicted ratings on how much the user would like them, for example on a five-star scale
• find all good items, such that would satisfy some user needs
• recommend a sequence, such as a set of musical tracks
• just browsing, displaying items more likely to interest
• find credible recommender, letting the user test the recommender to see how well it works
• improve the profile, letting the user give information on their preferences
• express self, letting the user state opinions for purposes other than recommendations
• help others, letting the user inform others of their experiences
• influence others, letting the users promote or warn about items, though some of this behaviour might be considered malicious.

Recommendations can indeed be made for a single item, a simple list of items, or a sequence. Recommenders for single items or simple lists do not consider how the user's choice of an item could influence choosing the next items; there has not yet been much work on sequences, however, especially such that would consider more than the items just visited (for an overview, see Quadrana, Cremonesi & Jannach, 2018).

When designing a recommender system, two perspectives on its purposes have to be considered: the application point of view and the user point of view. From the application side, as Picault et al. (2011: 336-7) note, the purposes include:

• being a major service provided by the application, in charge of what content is consumed, for example in services for music and films
• making use of the “long tail” (Anderson, 2007), the plurality of items that are less known
• increasing user loyalty, for example by involvement in the recommendation process
• increasing revenues by promoting profitable products
• increasing system efficiency – easing the search process reduces the cost of running the system.

Users are normally given the chance to browse content in addition to being given recommendations. How these different navigation methods are integrated can greatly influence the user experience. This could be done by letting the user request recommendations completely separately from content browsing, perhaps on the home page or home screen, or giving recommendations only in the interaction context, such as when viewing an item. If the only way to navigate an application is to use a recommender, bad recommendations would be fatal for user satisfaction, but other options would reduce the impact.

Arguably the biggest issue recommenders face is rating sparsity, which occurs when there is not enough data on what users like. When there is no information at all on a user or an item, this is called a cold start. To tackle rating sparsity, there have been content-based recommender techniques using content representations of items to locate items with similar content to items the target user liked (Lops et al., 2011; Pazzani & Billsus, 2007). Other studies have relied on other types of user-generated information, such as tags (keywords written by users) (Marinho et al., 2011; Zhao et al., 2008), social relationships in social media (Kaya & Alpaslan, 2010; Beilin & Yi, 2013; Chen, Zeng & Yuan, 2013; Yang et al., 2012), or demographic information (Pazzani, 1999). Jeong et al. (2013) use the selection of optimal personal propensity variables, instead of all the available ones. Personal propensities can be defined as something typical of an individual that can be used characterise them. However, none of these methods can solve the problem of high data sparsity levels, particularly the problem of there being little data on the user (Chen, Chen & Wang, 2015: 100).

Recommender systems generally outperform human recommenders, but people generally prefer human recommendations, and in subjective domains in particular, people seem more reluctant to take predictions from a machine (Logg, 2016). Recommender interface elements affecting trust, or the credibility of the system, include explanations, product comparisons and automated repair functionalities (Felfernig & Gula, 2006), and of course the perceived overall quality of recommendations (Herlocker et al., 2004). Several studies have found that increased control over the process increases trust (West et al., 1999, Komiak, Wang & Benbasat, 2005, Wang, 2005) as well as general satisfaction (Pereira, 2000; McNee et al., 2003) with the system. Recommending a few items the user already likes might also be useful initially, letting the user know that the system has some understanding of what the user likes. Trust can be measured by asking users about it in a user study; alternatively, an online test could study how many recommendations were used, or how many users returned, though other factors might be hard to separate. Offline experiments would not work, as trust is built through an interaction between the system and a user (Shani & Gunawardana, 2011: 285).

One way of increasing trust is to use social or community-based recommenders that recommend items based on the preferences of the user's friends. People would appear to rely more on recommendations from their friends than from similar but anonymous individuals (Sinha & Swearingen, 2001). Merging recommendation systems and social networks has gained popularity (Tavakolifard & Almeroth, 2012). However, research results about recommendations based on social networks are mixed: some have found social network data more useful than preference similarity data (Guy et al., 2009), or that adding social network data to ordinary CF improves recommendations (Groh & Ehmig, 2007); others (Golbeck, 2006; Massa & Avesani, 2004) have found them better than other CF approaches only in specific situations such as controversial items with varied ratings, or cold starts.

Using social networks also risks introducing what is called a filter bubble, as studied in Nguyen et al. (2014) and coined by the internet activist Eli Pariser (2011): algorithms tailoring information to people create a personal ecosystem of information based on user information, separating from diversity and confirming what people already prefer without exposure to anything unfamiliar to the user. Indeed, personalisation is typically based on similarities, contributing to this problem. As Ziegler et al. (2005) demonstrate, diversification could improve user satisfaction, and as Swearingen and Sinha (2001) found observing participants using several commercial recommendation systems, very novel and unexpected recommendations were well received. One way to avoid monotonous recommendations is to "inject a note of randomness" (Shardanand & Maes, 1995). This could be done by using genetic algorithms, which evolve with iteration; this also helps with adapting to changing user interests (Sheth & Maes, 1993). Lathia et al. (2010) recommend switching the algorithm over time to re-rank the results of frequent visitors, so that the system would be temporally evolving. Since diversity may take away from other properties such as accuracy, curves can be computed to evaluate the decrease in accuracy vs. the increase in diversity (Shani & Gunawardana, 2011: 288).

8.2. Assessing Recommender Systems

Once the goals of a recommender are set, targets for the performance of the system can be set along a number of criteria. According to Picault et al. (2011: 338), some key performance criteria could include:

- correctness metrics, such as accuracy, precision and recall, are commonly used but not sufficient to evaluate user satisfaction (McNee, Riedl & Konstan, 2006)
- transparency and explainability: the user understanding how the recommendations have been made can increase trust, but is difficult with collaborative filtering
- serendipity, or positive surprises, which are also hard to achieve with collaborative filtering
- risk-taking, like serendipity, is about recommending items that have lower odds of being liked, but might be very much liked, or hard to find
- response speed / performance: often the speed of the application is more important than the accuracy of the results, and sometimes it might be better to precompute the recommendations
- reliability, more important in important decisions
- robustness to attacks, often crucial when recommending products from different providers.

Traditionally, when evaluating recommender systems, researchers have relied on offline experiments as a common methodology. Offline experiments involve using existing data to estimate the prediction error of recommendations generated by the recommender system. In these experiments, the recommender system is tested using historical data, such as user-item interactions or ratings, to simulate how it would perform in real-world scenarios. The system generates recommendations based on the historical data, and the accuracy of these recommendations is then evaluated against a set of predefined metrics or criteria. Offline experiments provide a controlled environment for evaluating recommender systems, allowing researchers to assess their performance without directly impacting users. However, they also have limitations, such as the inability to capture real-time user feedback or interactions, which may affect the system's performance in practical settings. The accuracy of predictions is typically assessed using well-established information retrieval (IR) metrics, including Mean Absolute Error (MAE), precision (a measure of errors made in classifying samples into a particular class), and Normalised Discounted Cumulative Gain (NDCG), a logarithmically discounted measure considering the positions of recommendations (Ricci, Rokach & Shapira, 2011: 18).

Mean Absolute Error (MAE) provides a straightforward measure of the average absolute difference between predicted and actual ratings, offering insights into the overall accuracy of the recommendation system. Precision, on the other hand, assesses the correctness of the recommendations by measuring the ratio of correctly predicted items to the total number of recommendations made. This metric is particularly valuable in understanding the system's ability to avoid making incorrect suggestions. Normalised Discounted Cumulative Gain (NDCG) is a more nuanced metric that considers the position of items in the recommendation list. It applies logarithmic discounts to the gain of items based on their positions, giving higher importance to items placed at the top of the list. This reflects the understanding that users are more likely to interact with items presented early in the recommendation list.

These metrics collectively provide a comprehensive evaluation framework for recommender systems, enabling researchers and practitioners to gauge the accuracy, precision, and overall effectiveness of the recommendations generated by these systems. However, it's important to note that offline experiments have inherent limitations, and their results may not fully capture the real-world dynamics and user interactions that occur in online, dynamic environments. Hernández Del Olmo and Gaudioso (2008) criticised the existing accuracy and ranking metrics as overparticular, suggesting a new category of metrics for measuring a recommender's capability to make successful decisions, focusing on not just which items to recommend, but also when and how to do this, also taking into account the interactivity of a recommender system, which had not been evaluated before.

Some may point out limitations in such methods; others find that the quality of recommenders can never be measured since there are too many objective functions

(Jannach et al., 2011: 166). Recommender systems can have goals beyond the accuracy of the algorithms: they are there to create an enjoyable, personalised experience that helps with user retention, and of course with sales and profitability. The system also has to create interest in the recommended items and trust in the recommender, have a system logic with at least some transparency, point users towards new items, provide details about recommended items, and present ways to refine recommendations (Swearingen & Sinha, 2001), and such factors should also be sought to be measured. Even if the algorithm is good at predicting ratings, it might be disliked by the users for many reasons such as slowness, necessitating a user-centric evaluation, which can be done online after the system has been launched, or as a focused user study. In online evaluation, real users use the system, not knowing details of the experiment. Several different versions of the systems can be tried on different users. If an online evaluation is unfeasible or too risky, a focused user study is done, with a few users asked to do different tasks with different versions of the system. Both quantitative and qualitative information about the systems can then be collected (Ricci, Rokach & Shapira, 2011: 16).

User evaluations of recommender systems are noticeably influenced by the design of the system input, encompassing how preferences are elicited and the degree of control afforded to users in the recommendation process. In their study, Xiao and Benbasat (2007) discovered that the method employed for preference elicitation, whether implicit or explicit, significantly shapes user perceptions of the system. Implicit elicitation methods were associated with greater perceived ease of use and satisfaction, while explicit methods were deemed more transparent and conducive to better decision quality.

The value users attribute to the preference elicitation process is further nuanced by factors outlined by Gretzel and Fesenmaier (2007). Topic relevance, transparency in the elicitation process, and the level of effort required for providing input were identified as crucial aspects enhancing user perceptions. The relevance of the questions posed during the process not only indicates a consideration for user interests but also contributes to user satisfaction and facilitates constructive feedback.

The concept of locus of control (LOC), introduced by Duttweiler (1984), plays a pivotal role in user satisfaction with recommender systems. Users with an external locus of control, exhibiting less interest in controlling the recommendation process, may find predefined and static dialogues more suitable. On the other hand, individuals with an internal locus of control, desiring more control, tend to prefer flexible dialogues. Flexible systems actively propose varied parameters and feature settings, catering to users who wish to exert greater influence over the recommendation process (Mahmood & Ricci, 2007; Tiihonen & Felfernig, 2008). This adaptability in dialogue style aligns with the diverse preferences and expectations users bring to the recommender system interaction.

8.3. Collaborative Filtering

In collaborative filtering (CF), sometimes referred to as people-to-people correlation, the only input is a matrix of ratings given to items by users. Typically, they produce as an output a prediction about how much the user will like an item, and a list of recommended items. In collaborative filtering, it is expected that people sharing some similar interests would be likely to have other similar preferences as well, an approach that could easily be questioned, but appears to work: Collaborative filtering is generally considered to make more successful predictions than the other common methods (e.g. Lury & Day, 2019: 22).

There are two types of collaborative filtering recommenders: user-based and item-based (item-to-item). A user-based recommender searches similar users and makes a prediction based on their similar preferences (e.g. Adomavicius & Tuzhilin, 2005). The idea is to recommend items liked by similar users, calculated from the matrix of user ratings for items. A common way of doing this is a user-based nearest neighbour recommendation, which, given a ratings database and a user ID, identifies other users with similar preferences, often using the common statistical measure of Pearson correlation. Then, a prediction for every product the user has not seen is made based on the ratings given by such peer users (Jannach et al., 2011: 13-4).

In item-to-item recommendation, the items are compared first, but incorporating user preferences, making the idea to recommend more items liked by people who liked an item presently displayed. The item-to-item approach is based on ratings for the item and ignores user and item attributes, making it necessary to have enough ratings for the item (Schein et al., 2002). Large ecommerce sites must handle millions of users and items, making it impossible to compute predictions on a vast number of potential neighbours in real time. Therefore, they typically use item-based recommendation, as it is more apt for offline preprocessing, allowing computing recommendations in real time (Sarwar et al., 2001). This system has also been used by Amazon (Schafer, Konstan & Riedl, 2001; Linden, Smith & York, 2003; Koenigstein & Koren, 2013). It is also possible to use both user similarity and item similarity in ratings databases; Wang De Vries and Reinders (2006) take use of ratings for similar items made by similar users.

The ratings do not have to be explicit: both user-based and item-based collaborative filtering methods using inferred ratings have been found comparable to those using real ratings; inferred ratings are better in user-based collaborative filtering than in item-based collaborative filtering (Zhang et al., 2013). Opinions on specific aspects appear to be more useful than overall opinions in CF (Ganu, Kakodkar & Marian, 2013), and models that take review elements into account along with ratings have been found more accurate than the standard models that do not (e.g. Wang, Liu & Yu, 2012).

Collaborative filtering approaches can also be grouped into two general classes: memory-based (Resnick et al., 1994) and model-based (Breese, Heckerman & Kadie, 1998). Memory-based systems use the entire user database (collection of rated items) to make predictions; model-based ones use the database for estimating or learning a model, which is then used for predictions. They aim to learn the latent factors representing users' inherent preferences over an item's multiple dimensions (Koren, Bell & Volinsky, 2009). Memory-based approaches have the ratings database in memory, used directly for creating new recommendations. Model-based techniques, however, process the raw data offline, which is more efficient for large databases, but possibly less precise. User-based systems are typically memory-based (Jannach et al., 2011: 26).

Collaborative filtering algorithms can recommend any type of items, regardless of their content, which is one reason why they are widely used, with good success. Their main challenges would be improving their scalability and efficiency, as well as simply improving the quality of the recommendations (Sarwar et al., 2000, Sarwar et al., 2002). Indeed, collaborative filtering systems perform well when there is sufficient rating information (Su & Khoshgoftaar, 2009), but not when there is rating sparsity because of a poor coverage of recommendation space (Garcia Esparza et al., 2010) or a difficulty with letting users show their preferences as scalar ratings on items (Leung, Chan & Chung, 2006). Therefore, collaborative filtering might not be an appropriate choice for services with only a small number of users. New users or items without any ratings can be a particular problem, known as the cold-start problem. Furthermore, sometimes a user can have unusual tastes, leading to no good recommendations. Such users are known as grey sheep, or in extreme cases where no similar users at all can be found, black sheep. Sometimes popular items might end up being recommended to everyone. There is little space for novelty, and less popular items that might be relevant will be overlooked (Zhou et al., 2010).

Sometimes, a specific case of the cold-start problem might occur when a collaborative filtering recommender does not have a user's preferences in one domain, but only in other domains. For example, the user should be getting recommendations for films, but the system only has their music preferences. One method for overcoming this issue is Cross-Domain Collaborative Filtering (CDCF) (Berkovsky, Kuflik & Ricci, 2007). In CDCF, the system searches users whose preferences are similar to those of the target user in the source domains, the ones where information is available, then filtering and recommending items from the target domain preferred by those users. However, in their work, Berkovsky, Kuflik and Ricci have not actually used properly different domains, only films, splitting it into domains based on genre, for example making recommendations on comedy films based on users who liked similar action films as the target user. In their results, the quasi-cross-domain-based recommendations had results even better than the regular collaborative filtering ones. Relatedly, Winoto and Tang (2008) mapped between domains in a user study of 144 university students. They found high correlation between songs and films related to each other, and that users who liked books of a certain genre also enjoyed TV series from the same genre. Contrary to Berkovsky, Kuflik and Ricci, they found that having items from the target domain was indeed better than having them from other domains.

Collaborative filtering typically assumes user preferences to remain the same – if a user viewed or bought something, or gave it a good rating, they are also expected to have similar interests in the future (Jannach et al., 2011: 23). Nevertheless, interests change: having bought a lot of furniture at one time does not necessarily mean that you will remain interested in furniture! However, this is even more of a problem with content-based filtering, described below.

8.4. Content-Based Filtering

In collaborative filtering, nothing but the user ratings needs to be known. This leaves out the option of making recommendations based on the characteristics of the items the user has liked. This would be called content-based filtering (CB). It relies on comparing content of items rather than other users' opinions, seeking to show the user more of what they have liked, such as more comedy films featuring the same actors. For this to be possible, the system has to be able to extract information on the items through keywords, with their similarity often computed using term weighting such as TF-IDF (term frequency–inverse document frequency, a numerical statistic on how important a word is to a document) (Salton & Buckley, 1988).

Content-based filtering does not need many users for good recommendations accuracy, and new items do not need to wait for ratings but can be immediately recommended once their attributes are available. There is no need for data on other users, avoiding the cold-start and sparsity problems. Users with unique tastes can get recommendations, as can new and unpopular items, avoiding the so-called first-rater problem. Explanations on what attributes caused the items to be recommended can also be helpful for the user. On the other hand, new users can be an issue, as in collaborative filtering: enough items have to be rated before the system can have a grasp of what the users like.

Shallow content analysis is an issue in content-based filtering – looking at the text content may not be enough to quantify how interesting something like a web page might be, but other factors such as usability or aesthetics also make a difference (Balabanović & Shoham, 1997). As Shardanand and Maes (1995) point out, when document characterisation is based on keywords, recommenders cannot tell the difference between well and poorly written articles. Additionally, the recommended documents might just not be long enough for proper distinctions (Jannach et al., 2011: 75). Therefore, the recommended items could be of simply low quality, given that their purpose or topic is all that matters.

Furthermore, the item attributes are not always easy to get. Technical descriptions of the characteristics, such as genre with films and books, are often provided with the item, but qualitative, subjective characteristics are more of a challenge (Jannach et al., 2011: 51). If it is not possible to automatically extract descriptive characteristics, one option would be manual annotation. This would in many circumstances be too costly, but luckily web users tend to do this voluntarily, tagging content they or others have provided. Nevertheless, content-based filtering is still not a social approach, and the quality judgements of other users cannot be exploited, unless somehow included in content features. Overspecialisation can also easily become a problem with a system just recommending more of the same to the user.

8.5. Knowledge-Based Systems

Knowledge-based systems (KB) seek to show the user items that match with the needs that the user has expressed. They match the user requirements with a knowledge base about the domain in question, and recommend items they find the most appropriate for the user considering their preferences. Such systems do not just filter items for the user, but are highly interactive, thus being different from collaborative and content-based filtering that do not require the user to express preferences. Burke (2000) describes them as conversational systems. However, the distinctions aren't always clear, and sometimes content-based recommenders might resemble knowledge-based recommenders, or indeed any other recommenders. One difference is that collaborative filtering typically extracts item information automatically, but knowledge-based systems tend to rely on externally provided information (Jannach et al., 2011: 78). Knowledge-based systems are typically the best choice for more complex purchases: recommendations on rare purchases such as houses would have little data to use, and the data could be outdated; user preferences could have changed, and so could the world: five-star recommendations for a ten-year-old computer could be highly misleading.

Knowledge-based recommenders come in two basic classes: constraint-based and case-based. In both, the user must first specify their requirements, and the system then attempts to find something suitable. Ideally, fixes for inconsistent requirements are automatically proposed if no solutions are available, and recommendation results are explained. They differ in how case-based systems focus on retrieving items matching with the user's needs as well as possible, using different types of similarity measures, but constraint-based systems rely on predefined knowledge bases containing explicit rules on how to relate customer requirements with item features (Jannach et al., 2011: 82). Constraint-based methods are particularly suitable for recommending complex products such as financial services or electronic consumer goods. Knowledge-based systems generally tend to work better than others at the beginning of their deployment, but if not equipped with learning components, they may fall behind other shallow methods able to use the logs of the human-computer interaction, as in collaborative filtering (Ricci, Rokach & Shapira, 2011: 13).

8.6. Context, Emotions and Personality

8.6.1. Context

Recommenders can be improved with contextual information, though it can also add to the complexity. In the literature on context-aware systems, context used to be defined as the location of the user, the identity of the people nearby, the objects around, and the changes in these elements (Schilit & Theimer, 1994), but other factors have since been added to this definition. Brown, Bovey and Chen (1997) include the date, the season, and the temperature; Ryan, Pascoe and Morse (1997) the physical and conceptual statuses of interest for a user; Dey, Abowd and Salber (2001) the user's emotional status, further redefining context as any information that can characterise and is relevant to the interaction between a user and an application (Adomavicius & Tuzhilin, 2011: 221).

There is also the question of the social environment: is the system being used alone or with other people? Group recommendations can be made by, for example, merging individual preferences to create a group profile for the content retrieval process (Ardissono et al., 2003; Bolger, Davis & Rafaeli 2003), applying a consensus mechanism to co-operatively define a shared content retrieval policy (McCarthy et al., 2006) or other methods described by Cantador (2008). Nevertheless, even in a group, individual recommendations may be needed (Bernhaupt et al., 2008; for this example, some solutions have been proposed in Bonnefoy et al., 2007 and Lhuillier et al., 2006).

8.6.2. Emotions

A few recommenders have used categorising emotions. Plutchik (e.g. Plutchik & Conte, 1997) developed an emotion categorisation applied by the now-defunct movie recommendation environment MovieProfiler. The search engine could search items based on an emotional profile specified by the user applying a case-based approach, retrieving the most similar items. A user selects on a five-point scale which emotions should be evoked by the film. The users can then evaluate the films regarding fear, anger, sorrow, joy, disgust, acceptance, anticipation and surprise (Jannach et al., 2011: 249).

Another instance is the Emotion Sensitive News Agent (ESNA) (Al Masum Shaikh, Prendinger & Ishizuka, 2010), which categorises news stories from different RSS sources into eight emotion categories. The text is assigned a positive or negative sentiment with numerical values, and the recommender also considers the cognitive and appraisal structure of emotions in taking into account the user's preferences.

Koelstra et al. (2012) proposed a music video recommender that translates a user's bodily responses to emotions, helping to understand the user's taste and then to recommend a music video matching the user's current emotion, implicitly tagging the videos using affective information. Nevertheless, measuring emotions is fraught with issues (e.g. Dasborough et al., 2008), and indeed Koelstra et al.'s method for creating their emotion database and giving the relevant recommendations, using EEG, peripheral physiological signals and face video, in conjunction with self-assessment, would be rather impractical for a commercial recommender.

8.6.3. Personality

There are a few cases where ad hoc approaches to personality have been explored, such as in a restaurant recommender by Gonzalez, Lopez and Rosa (2002), which uses a quiz to capture user characteristics, such as patience for waiting to be served or curiosity for exotic food. However, in recent years, recommender systems based on personality frameworks have become a more common theme for research. Usually, either the FFM or the MBTI are used, but other models have also been occasionally attempted, such as in Quijano-Sanchez et al. (2011), who use Thomas-Kilmann Conflict Instrument personality model. Dhelim et al. (2022) compared different frameworks, namely the FFM, the MBTI, Eysenck and HEXACO. In the cold start phase, Eysenck and the MBTI performed the best, while later the FFM and HEXACO were more accurate. In fact, a common reason cited for using personality-based recommender systems is their performance in cold starts, such as in Tkalčič et al. (2011), who suggested using personality information in neighbourhood measurement to help with cold starts. MBTI-based collaborative filtering has been found specifically useful for working with sparse data (Yi, Lee & Jung, 2015).

Hu and Pu found that in their studies, using the FFM increased user loyalty and decreased cognitive effort (Hu & Pu, 2009), and outperformed the traditional rating-based collaborative filtering method (Hu & Pu, 2011). In the latter study, they proposed three recommendation approaches: (1) using only the user's FFM personality; (2) using a linear combination of both personality and rating information; (3) using a cascade mechanism to leverage both resources. Similarly, Orestis & Christos (2017) proposed a film recommender system that was 50% based on the user's FFM personality traits and 50% on ratings. Quijano-Sanchez et al. (2011) added social trust with other users into the mix; Balakrishnan et al. (2018) used the FFM and demographic information. Wu et al. (2018) uses the FFM to estimate the user's diversity preferences. Alharti (2015) uses a collaborative filtering recommender that creates a personality profile for books and films, then making recommendations based on similarity. Roshchina, Cardiff, and Rosso (2011) utilised the Mairesse tool (Mairesse et al., 2007) to identify individuals with similar FFM personalities to a user and presented hotel reviews authored by those users. This approach taps into existing reviews to connect users with similar personality traits, providing a basis for recommendations aligned with their preferences.

Various links between specific Five-Factor Model traits and recommender preferences have been found. Wu, Chen and He (2013) got positive reactions from users by adjusting recommendation diversity to the user's personality, using some surprising correlations from Chen, Wu and He (2013), such as that more Neurotic users like diverse directors, and low-Agreeableness users prefer diversity with the country of the film. Ferwerda et al. (2016) found that Conscientiousness is linked to a preference for high diversification, and Agreeableness to medium diversification. They do not discuss the other FFM traits. Karumur, Nguyen and Konstan (2018) found that Big Five traits correlate significantly with newcomer retention, intensity of engagement, activity types, item categories, consumption versus contribution, and rating patterns, but not with recommendation diversity. They suggest that since people with high Openness tend to give higher ratings and few half-star ratings, the system could give them suggestions motivating them to give their true ratings, or the algorithms should take into consideration that there are such systematic rating biases. This would be related to how personality also influences what persuasive strategies are effective in recommender systems (Sofia et al., 2016).

In music, a personality-based approach was more accurate than a standard ratings-based system (Tkalčič & Chen, 2015). Onori, Micarelli and Sansonetti (2016) found that music recommenders based solely on personality had comparable performance with state-of-the-art recommender algorithms, but with particularly high diversity. Ferwerda et al. (2019) recommend personalised user interfaces to consider personality and expertise, possibly extracting them from social media, finding that they influence what kinds of music people like. Having the users do personality quizzes shouldn't be a problem either: users enjoyed using them to get music recommendations, and users with low domain knowledge saw the personality-based recommender as more useful than domain expert users (Hu & Pu, 2010). Many other studies have also used personality in recommendation systems for music (Ferwerda et al., 2017a; 2016; Moscato et al., 2020; Liu et al., 2020; Cheng et al., 2016; Schedl et al., 2016; Ferwerda & Schedl, 2016; Hu et al., 2010; Zhou et al., 2011; Gupta et al., 2020; Melchiorre et al., 2020; Bansal et al., 2020; Kouki et al., 2020).

With research confirming connections between film choices and personality (Song et al., 2009; Cantador et al., 2013; Golbeck & Norris, 2013; Karumur et al., 2017), films have become a common subject for personality-aware recommender systems. Many studies have also used personality in recommendation systems for games (Yang et al., 2019; de Lima, Feijó & Furtado, 2018; Chan et al., 2018; Hill et al., 2015; Abbasi et al., 2020; Yang et al 2017). Other topics have been various, including advertising based on the MBTI (Yang et al., 2022), marketing content generation based on the MBTI (Farseev et al., 2021), conference presentation recommendations for viewers with similar FFM traits (Asabere et al., 2020) and webtoons (digital comics) with the MBTI (Yi et al., 2016).

The approach by Yi et al. (2016) found that people with the same MBTI type selected similar emotional words to describe the same comics. This led to the conclusion that those with the same MBTI personality profile have similar movie preferences and similar interpretations of their emotions. They found that this helps with scalability by effectively having just 16 types of users, eliminating the usual recommender system need to calculating similarity between users. However, they also found a trade-off between recommendation accuracy and scalability, implying that the recommendations were not accurate, but the new user problem was solved.

Similarly, Song et al. (2019) proposed a collaborative filtering recommender using emotional word selection and the MBTI, finding that users with the same MBTI type, at least on the I/E (Extraversion) dimension, selected similar emotional words to describe the same films and had similar movie preferences. However, the found that the results for preferences weren't clear enough since they had been using very popular films, and recommended trying the same with controversial films. Tuedon (2020) then also used keywords based on the favourite films of participants with their MBTI types, with good satisfaction with generally less popular films.

8.6.4. Personality Recognition

Creating user personality profiles for recommender systems involves various approaches, and researchers have explored diverse methods to understand and leverage user traits for personalised recommendations. One approach, as demonstrated by Wu and Chen (2015), involves observing user behaviour to make personality judgements, enhancing collaborative filtering in film recommendation systems.

A more prevalent strategy is leveraging users' social media data. Khan et al. (2020b), for instance, employed Twitter and IMDB data to extract users' personality traits and values using IBM Watson's personality insights API. IBM Watson's tool is known for its capability to analyse textual data and provide insights into various psychological attributes based on linguistic patterns. This step allowed them to gain a comprehensive understanding of users' personalities by analysing their social media content. The researchers then extended their analysis beyond social media, delving into users' film preferences. To accomplish this, they employed the Linguistic Inquiry and Word Count (LIWC) tool. LIWC is a linguistic analysis tool that categorises words based on their psychological and emotional dimensions. In this context, Khan et al. used LIWC to examine linguistic attributes within film storylines.

Caridad Martín Sujo (2023) suggests a recommender system where novels similar to the user’s writing on Twitter are recommended, based on a system by Martín Sujo and Golobardes i Ribé (2022), which also returns the expected MBTI types of the characters in the novels. In their evaluation of various embedding techniques for text similarity, the researchers acknowledged that BERT (Bidirectional Encoder Representations from Transformers) is considered one of the most advanced embedding models in natural language processing. BERT, developed by Google, employs a transformer architecture that captures contextual information from both directions, allowing it to understand the meaning of words based on their surrounding context in a sentence.

Despite BERT's advanced capabilities, the researchers concluded that it is not fully suitable for calculating text similarity. In light of this, the researchers opted for simpler NLP techniques for their text similarity calculations. Specifically, they mentioned using Jaccard similarity and Word2Vec embeddings. Jaccard similarity is a straightforward method that measures the similarity between two sets by comparing their intersection and union. In the context of text, it can be applied to tokenised words to quantify the similarity between two texts. Word2Vec, on the other hand, is an embedding technique that represents words as vectors in a continuous vector space, capturing semantic relationships between words. This method allows words with similar meanings to have similar vector representations. The researchers chose Jaccard similarity over Word2Vec, citing execution time as a deciding factor. Jaccard similarity is computationally less intensive compared to more complex embedding models like BERT, making it a pragmatic choice for their specific requirements.

Christodoulou et al. (2022) also employed BERT for personality prediction. They used the Personality Café MBTI dataset (Keh & Cheng, 2019) to construct a restaurant recommender system. This highlights the integration of sophisticated machine learning models to derive nuanced personality insights, enabling more accurate recommendations.

Widdeson & Hadžidedić (2022) utilised the MBTI and LIWC in a multi-domain recommender based on Amazon reviews. This approach harnesses automatic recognition techniques to infer users' MBTI types from linguistic patterns within reviews, enhancing the recommender's ability to tailor suggestions across diverse domains.

In a related study, Szmydt (2021) incorporated users' FFM personality traits, interpreted from Amazon reviews, into a collaborative filtering model. This demonstrates the versatility of utilising personality insights extracted from user-generated content to refine collaborative filtering algorithms, thereby providing more context-aware recommendations.

These studies collectively showcase the evolving landscape of personality-based recommender systems, utilising diverse data sources and cutting-edge technologies to understand and cater to users' unique preferences and traits. Next, we will discuss the topic of personality recognition in more detail, focusing on doing it based on text.

9. Text-Based Personality Recognition

Personality prediction is one of the most difficult author profiling tasks in computational stylometry. It involves detecting personality traits on the basis of writing style. Most commonly, one of two personality taxonomies are used, the Five-Factor Model, or the Myers–Briggs Type Indicator.

The study of language has been a longstanding focus in psychology. In the mid-twentieth century, researchers began categorising word clusters into "dictionaries" to gauge individuals' needs, like affiliation, achievement, and power. For instance, those using words such as "win", "success", and "goal" were perceived as motivated by a pursuit of accomplishment, a notion supported by various research studies. However, many automated text analysis systems developed in the latter half of the twentieth century faced a significant drawback – opacity. These systems often relied on intricate rulesets or idiosyncratic prioritisation of words, limiting their interpretability. Specific psychological theories underpinning dictionaries also posed challenges when applied outside specific use-cases. These systems, though ambitious, became hindrances rather than providing accessible techniques for broad studies in human psychology (Boyd, 2017).

The understanding that language can be quantified to reveal insights into a person's psychology has significantly impacted psychological text analysis. Traditionally focused on content words, the late 1990s saw a shift towards considering how individuals express themselves. The development of user-friendly applications like Linguistic Inquiry and Word Count (LIWC) marked a crucial step in this direction. LIWC, introduced in the late 1990s, incorporates a dictionary with mappings for approximately 80 word categories, covering both content and function words. Its strength lies in comprehensive coverage, facilitating nuanced language analysis. LIWC has undergone iterations, been translated into multiple languages, and is widely used in cross-cultural research. Its user-friendly interface and robust development using psychometric techniques make it a valuable resource for researchers exploring the psychological dimensions of language.

In language research focused on personality, self-reports are often used to correlate language use with individuals' responses to FFM questionnaires. Many of these studies utilise psychological measures generated by LIWC. An initial study by Pennebaker and King (1999) discovered correlations between FFM personality measures and language use in students writing various types of stories and narratives. For instance, individuals with higher self-reported Neuroticism used fewer positive emotion words and higher rates of negative emotion words. Conversely, those scoring higher on Extraversion tended to use more social words. Subsequent research has confirmed that LIWC measures can be employed to estimate someone's personality, even extending to automatically assessing the personality of fictional characters based on their language.

Moreover, research using the Meaning Extraction Method (MEM) has identified patterns of word use that contribute to understanding personality. In a study by Chung and Pennebaker (2008), common social themes like sociability, maturity, and psychological stability were identified by extracting themes from individual writing samples. Analysing these themes in people's writing revealed a connection between the frequency of theme invocation and their scores on self-reports of the FFM. For example, those who used more words from the sociability theme tended to score higher on Agreeableness, while those using words from the maturity theme tended to score higher on Conscientiousness.

In automatic personality recognition, the use of LIWC dictionary text analysis has been common, notably used by Mairesse et al. (2007) to associate word frequencies with personality profiles. This approach has extended to social media data analysis, as seen in Celli and Rossi (2012). Alternatively, with the MBTI, studies focus on specific words or word classes correlating with personality traits, such as Extraversion linked to exclamation marks and social words (Schwartz et al., 2013). Recently, machine learning approaches have gained popularity in this domain.

A main problem in predicting personality from linguistic input has been the limited amounts of labelled data. Early FFM-based datasets (Argamon et al., 2005; Mairesse et al., 2007; Luyckx & Daelemans, 2008) often consisted of essays written in formal language, which inhibits the expression of personality. Other sources used included emails (Oberlander & Gill, 2006), conversations from electronically activated recorders (Mehl et al., 2001; Mairesse et al., 2007), blogs (Iacobelli et al., 2011), or Twitter (Quercia et al., 2011; Golbeck et al., 2011).

Among the 60 papers on text-based personality prediction Fang et al. (2022) survey, they found the FFM featured in 45 papers, and the MBTI in 14. As they note, the FFM uses a continuous spectrum, as opposed to the dichotomous approach of the MBTI. It has a much stronger empirical basis than the MBTI, making use of large-scale quantitative analysis of natural language and survey data, with extensive development and validation processes. The MBTI, on the other hand, is driven by theory, lacks empirical support and has only four proprietary questionnaires that have not been conclusively tested (Pittenger, 1993; Nowack, 1996; Grant, 2013). The FFM has been thought to be more reliable (Costa & McCrae, 1992) Štajner and Yenikent (2021) note that linguistic characteristics of the FFM have often been studied, but they were not aware of studies on linguistic characteristics of different MBTI types. Indeed, while the FFM was originated in lexical analyses (Cattell, 1946; Costa & McCrae, 1992), the MBTI fundamentally makes use of the behavioural implications in theoretical and professional contexts. Nevertheless, better performance with algorithms trained on the MBTI than the FFM has been reported, and that the algorithm used has a large difference to results with the FFM (Celli & Lepri, 2018). However, the judgement (J) vs. perception (P) dimension has been found particularly hard to predict from text (Plank & Hovy, 2015; Lukito et al., 2016; Verhoeven et al., 2018; Choong & Varathan, 2021).

Earlier work using social media data was generally smaller scale, but more recently, using larger social media datasets has become more common. On the FFM side, as seen in Chapter II.4.1, the huge MyPersonality dataset was used to a great effect (Schwartz et al., 2013a; Schwartz et al., 2013b; Park et al., 2015; Kosinski et al., 2015).

However, Myers-Briggs personality types have the advantage of being widely popular and therefore readily available based on users' self-reporting on social media. Plank and Hovy (2015) took use of this to collect a dataset of over 1.2 million status updates on Twitter, labelled with their MBTI. Instead of resorting to personality lexicons or similar resources as used to be typical, they used an open-vocabulary approach and used logistic regression over word n-grams and details of the user such as gender and followers.

Verhoeven et al., (2016) introduced the TwiSty corpus, a large Twitter dataset of 34 million tweets written by 18,168 users, labelled with MBTI information in six languages: German, Dutch, French, Italian, Portuguese, and Spanish, with most of them in Spanish. The corpus is heavily imbalanced across the MBTI classes, for example with many Introverts in the data, which is typical in these datasets, for example Tuedon (2020). Indeed, it has been noted that Extraverts tend to prefer offline modes of communication, while Introverts may find online communication easier and more accessible (Goby, 2006). Similarly, Picazo-Vela, et al. (2010) noted that those high on Neuroticism or Conscientiousness in the FFM were more likely to write online reviews.

Lukito et al. (2016) also use Twitter data, but in Indonesian. Kumar & Gavrilova (2022) also used tweets, but with deep-learning-based language models, such as BERT and USE, combined with a contextualised weighting mechanism. Though usually they would be removed, they found stylistic attributes such as frequency of URLs, mentions, emoji, and hashtag helpful. Katiyar et al. (2020) studied the MBTI with a set of 40 Twitter and Stack Overflow users, finding that Naive Bayes classification with TF-IDF counts is able to infer both personality traits and technical skills from text for the purpose of job recruitment.

Gjurković and Šnajder (2018) introduced the MBTI9K corpus, a collection of 354.996 Reddit posts written by 9,872 users in English – also heavily imbalanced across the MBTI classes. They preferred using multi-layer perceptron (MLP) classifiers using a range of alternative text features such as word and character n-grams. Wu et al. (2020) consider author-dependent word embeddings for author profiling classification and introduce a model called Author2Vec. They use a logistic regression classifier built from a subset of the MBTI9k corpus. Santos and Paraboni (2022) also describe a series of experiments fine-tuning BERT with MBTI9K.

Jiang, Zhang and Choi (2020) created the dialogue dataset FriendsPersona for automatic personality recognition in screenplays with the dialogue extraction algorithm MainSpeakerFinder, using both attentive networks and contextual embeddings with BERT and roBERTa.

Keh and Cheng (2019) introduced the application of BERT models in the realm of author profiling. The use of pre-trained language models, especially those based on architectures like BERT (Bidirectional Encoder Representations from Transformers), represents a paradigm shift in natural language understanding. BERT models, pre-trained on large corpora, capture contextual nuances and intricate language patterns, enabling them to grasp the subtleties of human expression. These models learn to represent words in the context of the entire sentence, considering both left and right context—a feature particularly advantageous for tasks like author profiling, where understanding the context is crucial. By leveraging BERT models for author profiling, Keh and Cheng harnessed the wealth of linguistic knowledge embedded in these pre-trained representations. The transfer learning aspect of pre-trained models allows them to adapt to specific tasks with relatively smaller labelled datasets. This adaptability is particularly beneficial in scenarios where collecting extensive labelled data for a specific profiling task might be impractical or resource-intensive.

Many recent studies have made use of their MBTI dataset, which featured 50 posts each from 8675 users of the Personality Café forum, together with their self-reported MBTI personality types. Das and Prajapati (2020) compare boosting, bagging, and stacking ensemble methods with concatenated TF-IDF counts and word embeddings. Cui and Xi (2017) tried out various models and seem to have found deep learning the best. Abidin et al. (2020) use random forest and text statistics features such as sentence length and punctuation etc. Khan et al. (2020a) and Amirhosseini and Kazemian (2020) use XGBoost ensemble learning. Sugihdharma and Bachtiar (2022) used Convolutional Neural Networks (CNN). Maulidah and Pardede (2021) used the Long Short-Term Memory (LSTM) algorithm with random oversampling. Choong and Varathan (2021) focused on Judging-Perceiving prediction, finding LightGBM and SVM the best. Mehta et al. (2020) found that when using BERT, it is better to use it with language modelling features rather than conventional psycholinguistic features.

Jain, Kumar & Beniwal (2022) also use BERT with the same dataset, and Kumar, Beniwal, and Jain (2023) propose evaluating the performance of various Support Vector Machine (SVM) kernels and using an ensemble of them with a variety of voting techniques. Wen et al. (2023) find that classification heads for fine-tuning pre-trained language models are often insufficiently trained when annotated data is scarce. Therefore, they propose to tune the models through personality-descriptive prompts based on the lexical hypothesis of personality, which suggests that personalities are revealed by descriptive adjectives. Yang et al. (2021a) use it to propose a multi-document Transformer that considers other posts during encoding each post, and a dimension attention mechanism to obtain trait-specific representations. Yang et al. (2021b) propose a psycholinguistic knowledge-based tripartite graph network that uses interactions between posts through psychologically relevant words and categories. Yang et al. (2023) propose a dynamic deep graph convolutional network to avoid the order of the posts affecting the results. Most relevantly to us, Fernau et al. (2022)

created a chatbot that mirrors the user's personality, judged from their chat messages utilising a pre-trained language model based on the dataset, finding that contrastive learning approaches outperform previous methods. The personalisation was also found helpful in a user study.

However, the studies using the Personality Café dataset (Keh & Cheng, 2019) generally do not provide enough statistics or detail for replication, or even properly judging the results in such an imbalanced dataset in which it would be easy to achieve approximately 80% accuracy in individual dimensions by just making the same prediction every time. Some of the results, particularly in Khan et al. (2020a), do not seem realistic, claiming nearly 100% accuracy in classifying personality traits, which is far more than could be expected from the task. A possible explanation could be data leakage occurring in resampling or data preparation.

Many MBTI studies have barely managed to outperform the majority-class baseline, achieved by simply predicting the majority class. A comparison of performances of FFM and MBTI computational models trained on Twitter data found that type of architecture and settings had little or no effect on the results (Celli & Lepri, 2018), which could indicate that Twitter data might not contain sufficient amounts of lexical signals. Linguistic cues found in short texts do not appear to directly correspond to the results of the questionnaire results, even when there are sufficient signals in the text and human annotators agree about them, the only exception being the Extraversion/Introversion dimension where the agreement between the annotators and questionnaires reaches 75-77% (Štajner & Yenikent, 2021). Indeed, in the FFM, Extraversion does have a good linguistic correspondence, and it is indeed the highest correlated dimension between the MBTI and the FFM models (Furnham, 1996).

10. Reading Preferences

10.1. Interest

Interest is very much central for reading preferences. But what kind of texts arouse the most interest? Sadoski, Goetz & Rodriguez (2000) found that concrete texts are more interesting, as they are easier to understand and remember. Schiefele (1996) found that when high school students were reading texts below their grade level in reading difficulty, verbal abilities were negatively correlated with interest and enjoyment. From this, we could make the rather obvious-seeming conclusion that a text's difficulty should be at an appropriate level for its reader's comprehension.

The emotional psychologist Paul J. Silvia (2006) lists the following variables that have been found to affect interest in various types of texts: coherence, ease of comprehension, prior knowledge, themes of death, simple vocabulary, suspense, sexual themes, vividness, author voice, concreteness, meaningfulness, imagery, readers' connections, surprisingness, importance, character identification, power themes, familiarity, unexpectedness, emotiveness, and engagement. He finds that only a few of them had been extensively tested, but that ease of comprehension and coherence have typically been found to be the most important factors in interest, as seen in multiple studies (Wade, Buxton & Kelly, 1999; Schraw, 1997; Schraw, Bruning & Svoboda, 1995).

However, Silvia (2006; 2008: 58) later raises objections to the inclusion of stimulus feature lists. He points out that according to appraisal theories of emotion, emotions stem from subjective evaluations of events, relying on people's interpretations rather than objective facts. Consequently, Silvia emphasises the significance of these appraisals in generating interest, particularly two specific types of appraisals.

The first appraisal pertains to the evaluation of novelty-complexity, wherein individuals assess an event as being new, unexpected, intricate, challenging to process, surprising, enigmatic, or unclear. The second, which may be less apparent, involves an evaluation of an event's comprehensibility. Appraisal theories categorise this appraisal as related to one's assessment of coping potential, where individuals consider whether they possess the necessary skills, knowledge, and resources to effectively manage the event (Lazarus, 1991). In the context of interest, individuals find themselves grappling with an unforeseen and complex event while attempting to comprehend it.

In essence, if people appraise an event as both novel and comprehensible, they are likely to perceive it as interesting. These considerations may bring to mind the ideas put forth by Berlyne (1971), who suggested that art elicits pleasure by achieving an optimal level of arousal through characteristics like novelty, complexity, surprise, uncertainty, and incongruity.

However, while novelty can indeed spark interest, it may not necessarily lead to greater enjoyment when compared to its counterpart, familiarity. Research extending back over a century (Meyer, 1903) has consistently shown that individuals tend to develop a stronger preference for music after repeated exposure to the same piece. Many subsequent studies have confirmed these findings (Huron, 2008: 131). However, it's important to note that there is likely a threshold beyond which hearing the same piece repeatedly could become tiresome or even irritating.

The evidence in favour of a preference for the familiar is compelling and applies not only to humans but also to animals (For a review, see Bornstein, 1989). Surprisingly, familiarity can yield unexpected outcomes. For instance, Derrick's (2012) research revealed that rewatching sitcom episodes can enhance one's motivation and self-control. This underscores our inclination towards things we are already familiar with, a phenomenon commonly referred to as the exposure effect or mere exposure effect, as described by Zajonc (1968). Notably, both familiarity and prior knowledge were included in Silvia's list of factors as well.

Indeed, the question of whether familiarity or novelty is preferable appears to lack a definitive answer. It likely hinges on various contextual factors and the reader's individual preferences. For instance, if an author aims to elicit a sense of predictable satisfaction from the reader, they must craft a narrative that follows established conventions and patterns. Conversely, if the goal is to surprise and intrigue the reader, the author must introduce unexpected twists and turns. In both scenarios, the reader's experience hinges on their expectations. These expectations are shaped by a multitude of factors, including literary conventions, recurring themes in other works, the narrative's own prior developments, and the reader's real-life experiences. In essence, the interplay between familiarity and novelty in literature is dynamic and context-dependent, with no universal rule governing which is superior. The key lies in understanding the desired emotional response and tailoring the narrative accordingly, keeping in mind that readers may have varying preferences for predictability and surprise.

Furthermore, the state of mind is crucial in a person’s interests. Arousal regulation pertains to how individuals select media content to attain their desired level of arousal, aiming to alleviate feelings of boredom or stress. This concept is substantiated by Bryant and Zillmann's (1984) experiment, where participants, induced with either boredom or stress, were given the freedom to choose from a range of television programs. In accordance with the principles of arousal regulation, those in the bored group opted for stimulating and exciting programs (such as game shows or highlights from a football game), while participants experiencing stress chose more soothing content (such as classical concerts or nature programs).

10.2. Empathy for Characters

Another essential element contributing to the enjoyment of literature revolves around the reader's ability to empathise with the characters portrayed in the narrative. The question arises: with whom do we establish this empathetic connection, and to what extent should this connection be forged?

The intensity of empathetic responses for real people is greatly influenced by various factors, including the individual's relationship with the person in distress, the perceived similarity between them and the person suffering, and the likability of the person in distress (Batson, Early & Salvarani, 1997; Kozak, Marsh & Wegner, 2006). Additionally, group membership plays a significant role in shaping the empathetic response (Yabar et al., 2006). For instance, the pain network in the brain is more activated when witnessing the suffering of loved ones compared to strangers (Cheng et al., 2010). Conversely, this activation decreases when observing someone who has behaved unfairly before (Singer et al., 2006).

One might instinctively respond that our capacity for empathy for fictional characters is most pronounced when we encounter characters who share similarities with ourselves. Indeed, a study by Jose and Brewer (1984) observed that children, when reading a story, exhibited heightened anxiety and affinity for characters they perceived as resembling them closely. Conversely, Koopman (2016: 266-267) noted that depressed readers did not feel deeply immersed in a narrative centred on depression because they discerned disparities between themselves and the characters.

These scenarios introduce us to the concepts outlined by Bullough (1912) regarding under-distancing and over-distancing. Striking the right balance, often referred to as achieving an optimal distance, is paramount when it comes to appreciating and reflecting upon a literary work. The definition of this optimal distance can vary significantly from one individual to another.

In situations where there is too little distance, the potential outcomes can be overwhelming. This can result in experiencing excessive empathy and, in some cases, even emotional distress. In certain instances, individuals might distance themselves from the material as a defence mechanism, akin to the "denial" defence often observed in empirical studies (refer to Baumeister, Dale & Sommer, 1998). An example of this phenomenon is apparent in

Koopman's (2016: 68-93) study, where it was found that university students without children were more profoundly affected by a book depicting the loss of a child compared to parents. This difference in emotional response could be attributed to factors such as age or psychological resistance on the part of the parents, who may have chosen to distance themselves from the content to cope with its emotional impact.

Similarly, Aristotle posits that we do not experience pity for those who are too intimately linked to us because their suffering essentially becomes our own (*Rhetoric* 2.8, 1386a18-23). Conversely, an excessive emotional detachment would impede our capacity for pity. Consequently, ideal fictional characters could be perceived as occupying a position in harmony with the Aristotelian concept of "the mean" concerning our empathetic responses. Aristotle extensively explores this idea in his ethical philosophy, where he contends that virtue lies in finding the desirable middle ground between excess and deficiency (e.g. *Nicomachean Ethics* 2.6.1106b15-29).

Miall (2011) contends that it is the characters' motives, rather than their inherent traits, that primarily drive affective engagement and reader self-projection into these characters. This perspective could potentially elucidate why people can form attachments to characters they encounter only briefly, such as in theatre or film.

Bortolussi and Dixon (2003: 240) propose the concept of "transparency," which refers to the evaluation of characters' behaviour as logical and reasonable, as a contributing factor to the process of reader identification. Keen (2006) suggests that we may need to set aside some conventional value judgments regarding literary techniques. For instance, critics' preference for characters with psychological depth should not necessarily hinder empathetic responses to flat, minor, or stereotyped characters. Gerrig (1990) adds that we tend to make category-based judgments about fictional characters, placing greater emphasis on the attributes and dispositions we attribute to characters rather than their actual behaviour. Therefore, easily comprehensible and accessible flat characters may play a more significant role in engaging the audience than commonly believed.

On a different note, Hogan (2001) argues in favour of categorical empathy, which hinges on characters aligning with one's group identity. Alternatively, situational empathy depends on readers having undergone similar experiences. However, it's worth noting that none of these claims appear to have been subjected to experimental testing (Keen, 2006).

Nevertheless, it appears that establishing a connection with characters does not demand an extensive set of criteria, and sometimes, subtlety and understatement can be valuable. In fact, empathy can potentially diminish when readers are overly exposed to a character's inner thoughts or narrative voice.

The manner in which writers employ language plays a pivotal role in shaping readers' empathetic responses. While for some readers, adherence to formulaic conventions might enhance empathetic resonance, others may find that less conventional and more striking representations stimulate foregrounding and amplify empathetic reactions (Miall & Kuiken, 1999).

Pette (2001) observed in her study that readers who focused on the stylistic aspects of a novel became equally emotionally invested as those who primarily engaged in identification with the characters. However, these "aesthetic" readers tended to oscillate between being more and less emotionally moved compared to their identificatory counterparts. Many identificatory readers reported instances where they paused and created emotional distance from the characters to prevent becoming overly emotionally involved.

According to Cupchik (2002), encountering a work that is excessively realistic can be emotionally overwhelming. Conversely, when stylistic features are too prominent or unconventional, readers may struggle to establish a strong emotional connection. This perspective suggests that novel language usage may lead to over-distancing. However, Koopman, Hilscher, and Cupchik (2012) conducted an experiment that yielded intriguing results. They found that readers exhibited greater empathy toward a rape victim depicted in an aesthetically enhanced text by Gloria Naylor, compared to a similarly explicit but less aesthetically adorned text by Virginie Despentes. Nevertheless, it remains unclear whether this effect stemmed from preventing over-distancing or under-distancing, or just the quality of the language per se. Therefore, it is plausible that the prominent artistic use of language could enhance empathetic responses, but the extent of this influence may vary depending on the individual reader and their preferences.

10.3. Painful Responses

Empathy can be a double-edged sword, capable of causing emotional distress as well as fostering connections with individuals deemed worthy of it, particularly those who bear some resemblance to the empathising person. Functional neuroimaging studies have shown that perceiving another person's suffering can trigger similar neural responses to experiencing the suffering firsthand, encompassing both physical and emotional pain. This phenomenon is observed when viewing facial expressions of pain, witnessing physical injuries, imagining the pain of others, encountering social exclusion, or even receiving cues that someone will undergo painful experiences (Lamm, Decety & Singer, 2011). Moreover, studies have revealed that individuals who are more sensitive to physical pain also tend to be more sensitive to social pain (Asmundson, Norton & Jacobson, 1996; MacDonald & Kingsbury, 2006; Ciechanowski et al., 2002; Ehnvall et al., 2009; Waldinger et al., 2006), suggesting shared sensitivity to both types of distress (Eisenberger, 2012: 8).

Research involving functional magnetic resonance imaging (fMRI) has uncovered nuanced aspects of empathy. For instance, participants were found to be more sensitive to the pain experienced by individuals who contracted AIDS due to a blood transfusion as opposed to those who acquired it through drug use (Decety, Echols & Correll, 2009). Another study demonstrated that the neural response to perceiving others in pain is weaker when the individuals are from a different ethnic group (Xu et al., 2009). In the realm of sports, football fans reported higher pain ratings and empathetic concern when watching fans of their own team as opposed to those of a rival team. Furthermore, they exhibited greater activation in the anterior insula and were more willing to share the pain of fellow fans (Hein et al., 2010). Additionally, an investigation revealed that failures by in-group members led to pain, whereas the shortcomings of out-group members elicited pleasure (Cikara, Botvinick & Fiske, 2011). It's worth noting that the personal context of the viewer also influences affective arousal. For example, physicians were found to be less emotionally affected by the pain of others compared to individuals outside the medical profession (Cheng et al., 2007).

These considerations have clear implications for fiction as well. The more emotionally charged the content of a story, the more it activates brain regions associated with both cognitive and emotional empathy (Altmann et al., 2012; Hsu, Jacobs & Conrad, 2015). Moreover, as a text becomes more story-like and evokes greater emotional responses, it leads to increased activation in areas relevant to theory of mind processing (Wallentin et al., 2011). This lends support to the fiction feeling hypothesis (Jacobs, 2014), suggesting that more emotionally charged narratives tend to elicit higher levels of empathy and immersion.

In the context of reading, there is evidence of physiological changes reminiscent of fear responses. For instance, sentences describing fearful situations have been shown to produce a more substantial increase in heart rate compared to less emotionally arousing text (Vrana, Cuthbert & Lang, 1986). Experiments involving reading about frightening encounters led to accelerated heart rates and increased skin conductance (Lang et al., 1983). Unsurprisingly, individuals with a phobic fear of snakes exhibited particularly strong reactions to passages describing snakes.

However, it's important to note that what readers experience isn't precisely fear. Fear typically involves the activation of the amygdala, which is not consistently observed even when individuals watch horror films. In an fMRI study conducted by Straube et al. (2010), participants reported experiencing subjective feelings of anxiety, correlated with heightened activity in the dorso-medial prefrontal cortex (DMPFC), responsible for evaluating the emotional significance of stimuli and situations. Straube's team noted a lack of amygdala activation and hypothesised that this might be due to the sustained anxiety present in the clips, as opposed to sudden and unexpected threats. However, it's important to note that these clips were relatively short and intense, especially when compared to reading. Additionally, researchers explored the role of sensation-seeking in horror enjoyment and found that sensation-seekers, who are generally more aroused by stimulating material, displayed greater visual cortex activation when exposed to the clips. They also exhibited lower activity in the thalamus and insula when encountering less frightening content, suggesting potential under-arousal in ordinary circumstances. Conversely, individuals with low sensation-seeking tendencies exhibited higher baseline activation, which could deter them from seeking out challenging or stimulating experiences.

As such, it is more appropriate to describe the predominant emotion experienced while reading as anxiety rather than fear. These considerations underscore that readers can indeed experience pain when exposed to narratives featuring suffering. The intensity of this empathetic response is heavily influenced by the qualities of the characters portrayed. When a character is innocent and similar to the reader in some way, the reader's pain becomes intertwined with that of the character. However, if these conditions are not met, the empathetic effect may be diminished or even transformed into Schadenfreude.

11. Emotion Detection

Game practitioners sometimes conduct formal playtest session in artificial play environments for qualitative research, collecting and analysing subjective data through direct observations, interviews and think-aloud protocols. These methods have been found to good at representing accurate states, but they have shortcomings. True play experiences might be inhibited in artificial circumstances where someone is observing or questioning the players. While still playing, discussion might be distracted as well as distracting, affecting both the gameplay itself and its description. However, discussion after playing would be limited by how well the experience is remembered. Such sessions are also costly, and a more efficient method would be able to capture more people's experiences. Therefore, a lot of research has been done using quantitative methods using objective data, featuring approaches such as telemetry and psychophysiology. Telemetry logs players' in-game interactions to build player models. It is non-disruptive and can continuously capture objective gameplay statistics in natural settings. However, it cannot properly capture the player's experiences, being limited to the player's in-game actions and having little idea of how the player is feeling or what they are thinking, or what the motivations for the recorded actions are.

Psychophysiology uses physiological measurements to infer psychological states. The techniques include electrodermal activity (EDA), electromyography (EMG), electrocardiogram (ECG), electroencephalography (EEG), body temperature, pupil dilations and respiratory activity. EDA and EMG would appear the most popular, being as they are found good in measuring arousal and valence, respectively. Psychophysiology can measure player experiences continuously in real-time, but can also represent the real-life experiences of the player (Blom et al., 2014: 31). Unfortunately, most current approaches deal with expensive specialised equipment that are obtrusive, which are usually only viable in controlled laboratory settings.

Affective computing is a field concerned primarily with systems and devices for recognising, interpreting, and simulating human emotions (Kanjo, Al-Husain & Chamberlain, 2015). It has been used for personalising games, however most frequently, appears in academic studies rather than in practise. In addition to psychophysiology, one may track bodily expressions (motion tracking), with the assumption that particular bodily expressions are linked to expressed emotions and cognitive processes. This has included facial expressions, muscle activation (typically face), body movement and posture, speech, text, haptics, gestures, brain waves, and eye movement (Yannakakis & Togelius, 2018b). While such measurements can be very informative, a major limitation with most of them is that they can be invasive impractical and uncomfortable, as well as being problematic regarding privacy.

Affective systems can also often use self-reporting of emotions, a relatively feasible and lightweight method, but users may not be able or willing to report their true emotions. A better approach might be to use commonly available devices. Even though smartwatches and other such devices are widely available, many may still consider them too invasive for the sake of personalisation in terms of privacy, or even physically. Furthermore, the signals are not always being reliable or easily interpretable, especially when the user is reading, which is a less intense activity than gaming. Nevertheless, they could still prove to be a promising approach to gaining information helpful for personalisation.

12. Conclusions

This chapter has explored personalisation within the digital landscape, focusing on player modelling, psychological frameworks, player types, procedural content generation, interactive storytelling, and the pivotal role of recommender systems.

The exploration commenced with an analysis of the nuanced dynamics of personalisation, a dynamic force reshaping the digital landscape. It has become apparent that personalisation transcends mere colloquial usage to represent a profound paradigm shift in the interaction with digital content. Its applications in gaming and interactive storytelling have given rise to immersive experiences, wherein each user assumes the role of an author shaping their own narrative. In the domain of recommender systems, personalisation has evolved beyond content discovery to assume the role of a companion accompanying individuals on their digital journeys.

The discussion on player modelling and profiling illuminated the power of understanding users' preferences, but it also revealed the limitations of rigid player typologies in accommodating diverse narratives. Psychological personality frameworks such as the Five-Factor Model of personality (FFM) emerged as an interesting tool for personalisation. Their adaptability and effectiveness, particularly when faced with limited user data, offer a versatile tool for personalisation, promising to bridge the gap between user preferences and narrative content. In parallel, the Need for Affect (NFA) emerged as a lesser known yet pivotal measure of media preferences, casting a spotlight on the emotional landscapes users seek within narratives. The NFA provides a nuanced lens through which we can understand the varying degrees of emotional depth and intensity desired by users, paving the way for more emotionally resonant narratives. Furthermore, the Myers–Briggs Type Indicator (MBTI) was noted as a popular but less consistent framework, which nevertheless has the benefit of having more data available.

Recommender systems, discussed in detail, have emerged as the bridge between users and a vast repository of creative content. These systems hold the power to mitigate the challenges faced by new users, offering them personalised content from the very start. With robust user profiles, recommender systems can unlock the potential of collaborative filtering, further enhancing content recommendations.

Our foray into text-based personality recognition showcased the complexities involved in predicting personality traits from linguistic styles. The exploration of reading preferences elucidated the nuanced factors influencing textual interest, empathy for characters, and responses to distressing narratives.

Finally, we touched upon emotion detection, highlighting the challenges and promise in leveraging commonly available devices for enhancing personalisation through affective computing.

As we conclude, it becomes evident that the trajectory of personalisation in digital experiences is dynamic and promises continuous exploration, refinement, and adaptation to the evolving landscape of user expectations and technological advancements. Many fields, considerations and approaches were found relevant to our aims. The literature review underscores the need for a nuanced and multifaceted approach to personalising literary narratives. Traditional personality models like the FFM provide a strong foundation for understanding individual differences, but the integration of these models into narrative generation remains underexplored. The potential of narrative PCG, combined with insights from gamification and text style transfer, offers a promising pathway for developing personalised narrative experiences.

The thesis will focus on leveraging the FFM as a foundational framework for personalisation, integrating it into a narrative PCG system to tailor story elements such as plot and character to individual personality profiles. This approach aims to enhance engagement and resonance in literary narratives, providing a more immersive and personalised experience for users. Moreover, the thesis will address the challenges identified in the literature, particularly concerning the balance between narrative coherence and personalised content. By drawing on methodologies from gamification and text style transfer, the thesis will explore innovative ways to embed personality assessment within narrative experiences, ensuring both accuracy in personality representation and engagement in the narrative itself. It will also explore alternative models of understanding personality and preferences through MBTI and Need for Affect, and the potential of the approach to extend to recommender systems and understanding personality through writings.

In conclusion, the thesis will contribute to the field by bridging the gap between psychological models of personality and narrative personalisation, offering a novel approach to personalising literary narratives that is both theoretically grounded and practically applicable.

13. Summary

The literature review has studied many fields, topics and aspects pertaining to narrative personalisation, covering player modelling, psychological frameworks, player types, procedural content generation (PCG), interactive storytelling, the crucial role of recommender systems, and much more. This comprehensive examination has highlighted the transformative potential of personalisation, revealing its capacity to significantly enhance user engagement by tailoring content to individual preferences and needs.

The review began by analysing the broad concept of personalisation, underscoring its profound impact on the digital landscape. The discussion on player modelling and profiling emphasised the importance of understanding user preferences, yet also pointed out the limitations of rigid player typologies in accommodating the diverse narratives that users engage with; nevertheless, more appropriate measures for narratives could take inspiration from them, which will be discussed in Chapter VII. The exploration of psychological personality frameworks, particularly the Five-Factor Model (FFM), revealed its potential as a more versatile tool for personalisation, especially when user data is limited; this will be used Chapters IV and V. While the Myers-Briggs Type Indicator (MBTI) was acknowledged for its popularity, its inconsistent reliability was also noted, though it remains valuable due to the extensive data available for analysis, which will be examined in Chapter VI.

The examination of recommender systems highlighted their role in connecting users with personalised content, particularly for new users who benefit from immediate, tailored recommendations. The review also addressed the challenges of text-based personality recognition, which involves predicting personality traits from linguistic styles. This exploration revealed the complexities of aligning reading preferences with narrative elements such as character empathy and responses to emotionally charged content. These topics will be explored again in Chapter VI.

The literature review underscores the necessity for a nuanced and multifaceted approach to personalising literary narratives. While traditional personality models like the FFM offer a strong foundation, their integration into narrative generation remains an underexplored area. The combination of narrative PCG with insights from gamification and text style transfer represents a promising direction for developing personalised narrative experiences.
The thesis will build upon these insights by leveraging the FFM as a core framework for personalising narratives.

Furthermore, the thesis will address the challenges identified in the literature, particularly the need to balance narrative coherence with personalised content. By incorporating methodologies from gamification and text style transfer, the thesis will explore innovative strategies for embedding personality assessment into narrative experiences, ensuring both accurate personality representation and engaging storytelling. It will also consider alternative models like the MBTI and NFA, and the potential for these approaches to extend to recommender systems and personality assessment through text analysis.
Ultimately, the thesis aims to contribute to the field by bridging the gap between psychological personality models and narrative personalisation, offering a novel, theoretically grounded, and practically applicable approach to personalising literary narratives.

Chapter III: Method

The thesis employs a mixed methods approach, with both qualitative and quantitative approaches, leveraging various methods and techniques to explore the intersection of AI-driven personalisation and narrative content. These methods are designed to explore personalisation, storytelling, and audience engagement across different chapters and studies:

User studies (Chapters IV & V): Both the first and the second study depend on finding participants and having them do personality tests and share their experiences on reading the stories.

Interactive Narrative Creation (Chapter IV): The first study employs a creative approach by crafting an interactive narrative. This narrative serves as a tool to capture user choices and preferences. The creation of this narrative involves storytelling techniques and user experience design. User interactions within the narrative are carefully designed to reveal insights into their preferences. The narrative choices are aligned with the Five-Factor Model (FFM) and the Need for Affect (NFA) to create user profiles.

Personality Frameworks: The thesis utilises established personality frameworks, primarily the Five-Factor Model (FFM) and the Myers–Briggs Type Indicator (MBTI), as instruments for understanding user preferences. These frameworks are employed to assess personality traits and preferences based on user interactions, language style, and textual data.

Natural Language Processing (NLP): Natural Language Processing techniques are applied in the second study to automate personalisation. Various NLP models are trained and tested on participants to assess their effectiveness in adapting language style to user preferences. NLP involves pre-processing textual data, training machine learning models, and evaluating their performance.

Machine Learning: The third study explores the use of machine learning algorithms to classify users based on their MBTI types using textual data. Machine learning involves data preprocessing, feature engineering, model selection, training, and evaluation. The goal is to create a predictor that can estimate user personality traits from their textual input.

Data Collection and Analysis: Throughout the thesis, data collection and analysis play a pivotal role. Surveys, user interactions with the interactive narrative, and textual data are collected and analysed to draw insights into user preferences, personality traits, and emotional responses. Statistical analyses are applied to evaluate correlations and relationships between variables.

Ethical Considerations: Ethical considerations are integrated into the research process, ensuring responsible data handling and transparent practices. This involves obtaining informed consent from participants, anonymising data, and addressing privacy concerns.

These methods collectively form a robust framework for exploring AI-driven personalisation in narrative content. The combination of creative narrative creation, personality frameworks, NLP, machine learning, and ethical considerations allows the thesis to investigate personalisation from multiple angles and draw meaningful conclusions about its potential impact on user engagement and content recommendations.

Chapter IV: The Interactive Narrative

Abstract

In this chapter, we present a user study developed to explore the use of psychological frameworks for the personalisation of narratives. Further, we explore using interactive narratives to understand the user's personality and narrative preferences. The study consists of three sections: an interactive narrative, a personality test, and a personalised short story. Whilst it would appear that at least this interactive narrative could not be used as a personality test per se, it was able to capture some traits. The personalisation appeared to work remarkably well, especially regarding relating with the protagonist. In fact, personalisation based on either the personality test or the interactive narrative worked well, suggesting that the interactive narrative might have been able to capture personality better than the direct comparison with the personality test results would suggest. It was also found that Extraverted people appear to prefer reading narratives with less formal language, and Introverts prefer narratives with more formal language.

Introduction

What makes a good story? Any subjective answer to the question would, by definition, be down to the person answering the question. To present them a suitable story, we could find one matching their preferences, or, more intriguingly, make one fit them. Trying to understand the person using methods from psychology, the narrative could be made to have different variations for different personalities. But how do we get an understanding of the person's personality? Using their social media data would not always be possible or ethical. Using a personality test might not necessarily much fun to the user. Then, why not use a method that should be fun for anyone interested in narratives: a narrative? That is what this study does: presents the users with an interactive narrative designed to capture their personality using the Five-Factor Model (FFM) and the Need for Affect (NFA), and then personalises a narrative to match with their personality scores. This study seeks to consider various ways of personalising at the same time, focusing on written, non-interactive narratives, and testing how this affects the reader experience.

A critical component of this study is the innovative use of interactive narratives as a medium for personality assessment. Traditional methods of personality assessment, such as self-report questionnaires, can be intrusive or unappealing, potentially impacting the validity of the data collected. In contrast, using interactive narratives provides a more engaging and naturalistic method for personality assessment, potentially leading to more accurate and unobtrusive data collection. This method also aligns with the thesis's broader goal of enhancing the user experience through engaging and enjoyable means.

The study further aims to empirically evaluate the impact of personalised narratives on the reader's experience. This involves assessing various dimensions of engagement, such as immersion, enjoyment, and emotional response, to determine whether personalisation based on psychological profiles can indeed enrich the reader's interaction with the narrative. The findings are expected to provide empirical support for the theoretical propositions regarding the benefits of personalised content, thereby contributing to the academic discourse on narrative theory and digital storytelling.

By combining psychological theories of personality with narrative techniques, the study offers a comprehensive framework for personalising content in literature and media. This framework not only advances academic understanding but also provides practical guidelines for content creators interested in implementing personalised storytelling. Additionally, the research addresses ethical considerations and technical challenges associated with personalisation, ensuring a balanced approach that considers both user experience and the responsible use of personal data.

Methodology

The study had 59 participants (17 women, 36 men, 6 other or not disclosed), volunteers of all ages above 18 who were found by posting about the study on internet discussion boards on interactive narratives and other relevant topics in spring and summer of 2020.

Table 1: Participants by Age Group

Age group	Number of participants	Percentage of participants
18-24	16	27.12
25-34	23	38.98
35-44	12	20.34

45-54	6	10.17
65+	1	1.69
Prefer not to say	1	1.69

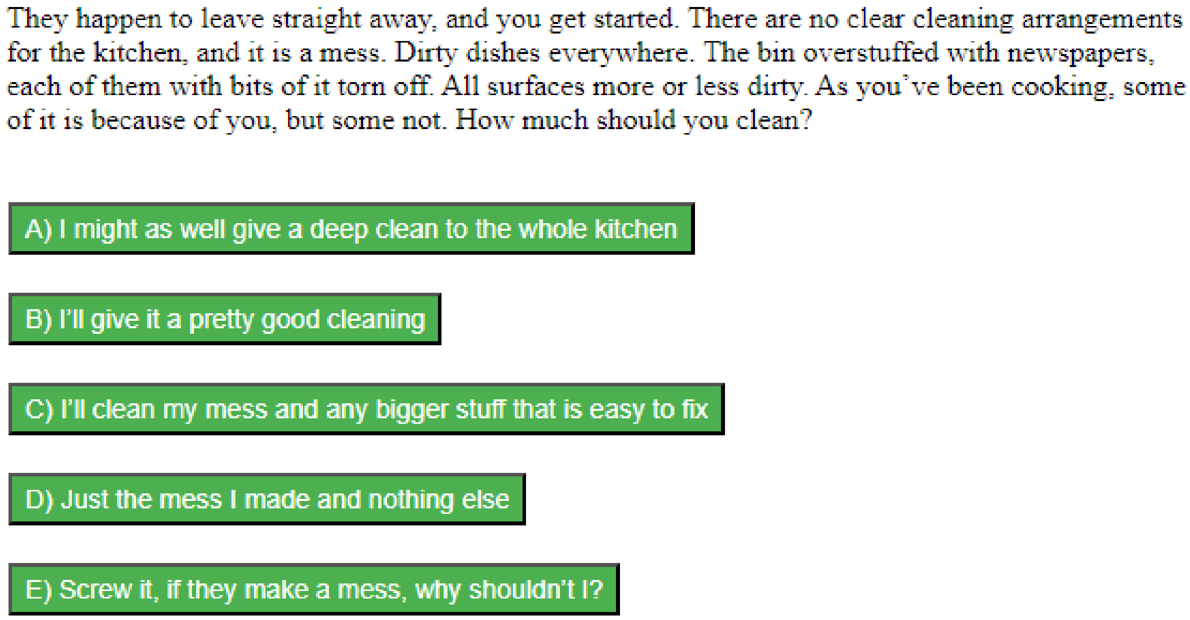
The first part is an interactive narrative specifically written for this study. It uses a second-person perspective where the user assumes the role of the protagonist and makes choices that determine what the protagonist does and how the plot advances.[1] All but one of the 25 questions simulate a personality questionnaire on a five-point Likert scale, with the possible options ranging from one extreme to another, measuring either one of the factors in the Five-Factor Model (FFM) or the Need for Affect (NFA). The questions relating to the FFM are about how the protagonist responds to the situation, for example in a very Extraverted or a very Introverted manner, and the questions measuring NFA are about deciding what happens next, ranging from something light-hearted to something very dark. The questions measuring FFM sought to emulate individual questions at the archive of FFM questions available at the website of International Personality Item Pool;[2] for example, the question presented below in Figure 2 assesses a person's Conscientiousness, particularly aspects related to responsibility, orderliness, and consideration for others, and can be matched with IPIP items such as X118 (Like order) or Q104 (Keep things tidy), and many others. On the other hand, the questions measuring NFA were not based on NFA questionnaires per se, and were a more liberal way of measuring how emotional they would like the narrative to be. Most choices do not affect the narrative trajectory, except for three questions, resulting in 2 x 3 x 5 = 30 different ways the story could end up.

[1] The exact instructions were:
"The first part of the experiment is an interactive narrative specifically written for the experiment and simulating a personality test. You are presented as the protagonist and must choose one of the presented options.

You will then take a 10-item Five-Factor Model questionnaire and a 10-item Need For Affect questionnaire. The purpose is to see how well personalisation could be done based on what sort of choices people would make in interactive narratives, and how closely their in-narrative persona matches with reality.

In the final part, you are presented with a short story written for the experiment and personalised to you. Users are split randomly into three groups. Group 1 will have the personalisation done according to the results of the personality test, group 2 according to the narrative, and one control group will get the opposite of what they would get in group 1. The users are asked how much they liked the story and how much they identified with the protagonist, which is used for measuring how useful the personalisation was."
[2] https://ipip.ori.org/

Figure 2: An example question from the interactive narrative.

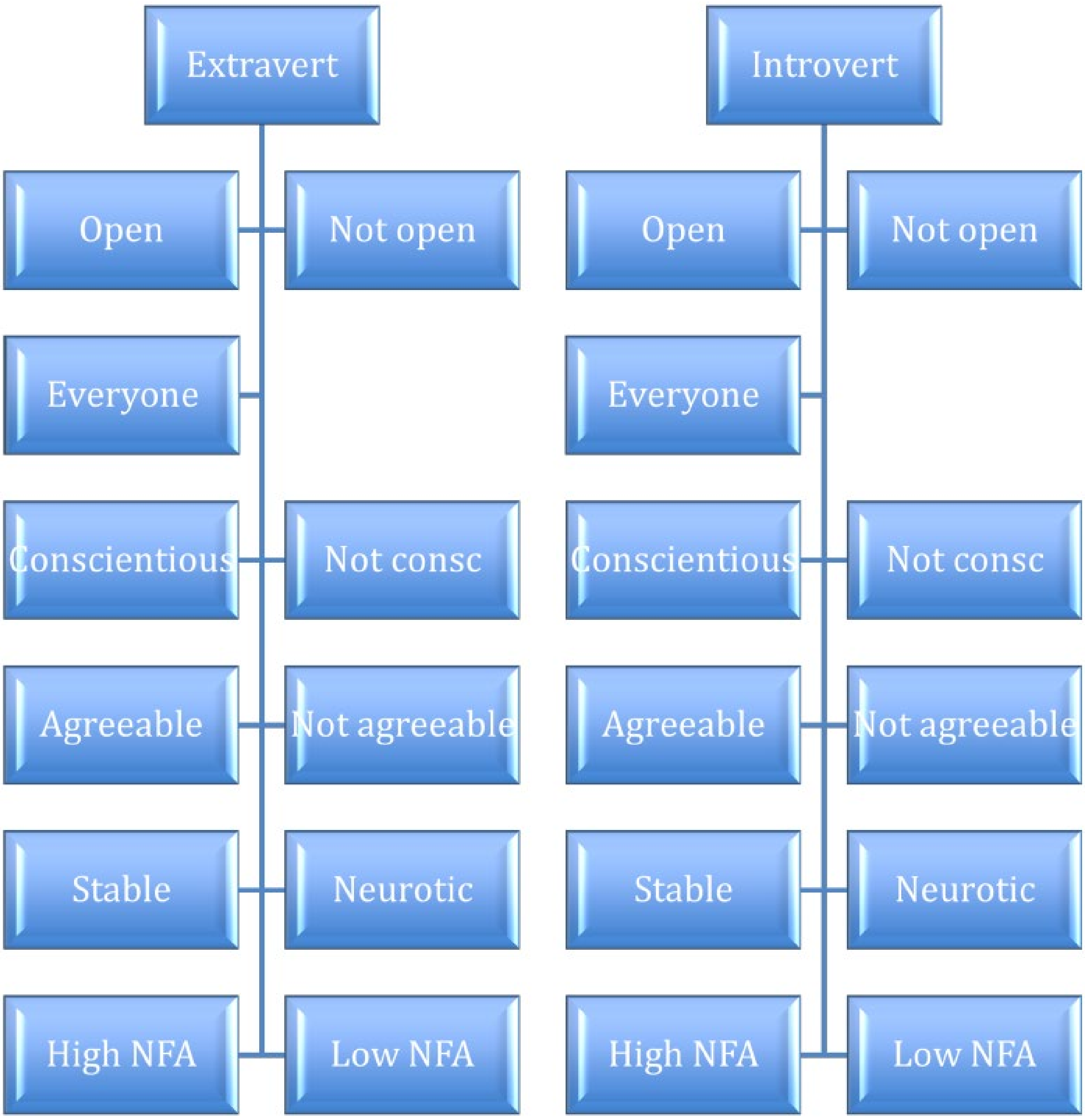


The participants then take a 10-item FFM questionnaire (Rammstedt & John, 2007) and a 10-item NFA questionnaire (Appel & Maio, 2012). Each question in both the narrative and the personality test is given a score from 0 to 4, and the total score for each trait is scaled in a normalised range from 0 to 1. Finally, the scores for the questions in the narrative are readjusted with the use of ridit scores approach (Bross, 1958) to take into consideration how specific questions tend to get answered. Here too, a normalised range from 0 to 1 is used.

In the final part, the users are presented with a short story written for the experiment and personalised for them. Like the interactive narrative, it was manually authored, with different versions prepared, section by section. Users are defined as being either high or low in a given trait and are given a version that matches with that. Group 1 have this done according to the results of the interactive narrative presented (henceforth referred to as IN), group 2 according to the personality test (PT), and one control group will get the opposite of what they would get in group 1. The users are asked how much they liked the story and its language and how much they identified with the protagonist, which is used for measuring how useful the personalisation was.

The story is about a group of friends going to a yoga retreat on a small island and then finding out that the rest of the world is suffering from a pandemic that turns people into zombies that have an overwhelming need to hoard toilet paper. There are two different versions from the start: high and low Extraversion. This affects the use of language throughout every section. The story is then split into six different sections, and, apart from the section that is the same for everyone in the same Extraversion pathway, there are two versions of each section under both Extraversion pathways depending on whether the user has a high or low score in a given trait. Each section displays one dimension of the protagonist's personality. For example, in the opening section, in the version meant for people high in Openness to Experience, the protagonist is happy about going for a yoga retreat, but in the low Openness version, they are reluctant and only agreeing to it under peer pressure. The protagonist's personality depends on the user's FFM scores other than Extraversion, and the ending on the NFA: a high NFA indicates a preference for a more emotional, tragic ending; a low NFA a less emotional, somewhat happy ending. Therefore, there are $2^6 = 64$ different variations of the story.

Figure 3: The structure of the personalised short story. Extravert and Introvert version are different throughout, and consist of sections depending on whether a trait is above or below 0.5/1.



FFM traits have been found to have various correlations with what sort of language people use, and here it is hypothesised that people would also like to read the sort of language they prefer to use. The most important FFM trait in this and many other respects has been found to be Extraversion (Mairesse et al., 2007), and therefore, to avoid complicating things, it is the only trait used for personalising language in this study. Since it is used for this purpose, it is not used in other forms of personalisation, e.g. adjusting the protagonist's personality, to separate the effect of personalising language from that of personalising the character. Nevertheless, since Extraversion is arguably the most important and the most widely understood FFM trait, it is the most likely one to affect identifying with the character, so if the other traits are helpful at all, Extraversion would be highly likely to be so as well.

Table 2: The 4 versions of a section of the short story displaying either Emotional Stability or Neuroticism.

Version	Passage
Stable, extraverted	During the next few days, Brian started acting oddly. Normally he was always talking, but now he would just stare into emptiness all the time and be unresponsive. Well, I say emptiness, but I mean the toilet paper stack, usually. We started wondering if he had caught the virus. The others were stressed out by the situation and had been having difficulties sleeping, but I didn't really mind so much. But that night I was the one awake for some reason. So it happened that I heard walking sounds leading in and out of the house, and I went to take a look, and for a good reason -- it was Brian taking the toilet paper to the boat. I immediately shouted to wake everyone up and ran to the boat. "What are you doing?" I shouted at him, as I stepped on the boat. "Going home", he said calmly, starting the engine. I kept yelling at him as the boat started moving, but the noise drowned out my voice. He was not reacting to me at all. I approached him to get him to stop, but he did not budge. I tried to stop the engine, but he pushed me off so I almost fell off the boat, and so one thing led to another -- he was the one in the sea now. Or perhaps "it" was in the sea now -- or how should I call a zombie? This is what I was thinking when he was there, drowning in front of my eyes, pleading for help.
Neurotic, extraverted	During the next few days, Brian started acting oddly. Normally he was always talking, but now he would just stare into emptiness all the time and be unresponsive. Well, I say emptiness, but I mean the toilet paper stack, usually. We started wondering if he had caught the virus. This was yet another thing to keep me up at night -- the stress was making me feel sick and had been keeping me up for a while. So I was once again struggling to sleep and heard walking sounds leading in and out of the house, so was more alert than usually. It turned out to be for a good reason -- it was Brian taking the toilet paper to the boat. I immediately shouted to wake everyone up and ran to the boat. "What the hell are you doing?" I shouted at him, as I stepped on the boat. "Going home", he said calmly, starting the engine.

	I kept yelling at him as the boat started moving, but the noise drowned out my voice. He was not reacting to me at all, and I was totally freaking out. I approached him to get him to stop, but he did not budge. I tried to stop the engine, but he pushed me off so I almost fell off the boat, and so one thing led to another -- he was the one in the sea now. Or perhaps “it” was in the sea now -- or how should I call a zombie? This is what I was thinking when he was there, drowning in front of my eyes, pleading for help, and I was frozen by panic.
Stable, introverted	During the next few days, Brian started acting oddly. Normally talkative, now he would just stare into emptiness all the time, unresponsive. Well, I say emptiness, but I mean the toilet paper stack, more often than not. We started wondering whether he had caught the virus. Being stressed out by the situation, the others had been having difficulties sleeping, but I didn’t really mind so much. Nevertheless, that night I was the one awake for some reason. So it happened that I heard walking sounds leading in and out of the house, and I went to take a look, and for a good reason -- it was Brian taking the toilet paper to the boat. I immediately shouted to wake everyone up and ran to the boat. “What are you doing?” I shouted at him, as I stepped on the boat. “Going home”, he said calmly, starting the engine. I kept yelling at him as the boat started moving, but the noise drowned out my voice. He was not reacting to me at all. I approached him to get him to stop, but he did not budge. I tried to stop the engine, but he pushed me off so I almost fell off the boat, and so one thing led to another -- he was the one in the sea now. Or perhaps “it” was in the sea now -- or how should I call a zombie? This is what I was thinking when he was there, drowning in front of my eyes, pleading for help.
Neurotic, introverted	During the next few days, Brian started acting oddly. Normally talkative, now he would just stare into emptiness all the time, unresponsive. Well, I say emptiness, but I mean the toilet paper stack, more often than not. We started wondering whether he had caught the virus. This was yet another thing to keep me up at night -- the stress was making me feel sick and had been keeping me up for a while. So when it happened that I was once again struggling to sleep, and I heard walking sounds

	leading in and out of the house, I was more alert than usually, and for a good reason -- it was Brian taking the toilet paper to the boat. I immediately shouted to wake everyone up and ran to the boat. “What the hell are you doing?” I shouted at him, as I stepped on the boat. “Going home”, he said calmly, starting the engine. I kept yelling at him as the boat started moving, but the noise drowned out my voice. He was not reacting to me at all, and I was terrified. I approached him to get him to stop, but he did not budge. I tried to stop the engine, but he pushed me off so I almost fell off the boat, and so one thing led to another -- he was the one in the sea now. Or perhaps “it” was in the sea now -- or how should I call a zombie? This is what I was thinking when he was there, drowning in front of my eyes, pleading for help, and I was frozen by panic.

The way the use of language is personalised here is based on previous studies that have shown that people with high Extraversion write using simple constructions; short sentences; few quantifiers; informal, affective language; the pronouns “we” and “which”; confident language featuring much words such as “want” and “need”; stylistic expressions such as “catch up” and “take care”; and a lot of semantic errors. Introverts, on the other hand, prefer the reverse: more long, formal and complex sentences; few errors; the pronoun “I”; negations; quantifiers; and less confident language such as “trying” and “going to” (Mairesse, 2007).

In the different passage versions in Table 2, some of these differences are on display. In crafting different versions of the narrative passages to reflect various personality traits, the aim was to subtly tailor the language and tone to resonate with the psychological profiles of the intended readers. In the versions meant for people with high Emotional Stability, the narrative employs a more composed and detached tone, reflecting a calm response to the unfolding events. The character's actions are described in a straightforward manner, emphasising a controlled and less emotional reaction. For example, the stable versions lack overt expressions of fear or panic, aligning with the trait of emotional stability, which is characterised by resilience and composure under stress. In contrast, the Neurotic versions include language that conveys heightened anxiety and emotional turmoil, such as "totally freaking out" or "frozen by panic." This choice reflects Neuroticism, where individuals are more likely to experience negative emotions intensely. The inclusion of phrases that express fear or anxiety aims to create a narrative that neurotic readers might find more relatable or immersive.

The Extraverted versions incorporate simpler sentence structures and a more informal tone, in line with the linguistic tendencies of Extraverted individuals. The use of confident language and colloquial expressions is designed to resonate with Extraverted readers, potentially enhancing their engagement with the narrative. For Introverted readers, the narrative uses more complex sentence structures and formal language. The inclusion of first-person pronouns, along with more tentative expressions, aims to create a narrative voice that aligns with the introspective and cautious communication style often associated with Introversion.

It can be noted that all versions do feature the verb “try”, preferred by Introverts; this is because the protagonist in the passage tried to do something but failed. The decision to maintain certain consistent language elements across all versions, such as the use of the word "tried," serves the purpose of narrative coherence. Whilst Extraverted people use the word less, they still use it. Maintaining a baseline of consistent language helps preserve the core narrative structure, ensuring that the story remains recognisable and coherent regardless of the reader's personality. This consistency is crucial for experimental control, allowing differences in reader response to be more confidently attributed to the variations in narrative tone rather than significant plot alterations. The FFM personalisation strategy is intentionally subtle, focusing on tone and emotional cues rather than overt changes in the storyline – which was the role of the NFA personalisation. This subtlety ensures that the narrative remains universally accessible while still offering nuanced differences that can resonate more strongly with certain personality types.

Results

As the Shapiro-Wilk test showed that, as expected, the personality data was not normally distributed, it was found better to use primarily Spearman correlations, though Pearson correlations were also checked for. The personality scores given by the interactive narrative (IN) had varying correlations with the personality scores given by the personality test (PT), slightly improved by the ridit analysis. The Spearman correlations between the results of the IN and the PT are displayed in Table 3 below.

Table 3: Spearman correlations of traits as judged by IN and PT, readjusted with ridit scores.

Trait	Spearman correlation	*p*-value
Extraversion	0.425	0.001
Emotional stability	0.323	0.012
Conscientiousness	0.155	0.241
Agreeableness	0.128	0.335
Need for affect	0.035	0.791
Openness to Experience	0.018	0.89

It appears the IN was able to make an approximate assessment of the user's Extraversion and Emotional Stability, but not the other traits. The NFA(IN), or the NFA according to the IN, did not have any significant correlations with anything, especially with the NFA(PT), but got close to significant correlations with Agreeableness(PT), $\rho=-0.233$, $P=0.076$, and with Openness(PT), $\rho=-0.211$, $P=0.108$, both negative but not quite significant at $p<0.05$.

The IN also appears good at judging Openness in those identifying as women ($\rho=0.469$, $P=0.058$), but for men the correlation is actually negative ($\rho=-0.273$, $P=0.107$); neither are quite statistically significant. Particularly for women, statistical significance was hard to reach due to the small number of participants who identified as women (17).

Some people took the experiment rather quickly and presumably carelessly, and one person confessed to just skim-reading the personalised narrative. Therefore, we exclude from the personalised short story analysis the 14 people who spent less than ten minutes on the test.

Table 4: Average scores by group, scale 1-5.

Group	Liking story	Relating	Liking language
1 [n=9] (IN)	3.56	2.78	3.55
2 [n=19] (PT)	3.37	3.00	3.58
3 [n=17] (Control, IN)	3.00	2.35	3.00
Kruskal-Wallis *p*-value	0.342	0.197	0.022

According to Kruskal-Wallis tests, the only statistically significant inter-group difference for the ratings for the story (Table 4) was with the language, $P=0.022$. However, the scores between the groups are not perfectly comparable; for example, group 2 got the sad ending much more than others, which could skew the results slightly in their favour, since that ending was more liked on average. This is because users tended to get high NFA scores in the PT (average 0.61), which defined the personalisation of group 2, but the results were more balanced in the IN (average 0.45), which was used in groups 1 and 3. Nevertheless, it is easy to note that the control group performed the worst in every aspect, suggesting the personalisation did improve the experience.

Table 5:Significant correlation pairs, comparing groups presented with different version of the short story.

		Personality according to interactive narrative			Personality according to personality test		
Trait	Rating	Corr (high-trait group)	Corr (low-trait group)	*p*	Corr (high-trait group)	Corr (low-trait group)	*p*
Extraversion	Relating with protagonist	0.387 [n=21]	-0.24 [n=24]	0.021			
Extraversion	Liking the language	0.136 [n=24]	-0.41 [n=21]	0.037	0.389 [n=21]	-0.638 [n=24]	<0.001
Conscientious-ness	Relating with protagonist	0.381 [n=29]	-0.176 [n=16]	0.044	0.109 [n=29]	-0.426 [n=16]	0.048
Stability	Relating with protagonist	0.104 [n=26]	-0.422 [n=19]	0.044			
Openness	Liking sad ending	-0.364 [n=25]	0.333 [n=20]	0.012			
NFA	Relating with protagonist more in sad ending	0.169 [n=25]	-0.382 [n=20]	0.038			

We can note that Extraversion according to the PT had 0.387 correlation with relating with the protagonist in the high-Extraversion version of the short story (n=21), but -0.24 in the low-Extraversion version (n=24), meaning that the more Extraverted the user is, the more they relate with the protagonist if the language used is Extraverted, but if the language is Introverted, the reverse happens: the more Introverted the user is, the more they relate! This indicates that the personalisation was very successful in this respect, with a clear correlation between Extraversion and liking relating with the protagonist if they use Extraverted language. The *p*-value for such a difference in correlations is 0.021. There were many such correlation differences, with only the statistically significant ones presented in Table 5. Notably, the correlations were much more obvious when looking at personality judgements according to the interactive narrative rather than the personality test results, which might suggest that the interactive narrative was more effective at capturing the preferences than the personality test.

Looking more at Table 5, we can note that adjusting the language depending on level of Extraversion worked well regarding liking the language, and, in the case of the PT, also with respect to relating with the protagonist (who was also the narrator). Personalisation based on Conscientiousness and Emotional Stability(IN) also worked particularly well in making the protagonist relatable. Openness(PT) indicated liking the happier, less emotional ending, and therefore would apparently have been better for personalising the ending than the NFA was, though the NFA(IN) appeared to work too, but did not quite reach significance (-0.028 vs. -0.45, P=0.079). The NFA(PT) seemed to have the opposite effect, which would have reached significance without removing the fast experiment takers. However, the NFA(IN) did work in making the protagonist relatable. Generally, doing the personalisation based on the IN worked in many ways much better than the PT, though many correlations weren't found quite significant and therefore weren't mentioned here.

Gender was also a major factor with the ending. For men, the ending made little to no difference, as versions with the happy ending were found just marginally better (3.27 [n=15] vs. 3.07 [n=15], Mann Whitney U P=0.24). However, for women, the sad ending was greatly preferred (3.71 [n=7] vs. 2.6 [n=5], Mann Whitney U P=0.014).

Table 6: Ratings by gender.

	Men			Women		
	Happy ending [n=15]	Sad ending [n=15]	Mann Whitney U *p*	Happy ending [n=7]	Sad ending [n=5]	Mann Whitney U *p*
Liking	3.27	3.07	0.24	2.6	3.71	0.014
Relating	3.0	2.4	0.071	2.4	3.0	0.11
Language	3.4	3.33	0.41	3.0	3.57	0.043

As noted above, the IN was able to make an approximate assessment of the user's Extraversion and Emotional Stability, but not the other traits. It should be noted that its way of measuring the NFA did not match with the way the NFA is tested, but with the way the authors of the NFA describe the preferences for art that people with high NFA are expected to have: the more emotional and intense, the better. This study would give some indication that this is not necessarily the case.

The scores given by the IN followed a more standard distribution than those from the PT, particularly with ridit scores. According to the PT, the participants had a particularly low average score in Extraversion (0.32) whilst being high in Openness to Experience (0.77), for example, which makes sense given the way they were recruited, but this could skew the results given by the IN, which had all of the average scores between 0.43 and 0.54 before ridit, and 0.45 and 0.50 after ridit. Some choices in the IN were far more popular than others, typically with bias in favour of the middle options. When the bias was away from the middle options, however, the ridit scores pulled the scoring closer to the middle – which is essentially the purpose of ridit scores.

In question 23, the user is asked whether to slip a housemate's medications into his drink, as he has been refusing to take them, but needs them. The choice is presented only to add to the user experience and the user's sense of control, and has no influence on the user's personality scores. Interestingly, almost half of the users (26/59) decided to do so, and of those who did, few (8/26) wanted to see him unwell afterwards, though this was very common (28/33) in the group that chose not to slip them! This was the only question where previous choices could have such influence on answers, being avoided specifically for issues like this. This gives some clue that while people often want to see suffering in interactive narratives, they don't want to feel like it's their fault.

Discussion

It was found that Extraverted people appear to prefer reading narratives with less formal language, and Introverts prefer narratives with more formal language, or specifically, the types of language Extraverts and Introverts have been found to write; this does not appear to have been tested before. Whilst it would appear that at least this interactive narrative could not be used as a personality test per se, it was able to capture some traits, specifically Extraversion and Emotional Stability. It is possible that with Agreeableness, Conscientiousness and Openness to Experience, people might indeed have a preference to act within fiction differently from how they would in reality; for example, someone who is Agreeable in reality might want to get a safe experience of what it is like to be rude. Whether they would then want to see protagonists behaving like this as well, or preferably like they would in reality, is an open question.

We should also consider the possibility that the personality tests did not measure traits ideally. Short versions were used to not bother the participants too much, but longer versions might have been more accurate. On the other hand, some people could have been rather uninterested in the personality test section and clicked through it rather carelessly, and making it longer could have exacerbated such a problem. It is therefore possible that interactive narratives could capture at least some aspects of personality even better than personality tests, and in fact the IN appeared to be better than the PT for personalisation. Questions in personality tests can be rather abstract and open to interpretation, even ambiguous, but in an interactive narrative, the user is put into a specific situation in a rather concrete manner. The NFA, however, did not appear to work as intended, except with the way the IN interpreted it. Therefore, perhaps the NFA(IN) could be re-defined simply as a preference for tragic rather than light-hearted themes in narratives – perhaps this could be called Preference for Tragedy, or PFT.

The personalisation with individual FFM traits also appeared to work well for relating with the protagonist. However, the effect could have been limited by the fact that the sections displaying the protagonist's personality were rather brief. Therefore, that a type of personalisation did not appear to work with this story does not mean that it could not work when done better, in a longer narrative, or with more participants, and that it did appear to work here could in some cases be down to just chance. Similarly, at least some of the questions in the interactive narrative could have been just poorly made. Therefore, more similar studies would be helpful. Other ways of personalisation could also be done based on the FFM, such as more novelty for people with high Openness for experience. NLP could be used for altering the writing style, vastly easing the process.

By an intriguing coincidence, while this study was being conducted, McCord, Harman & Purl (2019) also used interactive narratives, or as they call them, text-based fantasy games, to function like FFM personality tests. Their narratives were different in the sense that it was more like a game, aiming to complete a quest, whereas this narrative had no goals but perhaps enjoyment of the narrative. They used three different narratives, two of which made the user choose a course of action associated with one FFM trait over another, and one, like our narrative, set high or low levels of a trait against each other. In the former narratives, Openness and Neuroticism failed to get significance, but others were successful. In the one resembling ours more, it was Conscientiousness that was non-significant. Their Pearson correlations with IPIP-50 were 0.34 for Openness, 0.09 for Conscientiousness, 0.43 for Extraversion, 0.34 for Agreeableness and 0.22 for Neuroticism. In comparison, our test's equivalent Pearson correlations were 0.03, 0.17, 0.38, 0.11 and 0.34. Taking a look at all their correlations and ours, it looks like Extraversion has very consistent correlations, but the other factors have different results in each study.

Table 7: Pearson Correlations between narratives and FFM tests.

	McCord, Harman & Purl (2019) study 1	McCord, Harman & Purl (2019) study 2	McCord, Harman & Purl (2019) study 3	Our study
Openness	0.20	0.13	0.34	0.03
Conscientiousness	0.38	0.20	0.09	0.17
Extraversion	0.42	0.42	0.43	0.38
Agreeableness	0.36	0.26	0.34	0.11
Neuroticism	0.48	0.00	0.22	0.34

In the future, it should be studied how interactive narratives could better capture personality. More such narratives should be written, and the kinds of choices that best correlate with personality scores should be chosen for further usage. Such choices would not necessarily have to be on a Likert scale. Collaborative filtering could also be used, and with enough participants and questions, surprising links could be found, which in turn could help personality research as well. Finally, the user profile thus created could also later be used for a recommender system, particularly with narratives, but possibly with other domains, as well; as discussed in the literature review, it has indeed been found that the FFM can be useful in recommenders, particularly when there is little data available on the user (Tkalčič, 2011), as well as for increasing the diversity of recommendations (Onori, Micarelli & Sansonetti, 2016). Ultimately, the approach could be used to personalise just about every aspect of narratives, as well as to recommend and perhaps generate more.

Conclusions

In this chapter, the investigation into personalised narratives has revolved around two central elements: the interactive narrative and the personalised short story. These twin focal points have jointly yielded interesting insights into user preferences, personality traits, and narrative experiences, offering more understanding of the interplay between interactive storytelling, psychological frameworks, and user engagement.

While the interactive narrative functioned as a gamified platform, crafted around the Five-Factor Model (FFM) and the Need for Affect (NFA), the personalised short story emerged as a distinctive and immersive component. The short story, personalised to individual user profiles, turned out to be successful in reflecting the user's personality and emotional inclinations. Different parts of the short story wove together elements that aligned with the user's Extraversion, Agreeableness, Conscientiousness, Emotional Stability, Openness to Experience and Need for Affect. The gamification principles extended to the personalised short story, transforming it into a dynamic and participatory narrative that adapted to the user's unique psychological profile. This approach not only heightened user immersion within the story but also provided a rich source of data for in-depth analysis.

The amalgamation of the FFM and the NFA in this context adopted an approach reminiscent of a factorial or vectorial model, as elucidated in Chapter II.3. In this framework, multiple factors were encapsulated within a numerical profile, offering insights into various aspects of the user's characteristics. This profile, essentially a multifaceted representation, served as the foundation for implementing personalised experiences. As emphasised by Charles et al. (2005: 14), the process of determining the pertinent variables for inclusion in the profile isn't always straightforward. However, leveraging statistical techniques becomes instrumental in discerning the variables that hold significance in contributing to a comprehensive user profile. As a result, the derived user profile can become a dynamic tool for tailoring experiences, illustrating the synergy between psychological theories, statistical methodologies, and the ultimate goal of enhancing user personalisation.

The findings suggest that while Extraverted individuals tend to prefer narratives with less formal language, Introverts lean towards more formal language, aligning with the writing styles associated with each personality type. The study demonstrated the potential of interactive narratives in capturing certain traits, particularly Extraversion and Emotional Stability. Personalising the protagonist based on the user’s FFM personality turned out to work strikingly well, regardless of whether the FFM score was based on the interactive narrative or a traditional personality test, suggesting that the interactive narrative might have captured the users’ personality better than the personality tests might suggest. This highlights the opportunities in using narrative-driven approaches for personality assessment, raising questions about the adaptability of traditional personality tests in capturing nuanced preferences within fictional contexts.

Moreover, the personalised short story section, informed by the interactive narrative and personality tests, showcased varying user responses based on the levels of Extraversion and other personality traits. The language personalisation aligned with Extraversion levels, demonstrating a correlation between language style and user satisfaction. The study's approach offers a promising avenue for further research into refining interactive narratives for enhanced personality capture and tailoring narrative experiences to individual preferences.

Summary

Chapter IV utilises an innovative approach that combines interactive narratives, psychological frameworks, and personality tests. The study, designed around the FFM and the NFA, employs an interactive narrative to capture user personalities, followed by a personalised short story tailored to individual traits. The findings provide empirical support for the approach of the thesis, as it was found that changing language style and personality of the protagonist increased the enjoyment. This contributes to theoretical discourse and providing a practical approach for personality-based personalisation of narratives. The FFM performed even better than expected in every respect, but the use of the NFA turned out to need refinement. To advance the approach of personalising text style personalisation, however, using NLP rather than manual edits would often be more practical, which will be explored in the next chapter.

Chapter V: Text Style Transfer

Introduction

In our prior study, the process of editing language to personalise narratives was carried out manually, a meticulous endeavour that required significant effort. However, in the pursuit of creating longer narratives tailored to individual preferences, the integration of automated techniques such as text style transfer, a task in natural language processing (NLP), has emerged as a promising avenue. It involves the transformation of text while preserving its meaning. In the context of literature, it allows for the modification of linguistic style, tone, and even language complexity, catering to the unique preferences of readers. This study aims to investigate the fusion of text style transfer techniques with insights from psychology to uncover how written narratives can be systematically modified to resonate with individuals of varying personality traits.

A user study was structured in a way that seamlessly combined AI-driven text style transfer with psychological profiling. Central to this study is the utilisation of psychological frameworks, namely the Five-Factor Model (FFM), as in the previous study, as well. By administering a ten-question FFM personality test to participants, we aimed to gain insights into the preferences of people with different personality profiles. Subsequently, they were immersed in an AI-modified short story that had undergone text style transfer to align with specific personality traits. Participants were then asked to express their opinions on the AI-modified short story. Their feedback encompassed various dimensions, including emotional resonance, engagement, and overall satisfaction with the narrative. Through this feedback, we aimed to discern the effectiveness of AI-driven language adaptation in creating a more personalised reading experience.

The study aims to develop the way narratives are personalised. The integration of text style transfer in combination with psychological profiling can offer readers narratives that align with their unique cognitive and emotional preferences. This approach transcends the limitations of manual editing, making it feasible to personalise not only the language but also the overall narrative structure and tone. The findings can be applied to a broad spectrum of content, including novels, articles, and various forms of written communication. By harnessing AI-driven personalisation, content creators can ensure that their narratives resonate with readers across the personality spectrum, from those who prefer simplicity to those who seek linguistic artistry. As with any study involving personal data and AI-driven personalisation, ethical considerations are paramount. Protecting user data, ensuring transparency, and maintaining privacy remain essential pillars of this research.

Method

The study explores the integration of artificial intelligence and psychological insights to create personalised narratives that adapt to individuals' personality traits. By combining text style transfer techniques from natural language processing and the Five-Factor Model personality assessment, the research aims to determine how AI-driven language adaptation can enhance the personalisation of longer narratives. A user study is conducted, where participants undergo a personality assessment, read AI-modified short stories, and provide feedback on the tailored narratives. This method allows us to investigate the efficacy of AI-driven personalisation in creating a more engaging and personalised reading experience.

We set out to create different versions of the same story to different participants to compare their opinions to their FFM personality scores, their reading skills and their age and gender. Different language styles were tried out using text style transfer. Furthermore, different endings were used, with happy, sad and ambivalent versions edited manually, aiming to make minimal changes to the original, but resulting in different outcomes for the protagonist.

To do this, we needed to find a suitable text style transfer method, and for this, a literature review was conducted (see Chapter II.7.3.). To explore the state of the art at the time, the Fuzhenxin GitHub repository[3] was of great help, listing research papers related to the topic. We examined all the papers listed and also extended our review to include additional relevant sources. Our objective was to comprehensively evaluate these methods and determine their suitability for our research project. The aim was to find approaches that change the literary style producing good, readable output, and provide instructions or code that can be used to actually do the style transfer. However, many text style transfer approaches and papers were found not to really fit our definition of text style transfer seeking to change the style of a text without changing its meaning. Generally, the field consists of aiming to change sentiment, formality, genre, or the political slant or the gender of the writer. However, as noted in Chapter II.7.3., attribute transfer, such as sentiment, gender, and political transfer, does not in fact appear to change the style, but the meaning of a sentence,

[3] https://github.com/fuzhenxin/Style-Transfer-in-Text

which means that in our opinion, it should not be called text style transfer, and would not suit our purposes here, which are to change how formal or informal and how creative or conventional the language is. On the other hand, some papers and approaches were considered more relevant to our purposes, but upon closer inspection, the results of the text style transfer seemed not very fit for purpose. In many instances, there seemed to be no way to actually use the method described, and making the code public in particular was rare to see.

Among the array of text style transfer methods we investigated, we found the approach presented by Krishna et al. (2020) particularly promising and well-aligned with our research objectives. Their method stood out for its capacity to facilitate the transformation of writing styles while delivering results that did clearly change the style, with decent fluency. Their approach is to employ a novel strategy rooted in paraphrase generation to achieve the desired style adaptation. In essence, they utilise unsupervised paraphrase generation, a technique that involves the creation of pseudo-parallel data. This is accomplished by subjecting sentences from varying styles to a diverse paraphrase model. The underlying principle is to normalise the input sentence, effectively removing elements that are indicative of its original style.

With the input thus pre-processed, a specialised inverse paraphrase model tailored to the original style is trained. The role of this inverse paraphrase model is to reconstitute the input sentence while preserving its original style. In this way, the model can effectively reproduce the distinctive style of the original text without unwarranted alterations to its semantic content. This approach offers a promising means of achieving style adaptation while maintaining the core message and meaning of the text.

Their work also involved collecting a large dataset of 15 million English sentences spanning 11 diverse styles, including the works of James Joyce, romantic poetry, tweets, and conversational speech. We used two of their models, Shakespeare and song lyrics, to modify a short story, and trained one of our own based on their approach, using 1015 youth novels available for free at Smashwords[4].

The selection of different linguistic styles for the text style transfer was grounded in the expectation that these styles would resonate differently with individuals based on their personality traits. The original and youth styles were intended to represent formal and

[4] https://www.smashwords.com/

informal versions of prose, respectively. The original style, with its sophisticated vocabulary and intricate sentence structures, was expected to appeal to Introverted individuals who favour more complex and formal language. The youth style, in contrast, employed simpler and more straightforward language, aligning with the communication preferences of Extraverts who typically favour direct and uncomplicated expressions.

In addition to these prose styles, two more lyrical approaches were incorporated: Shakespearean and song lyrics. The Shakespearean style, characterised by its elaborate and archaic diction, was hypothesised to appeal not only to Introverts but also to individuals with high Openness to Experience, due to its uniqueness and historical richness. The song lyrics style, with its informal, emotive, and often rhythmic language, was anticipated to attract Extraverts who are drawn to spontaneous and affective communication. This style was also expected to engage those with high Openness to Experience, given its creative and unconventional nature.

Indeed, both the Shakespearean and song lyrics styles, due to their distinct and unusual qualities, were likely to appeal to individuals who enjoy novel and imaginative expressions, typical of those with high Openness to Experience. The youth style, being more conventional and straightforward, was presumed to resonate with people with high Extraversion and low Openness to Experience, who prefer clear and simple communication.

However, these expectations were approached with an exploratory mindset. The primary objective was to select styles that were significantly different from each other in terms of formality, complexity, and lyrical quality, to investigate how these variations influence reader engagement and experience. By examining a range of narrative tones and structures, the study aims to provide insights into the potential for personalised storytelling.

Their official code is available on GitHub[5]. The code used PyTorch 1.4+, HuggingFace's transformers library for training GPT-2 models, and Facebook AI Research's fairseq for evaluation using RoBERTa classifiers. Initially, I started working with the code in November 2020, when it had just been released, and did not have an extensive readme yet. Emails were exchanged with the researchers for more instructions. I modified their code to accommodate for processing longer passages, and to fix some problems encountered, such as the generated language going in loops and repeating words and phrases. Their original paraphrase generation worked line by line; I made it work sentence by sentence and

[5] https://github.com/martiansideofthemoon/style-transfer-paraphrase/tree/master

process whole txt files in one go, accommodating different encoding formats and considering punctuation better, producing cleaner text output. I also had to modify their code to suit Windows due to not having the permissions to install Ubuntu or WSL using the departmental desktop.

However, the desktop was not powerful enough for training a custom model, lacking CUDA. Initially, we attempted to do the training on Google Cloud Platform (GCP), but ran into issues with their lack of GPU availability at the time, apparently due to the then-widespread phenomenon of using them for cryptocurrency mining. Fortunately, the department then gave me access to a powerful desktop for the task, but unfortunately, getting to use it got delayed due to a lengthy sick leave. The desktop was equipped with a NVIDIA Quadro RTX 4000, reported to have GPU Memory of 8 GB GDDR6, 256-bit memory interface, up to 416 GB/s memory bandwidth, 2304 NVIDIA CUDA® Cores, 288 NVIDIA Tensor Cores, 36 NVIDIA RT Cores, single-precision performance of 7.1 TFLOPS and tensor performance of 57.0 TFLOPS. By this point, the plain text from the youth novels had been split into training, testing and development datasets, and a label dictionary had been built. Then, an image of the GCP VM was downloaded to be used on the desktop to carry on with the process. Next, the dataset was paraphrased using GPT-2, and inverse paraphrasers were trained. This required leaving the desktop processing for a long time. This got interrupted due to running out of disk space, which was solved by asking another student for permission to remove his old files. After this, the training took approximately four weeks to complete.

After this, it was time for finetuning. Due to it being summer, the department was largely closed, which meant I was only allowed access to the desktop once a week. Typically, I would work on the code that day and then leave the desktop running, then coming back the next week to find out the results. Initial attempts at finetuning failed due to the desktop running out of RAM. This was fixed by cutting out about 95% of the text data used for finetuning. This wasn’t considered a problem, as with large language models, the finetuning does not require as much data as we had gathered. The next problem was running out of disk space again, as the process involved creating many space-consuming checkpoints. To solve this problem, the number of checkpoints to be created was greatly reduced. Nevertheless, the finetuning still took another two weeks, approximately.

The models were then used to adapt a short story adapted from *The Cloak*, by Nikolai Gogol and translated by Thomas Seltzer. There were three different endings of the story: the original ending, where the protagonist dies and comes back as a ghost, considered the sad ending; one where his death was just a misunderstanding, considered the happy ending;

and one where whether he really died or not is left ambiguous, considered the ambivalent ending. The changes were done manually to the original version. In the original version, it is stated that “At length poor Akaky Akakiyevich breathed his last. They carried Akaky Akakiyevich out, and buried him”; in the happy version, this is replaced by “And so it was in this state that he got up, sold his coffin to some passerby, and left”; in the ambivalent version, this bit and the following paragraph is simply removed, leaving it unclear what really happened, as we jump straight from him being ill to him being replaced at work and rumours spreading of his ghost wandering around. Below, in Table 8, contiguous passages from the end of the happy version of the story are presented in all the different styles. The sentence added to turn the ending into the happy version, where his death was just a misunderstanding, is in bold.

Table 8: Passages from The Cloak.

Original	The next day a violent fever developed. Thanks to the generous assistance of the St. Petersburg climate, the malady progressed more rapidly than could have been expected, and when the doctor arrived, he found, on feeling the sick man’s pulse, that there was nothing to be done, except to prescribe a poultice, so that the patient might not be left entirely without the beneficent aid of medicine. But at the same time, he predicted his end in thirty-six hours. After this he turned to the landlady, and said, “And as for you, don’t waste your time on him. Order his pine coffin now, for an oak one will be too expensive for him.” Did Akaky Akakiyevich hear these fatal words? And if he heard them, did they produce any overwhelming effect upon him? Did he lament the bitterness of his life?—We know not, for he continued in a delirious condition. Visions incessantly appeared to him, each stranger than the other. Later on he talked utter nonsense, of which nothing could be made, all that was evident being that these incoherent words and thoughts hovered ever about one thing, his cloak. **And so it was in this state that he got up, sold his coffin to some passerby, and left.** Several days later, the porter was sent from the department to his lodgings, with an order for him to present himself there immediately, the chief commanding it. But the porter had to return unsuccessful, with the

	answer that he could not come; and to the question, “Why?” replied, “Well, because he is dead! he was buried four days ago.”
Youth	And the next day a new official sat in his place, handwriting by no means so upright, but more inclined and slanted than the previous. The sudden rumour spread like wildfire through St. Petersburg, that a dead man had taken to appearing upon the Kalinkin Bridge, and its vicinity, at night in the form of an official seeking a stolen cloak, and that, under. One of the department officials saw the dead man with his own eyes, and immediately took note of the man's face and name - Akaky Akakiyevich. This, in turn, inspired him with such terror, that he ran off with all his might, and therefore did not scan the dead man closely, but only saw how the latter was threatening him from a distance with his finger. However, we have totally neglected that certain influential personage who may truly be considered the cause of the fantastic turn taken by this true history. First of all, justice compels us to say, that after the departure of poor, annihilated Akaky Akakiyevich, he felt something akin to remorse. He had even resolved to send an official to him, to learn whether he was truly capable of assisting him, the thought troubled him to such an extent that a week later he had even. And when it was reported to him that the Akaky Akakiyevich, the darling of his school, had died suddenly of fever, he reeled, hearkened to the reproaches of his conscience, and was out of sorts for the.
Shakespeare	For one of his friends’s houses, Where he hath a goodly feast in hand, Wishing to divert his mind in some other and drive away The disagreeable impression, he set forth that evening For one of. After supper he drank a bottle of champagne—not a bad recipe for cheerfulness, as every one knows. In his time of pilgrimage, The champagne inclined him to divers adventures, And, not returning home, he resolved not to part with his garments, But to go and see a certain well-known lady, Of German extraction, of whose mother he is. Then the important personage descended the stairs, Stood on his sledge, and said to the coachman, “To Karolina Ivanovna”. ” So that the means.

	Some one else by the collar clutched him suddenly. Turning round, he perceived a man of short stature, in an old, woreèd uniform, and recognised, not without terror, A certain fellow, in a grave, of the name. The funeral face of the official's was as white as snow, And looked just as if it had been a pile of. I will not name the name of the important personage, For he's the most dreaded of all, When he saw the dead man's mouth open, And heard it utter the following sentences, While it breathed upon him the most hideous.
Lyrics	I've got you, that—by the collar! I need your cloak. Taken no trouble but reprimanded me. So now give yourself away your own". The pallid prominent personage almost died from fright. Whirlwinded his cloak hurriedly from his shoulders and yelled to his coachman in an unnatural voice, "Home at full speed! Nervous, carefully frightened, and cloakless, went home instead of Karolina Ivanovna's, reached his room somehow or other, and passed the night in the dreadst distress; so that the next morning over their own. But papa remained silent, and said not a word to any one of what had happened to him, where he had been, or where he had intended to go. This event made a deep impression. And even started to say, "How dare you? Do you realize who's standing before you? Less frequent to the under-officials, and, if he did pronounce the words, it was only after first having learned the bearings of the matter. But from that day forth an apparition of the dead official ceased to be seen. Evidently the prominent personage's cloak just fit his shoulders. His dragging cloaks from people's shoulders have been heard of at all events. But many active and solicitous persons could by no means reassure themselves, and asserted that the dead official still showed himself in distant parts of town.

There were 50 participants (27 male, 16 female, 7 other/undisclosed), excluding an additional eight (six male, two female) who spent less than seven minutes on the test. They were recruited on online forums, primarily Reddit, in late 2021 and early 2022. As it was found hard to find enough participants, and many complained that the story was too long, for later participants the story was shortened from about 4000 words to 3000, and most post-story questions, except for the last two, were asked in the middle of the story to maintain concentration and measure reading speed better. The participants were also offered £5 Amazon vouchers, subject to paying enough attention to the study, which was measured by how long they spent on different sections of the study. It was thought that the request for written comments and the question on whether the protagonist really died or not could have been helpful for analysing this as well, but turned out not to be necessary, as those trying to game the system did a very poor job at it, trying to do the study extremely quickly. However, most eligible participants opted not to ask for a voucher even when they would have qualified for one, and only 12 were given out. It was indeed found that the vouchers appeared not to increase participation in the study, but perhaps even reduce it by making the study appear less interesting in its own right.

As in the previous study, each participant took a 10-item FFM questionnaire (Rammstedt & John, 2007) where every question is given a score from 0 to 4, and the total score for each trait is scaled in a normalised range from 0 to 1. They were also asked about their English reading skills, age and gender. They were then presented with one of the different versions of *The Cloak*; the original style versions of the story were also presented as controls. Therefore, there were 12 different versions in total. The version of the story each user got was selected by random.

In the final part, they were asked questions about the experience:

- How do you find the coherence of the language in the story?
- Apart from coherence, how do you find the language in the story?
- Do you think the version you have seen was altered by AI?
- How closely do you relate to the protagonist?
- How did you find the story?
- In the version you saw, did Akaky really die?

Results

Though it was found a not very helpful exercise to use automatic metrics due to the futility of using them to evaluate text style transfer, as discussed in II.7.3, for the sake of completeness, a few metrics were used anyway to measure similarity and readability. The cosine similarity of the beginning fifth of the story was tested with the Python library Sentence-Transformers[6], finding that the Shakespeare model got a score of 0.8757, Lyrics 0.8404 and Youth 0.8180. Readability was tested with Flesch Reading Ease (Kincald et al., 1975), in which the higher the score on the 100-point scale, the easier it is to understand the document. For the original version, this was 62.31, for Shakespeare 87.76, for Lyrics 80.82 and for Youth 79.8. Nevertheless, like such measures tend to do, this tells only one side of the story, in this case, how many words there are per sentence and how many syllables in a word.

A simple measure for checking the quality of the language would be to simply count the number of errors in the text, which was done looking at the whole story on Microsoft Word, using its Spelling and Grammar function, and going through all of the corrections to see that they were indeed clear and obvious errors, ignoring spaces before closing quotes, which was a common but minor problem, and language that could be considered a question of style, such as “gonna”. No such errors were found for the original. 12 were in Shakespeare (excluding how there often was capitalisation after commas), 18 in Lyrics and 16 in Youth. This could be considered decent performance in a 4000-5000-word story, though it did not account for errors a human would notice, but Word doesn’t. Furthermore, this does not consider how the youth literature model, unlike the other ones, turned out to have rather poor named entity recognition. As a result, some of the names of the places and characters did not remain consistent throughout the story. This was dealt with by having the names manually fixed to maintain readability in this language style version.

In the end, 14 people got the original style version, 12 got Shakespeare, 11 got Lyrics and 13 got Youth. The versions using the original language were less likely to be thought to have been modified by AI, but were still thought so by average. In fact, some people wrote comments complaining the AI had turned the language into an unreadable mess, even though they were reading the original version, in which we had not spotted any errors. It was indeed found more coherent on average than the modified versions, but still had mediocre scores. The version finetuned on song lyrics was considered good language, almost equal to original, even if less coherent. The original was also found more relatable and as having a better story.

Table 9: Scores by version, scale 0-1.

Version	Confidence	Coherence	Style	Relating	Story

[6] https://www.sbert.net/

	this was AI				
Original	0.571	0.571	0.661	0.464	0.696
Shakespeare	0.917	0.229	0.375	0.354	0.354
Lyrics	0.773	0.25	0.568	0.295	0.455
Youth	0.808	0.25	0.461	0.308	0.327
Kruskal-Wallis p-value	***0.015***	***<0.001***	***0.046***	***0.260***	***<0.001***

Table 9, seen above, displays the average scores on the scale from 0 to 1 for each version of the story. For example, when asked “Do you think the version you saw was altered by AI?”, the user’s choices were “Definitely”, “Probably”, “Not sure” “Probably not” and “Almost certainly not”, which gave the scores of 1, 0.75, 0.5, 0.25 and 0, respectively. The highest score here was 0.917 for Shakespeare, indicating a generally universal agreement that it was AI-altered, and the lowest was 0.571 for the original, indicating that on average, the users were not sure, but slightly inclined to think it was AI-altered. The other scores are reflected the perceived quality of the story, “coherence” being the answer to “How did you find the coherence of the language in the story?”, with 1 indicating “Great” and 0 “Awful”, “style” being the answer to “How did you otherwise find the language in the story?”, relating the answer to “How closely did you relate to the protagonist?”, and “story” related to “How did you find the story?”. The users were also asked “In the version you saw, did Akaky really die?”, but this related to the plot version, not the style version, and so was not included in this table. We knew of course whether Akaky died in the version they saw anyway; the point of the question was to check their comprehension of what happened. The table also displays the Kruskal-Wallis p-value at the bottom, indicating whether the differences between the groups are statistically significant. At p <0.05, we find they are all significant, except for the

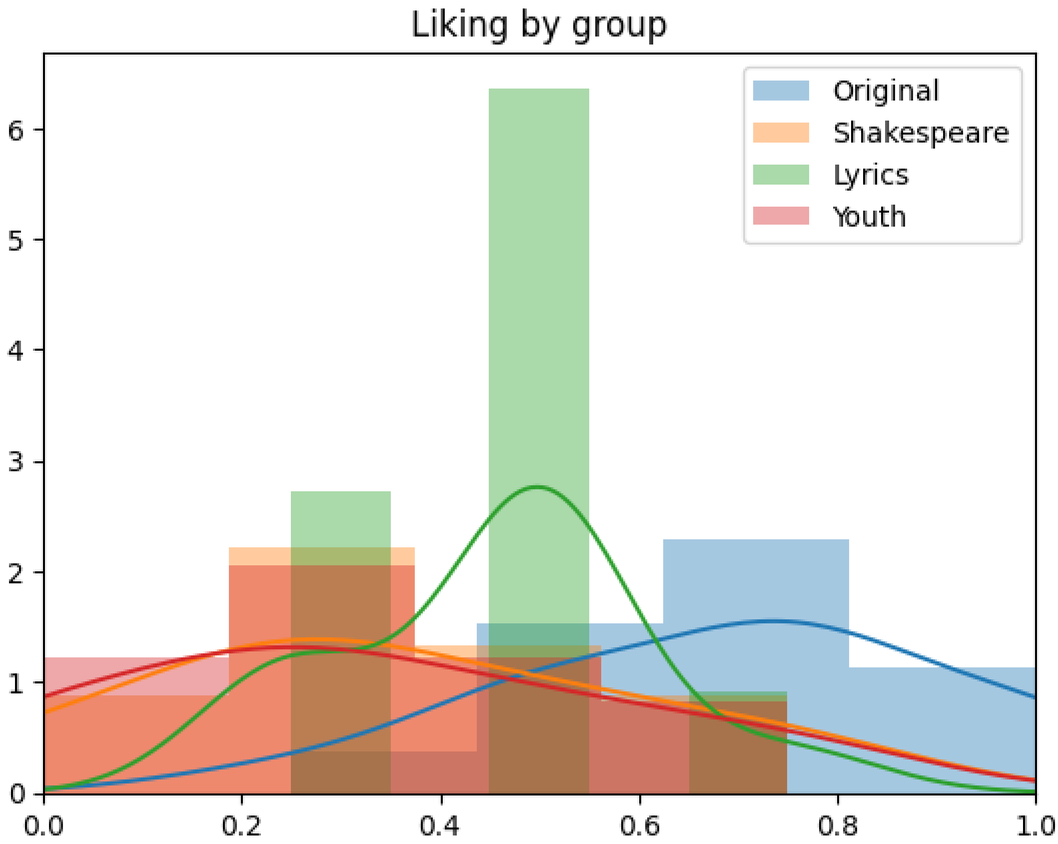
“Generally speaking, I could understand what was going on but hoo boy, could I tell the story was altered by a bot. For example, it started off in third person then for some reason, transitioned to first person at some point? There was an emotional distance from the story that was very weird.”
– A participant who read the Shakespeare version

relating part, where indeed the differences between the different groups were not large, though the original version saw slightly higher scores than the other ones.

> "Unfortunately, the narrative was so full of grammar errors that it became largely unreadable. I think I saw something in the last paragraph I skimmed that indicated the protagonist survived the whole thing, although by that point the plot itself was no longer coherent."
>
> A participant who read the original, unaltered version

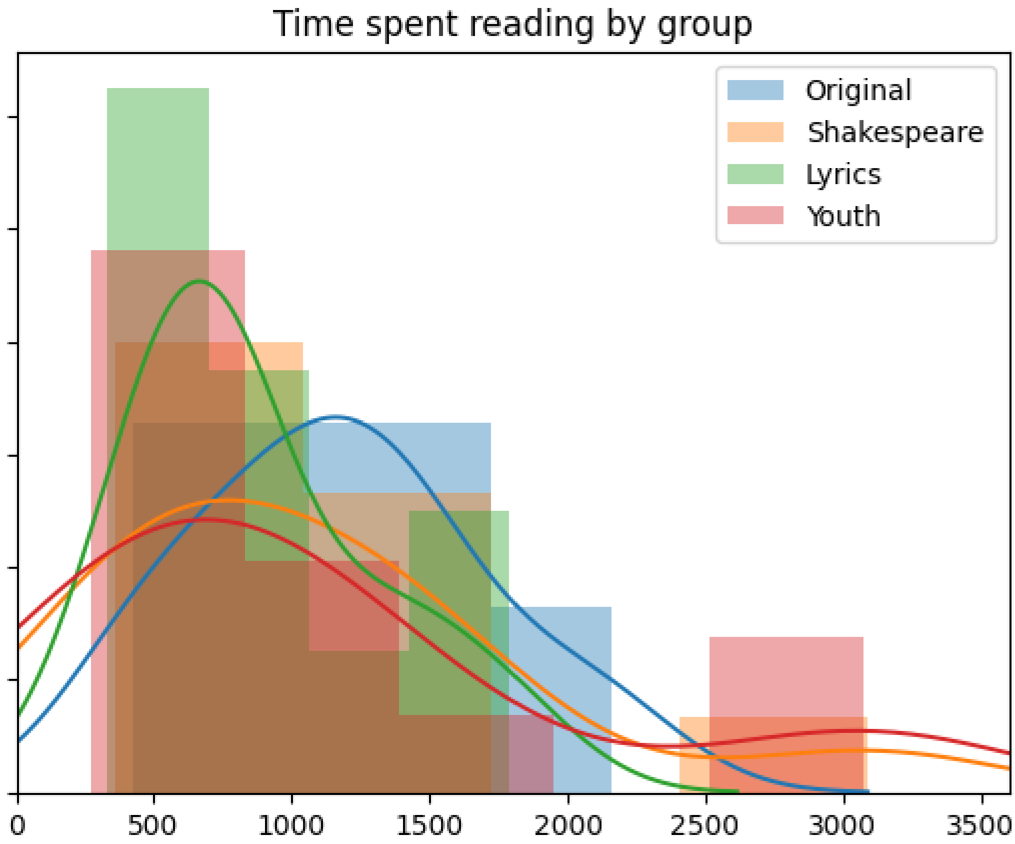
As in the previous study, the Shapiro-Wilk test showed that the personality data was not normally distributed, and therefore it was found better to use primarily Spearman correlations, though Pearson correlations were also checked for. This was done when comparing the results of the questions to each other. For example, finding the language coherent was very closely correlated with finding the language good (ρ=0.693, P<0.001), even though the exact question for the latter was "Apart from coherence, how do you find the language in the story?", indicating the issues shouldn't be so closely related. Relating with the protagonist was also related to finding the language coherent (ρ=-0.246, P=0.085), though this wasn't found statistically significant. Enjoying the story had a negative correlation with considering it to have been edited by AI, though this didn't reach significance (ρ=-0.255, P=0.073).

Figure 4: The ratings for liking the story by story version group. The bars display the number of users giving the rating shown on X axis; the lines show the moving average. Note that different groups had different numbers of participants.



For native English speakers, the average reading time was 1132 seconds; non-native speakers, it was the much higher 1467 seconds, excluding one outlier. The median time spent reading was the highest for the original version (1145 seconds, 5043 words for ambivalent version), followed by Shakespeare (883 seconds, 4300 words for ambivalent version), but shorter for Youth (738 seconds, 4843 words for ambivalent version) and Lyrics (722 seconds, 4386 words for ambivalent version). This adds up to reading speeds of 264 word per minute for original, 292 for Shakespeare, 394 for Youth and 364 for Lyrics, which could be taken as indications of their ease of reading. However, faster reading speed could also be an indication of lower interest in the story; the Spearman correlation between reading time and enjoying the story was 0.338 (P=0.016), indicating those who spent longer on the story enjoyed it more. Interestingly, Pearson correlation between them would be just -0.040, with a p-value of 0.787, but this was indeed considered a less appropriate measure, as the Shapiro-Wilk tests showed both the data as not normally distributed. However, these time measurements could be skewed by the fact that the story was later cut slightly shorter to attract more participants and to maintain their attention better. Nevertheless, for the later, shorter versions, the number also includes the time spent answering questions in the middle of the story, which should roughly balance it out.

Figure 5: Time in seconds spent reading the story by version. The number of users spending the time shown in X axis is displayed as bars; lines show moving average. Note that different groups had different numbers of participants.



There were no notable differences with gender, but age saw many: older participants found the language less coherent (ρ=-0.411, P=0.003) and otherwise worse (ρ=-0.468, P<0.001), and liked the story less (ρ=-0.412, P=0.003). Age also correlated with Emotional Stability (ρ=0.287, P=0.045), which appears to be a common finding (Donnellan & Lucas 2008). There were also strong intercorrelations between many personality trait pairs, as is to be expected: Extraversion-Agreeableness (ρ=0.289, P=0.044), Agreeableness-Conscientiousness (ρ=0.341, P=0.017), Agreeableness-Emotional stability (ρ=0.413, P=0.003), Conscientiousness-Emotional stability (ρ=0.401, P=0.004) and Conscientiousness-Openness to Experience (ρ=0.294, P=0.040).

36 of the participants were native English speakers, while 14 were not. Out of native speakers, the better they considered their reading skills, the more likely they were to consider the text was edited by AI (ρ=0.38, P=0.022). They were also less likely to relate to the protagonist, though this wasn't quite significant statistically (ρ=-0.309, P=0.067). No such

> *“I enjoyed reading the story, even though the build up was a little garbled, the ending was followable.”*
>
> - *A participant who read the Youth version*

effects were seen in non-native speakers. Instead, the more fluent they considered their English reading, the less coherent they appeared to find the language, though this was also just outside being statistically significant (ρ=-0.500, P=0.068) despite the strong correlation, but the non-native participant numbers were indeed low at just 14.

The versions with a happy ending were the least liked, though this was not statistically significant (score 0.41/1 for happy ending; 0.51 for sad; 0.5 ambiguous; Kruskal-Wallis p-value=0.369). No clear correlations were found with personality and the ending. In fact, by some strange coincidence, very few people got the ambivalent ending, with just one such case in the Shakespeare group, and four in the Lyrics group, and none in the others. It was notable that the participants had some difficulties telling whether Akaky had really died – this wasn’t surprising however, given the ambiguity of his return as a ghost, and how one could miss what really happened by skipping just a couple of sentences, and of course the fact that one ending was intentionally ambiguous. They were asked about this with four available options: that he died, that he didn’t, that it was left ambiguous, or that they weren’t sure. In the original version, 10/14 got it right and 3/14 the wrong way around, and one mistakenly thought his death was ambiguous. In the Shakespeare version, it seemed the readers struggled to understand what happened, as the answers seemed quite random, and only 3/12 got the answer right. Similarly in the Lyrics version, only 3/11 got it right. In the Youth group, 6/13 had it right, and on the 7 occasions where he died, 5 had it right, while in the 6 occasions where he didn’t die, just one had it right, and one the wrong way around, while two thought it was ambiguous, and two weren’t sure. Overall, only 22/50 had his death, the lack of it, or the ambiguity of it right, but just 8/50 had it the wrong way around, with the other 20 having different kinds of uncertainty over ambiguity.

The participants had very high average scores for Openness (0.8), which could have been encouraging them to participate in the study in the first place. Their other FFM traits averaged nearer to the middle (Extraversion 0.38, Agreeableness 0.57, Conscientiousness 0.55, Emotional Stability 0.40). This could have made it harder to find correlations with Openness. However, it appeared to be correlated with liking the story, though significance was only achieved with Pearson correlations, which we decided not to use, as discussed above; with Spearman, liking the story overall (ρ=0.247, P=0.083) and the Shakespeare version specifically (ρ=0.575, P=0.050) fell just outside statistical significance. Openness also appeared to be slightly correlated with liking the protagonist, though negatively so, which could have been caused by the protagonist appearing very much the opposite of open to experience. However, this was statistically significant only with the lyrics (ρ=-0.647, P=0.023) and Shakespeare (ρ=-0.634, P=0.036) versions.

> "Lots of typos and words that dont exist in the english language, hadn't I known NLP was involved I'd have guessed the story to be just a very bad translation from russian. I'd assume that that readers with a reading comprehension below native level would have difficulty making sense of it."
> - A participant who read the Lyrics version

Finally, and intriguingly, Extraverted people appeared to like the Shakespeare version, with a clear correlation between Extraversion and liking it (ρ=0.658, P=0.020). Therefore, though our previous study (Chapter IV) found that Extraverted people prefer less formality and Introverted people more formality, this study gave some indication to the reverse, though there could be many reasons for this, to be discussed below.

Discussion

Personalising narratives through NLP is a little-explored field with many prospects, and this study explored just one angle of it, using psychological models with text style transfer. In the previous chapter, it was found that the Five-Factor Model was highly useful in exploring a person's preferences in literature, and particularly for adapting the protagonist. However, this time, we focused on adapting the language and not the characters, which led to less conclusive results. The previous finding of less Extraverted people preferring more formal language should still apply however, but could not be replicated here, perhaps because of varying quality of different versions, and it being less straightforward which versions were more or less formal.

One notable aspect is the unexpected preference of Extraverted individuals for the Shakespearean style, contrary to the anticipated inclination towards less formal language. This deviation from previous findings raises questions about the nuanced relationship between personality traits and writing styles. It prompts speculation on whether Extraverts, known for their sociable nature, find allure in the expressive and dramatic qualities of Shakespearean language, even if it entails unconventional punctuation and capitalisation. It is possible that Extraverts, unlike Introverts, did not mind the fact that the text was full of unusual capitalisation and punctuation, possibly caused by having been trained by verse texts.

The correlation between Openness and enjoyment of the story, particularly in the Shakespearean and Lyrics versions, could suggest literary experimentation can captivate individuals with higher Openness scores, as has been found before in FFM research (McCrae & Costa, 1989). It could very well be that individuals more receptive to novel experiences and unconventional narratives find enjoyment in the linguistic creativity of these styles.

The age-related disparities in language perception and story enjoyment among participants introduce a layer of complexity. It prompts speculation on whether the observed differences are rooted in varying reading preferences shaped by generational influences, reading habits, or perhaps, scepticism towards AI-driven adaptations. Different generations have had different experiences and associations with styles like Shakespearean English (Pennebaker & Stone, 2003). Older readers expressing less satisfaction may signal a resistance to changes in traditional writing styles or a potential wariness of technology altering literary experiences. Perhaps a question on the attitudes of the participants towards AI would have been helpful. This could have been done at the beginning of the study, so as to reduce the effect of the experience they had with the story on the answer. On the other hand, asking such questions at the beginning could also affect their attitude to the story, acting as a sort of priming effect. Therefore, there is also an argument to be had in favour of focusing on the present experience rather than preconceptions.

Nevertheless, there are of course ways in which knowing more about a user’s preferences would be helpful. Another, a perhaps better way of adapting language for the sake of personalisation might be to identify the reader’s favourite authors rather than their personality type. Moreover, there is also the question of acquiring data on the users; readers might be more willing to share their favourite authors rather than doing personality tests. Such use could be combined with the use or recommender systems for suggesting non-personalised books to read, with the user's reading history in turn helping with personalisation, as well. Another suitable area of using personalisation, as well as an enjoyable way of figuring the person's preferences would be interactive narratives that personalise the generated text according to the user input. The most likely use would be with formality, figured from Extraversion or the formality of the input, or both. Perhaps these are avenues to explore next.

At the time of doing the study, it was based on the latest research, but at the time of writing this, it is already outdated, as ChatGPT and GPT-3.5 would be capable of doing style transfer far better. When we tried using them for this, asking ChatGPT to edit the same story used in the study, the results were very promising. The story did not need to be provided to it, as it was already familiar with it. Different authors' styles work well, with perfectly coherent language, and when changing the protagonist's personality, each version features a lesson for people with that kind of personality. Moreover, as people have become accustomed to ChatGPT and other large language models, the phenomenon of being sceptical of AI-edited language may have faded. It would be interesting to repeat the study using newer language models.

The study's outcomes underscore the importance of participant engagement and sustained attention throughout the research process. A more extensive participant base, spanning varied demographics and cultural backgrounds, might have provided nuanced insights into the intersection of personality traits and language preferences. Increased participation often leads to enhanced statistical power, allowing for more robust and reliable analyses. Higher participant numbers could have strengthened the significance of observed correlations, making it easier to draw definitive conclusions about the relationships between personality traits, reading experiences, and language preferences. Language fluency, especially among non-native English speakers, can significantly influence whether text is perceived as more challenging or less enjoyable (Dewaele, 2007). This introduced even more variability that would have required more participants to study.

The study also involved a prolonged reading session and intricate questions, demanding sustained attention from participants. Higher attention levels would likely result in more thoughtful and accurate responses, reducing the likelihood of misinterpretations and enhancing the overall quality of data collected. Some participants expressed concerns about the length of the story, potentially leading to reading it less carefully. This was in a way the reverse of the problem with the short story in the previous chapter, which was a bit short, not leaving much space to express the personality of the protagonist. Nevertheless, seeing the promising results of that chapter, it might be better to err on the side of shortness in a short story in any such future studies.

Higher participation and attention levels would not only enhance the external validity of the study but also contribute to a more thorough examination of the complex interplay between personality, language styles, and narrative experiences. Future research endeavours in this domain may benefit from strategies aimed at maximising participant engagement, ensuring a more representative and attentive study population.

Looking ahead, it would be worthwhile to consider these speculations as hypotheses for future studies. Exploring the intricacies of how personality traits intersect with writing styles and narrative elements could uncover deeper insights into the dynamics of personalised storytelling. Additionally, investigating the evolving perceptions of readers towards AI-driven language adaptation over time may provide valuable perspectives on the acceptance and integration of advanced technologies in literary experiences.

Conclusions

In this chapter, we explored the integration of text style transfer techniques with psychological profiling to create personalised narratives. The goal was to investigate how AI-driven language adaptation, informed by personality traits, could enhance the engagement and satisfaction of readers. The study employed a user-centric approach, combining AI-driven text style transfer with insights from the Five-Factor Model (FFM) personality assessment.

Our method involved a meticulous selection of a text style transfer approach, with a detailed exploration of existing literature and methodologies. The chosen method, based on the work by Krishna et al. (2020), demonstrated promising results in adapting writing styles while preserving the semantic content of the text. The implementation involved training models on a diverse dataset and modifying a short story to offer different endings manually. The process, although intricate, laid the foundation for assessing the effectiveness of AI-driven personalisation in narrative creation.

Results from the user study revealed interesting insights. Participants experienced different versions of the adapted short story, each aligned with specific writing styles and endings. The data analysis considered user perceptions, coherence, language style, relatability to the protagonist, and overall story satisfaction. Despite challenges such as participants sometimes mistaking the original for an AI-altered version, the study provided valuable insights into the nuances of AI-driven language adaptation.

Notably, the study found correlations between certain personality traits and preferences for specific writing styles. Extraverted individuals appeared to favour the Shakespearean style, which could be somewhat surprising given the previous study showing their preference for less formal style, but this could be explained by the quality of the text and Introverted people being concerned about typos and grammaticality more. Introverted readers might have more need to understand what they are reading than Extraverted users, who in turn might care more about the beauty of the language, but this is something that needs to be studied more. People high in Openness to Experience also enjoyed the Shakespearean and Lyrics versions, suggesting that people interested in novel experiences and unconventional narratives might be particularly fond of the linguistic creativity of these styles.

The study highlights the significance of participant engagement and sustained attention throughout the research process. A larger and more diverse participant pool could have enriched the understanding of how individuals respond to AI-driven text style transfer, capturing a broader range of perspectives. Increased participation offers enhanced statistical power, enabling more robust analyses and clearer conclusions regarding the relationships between personality traits, reading experiences, and language preferences. Moreover, prolonged reading sessions and intricate questions demand sustained attention from participants, potentially leading to more thoughtful and accurate responses. However, concerns about the length of the story may have affected participants' attention levels, potentially impacting the quality of data collected. Having to make changes to the study after difficulties finding participants and maintaining their attention was considered a problem, but a necessary action, and making sure that the results were still comparable was a priority. Future studies in this domain ought to aim to at maximise participant engagement, ensuring a wide and attentive study population.

As large language models have been taking huge leaps forward lately, the prospects of the approach to personalisation studied have only increased. With the use of newer models like GPT-4.0, a better quality of text style transfer can be achieved easily. This has in all likeliness also increased the interest and the responsiveness of the general public for such study; in this user study, there was still a lot of scepticism towards AI-based modification from the users, displayed in their complaints about the AI-modification even if they had read the original version. The age of participants also played a role, with older readers expressing less satisfaction with language coherence and overall story enjoyment. The attitudes towards AI-led personalisation may warrant more research.

Future personalisation systems might want to try out adapting language according to the reader's favourite authors rather than their personality type. A simple question on whether a reader would like to see a Shakespearean version or a Hemingwegian version could be a starting point. Personality could of course be studied at the same time, exploring the preferences of different personalities. This would also be to the benefit of recommender systems, not necessarily changing text, but searching for similar texts that already exist. However, more intriguing possibilities may lie in altering text, especially in interactive narratives, including games and chatbots, and the style of the text generated could also be made imitate the user's writing style, or what is expected to be their preferred style of writing. Indeed, predicting what type of writing the user may prefer is explored more in the following chapter.

Summary

This chapter has explored the integration of text style transfer techniques with psychological profiling to create personalised narratives, a key component of the broader thesis. The relevance of this chapter to the overall research lies in its investigation of how AI-driven language adaptation, informed by personality traits, can enhance reader engagement and satisfaction. By employing a user-centric approach, the study combined AI-driven text style transfer with insights from the Five-Factor Model (FFM) of personality, aiming to tailor narrative experiences to individual preferences.

The methodological choices made in this chapter, particularly the selection of text style transfer techniques based on the work of Krishna et al. (2020), are crucial to the thesis's aim of exploring the intersection of AI and personalised storytelling. The exploration of existing literature and the training of models on diverse datasets set the groundwork for evaluating the effectiveness of AI-driven personalisation in narrative creation. The findings from the user study, which revealed correlations between personality traits and preferences for specific writing styles, contribute to the thesis's goal of understanding the role of personalised content in enhancing user experience.

The chapter's focus on the challenges of participant engagement and attention during the study is also significant. The difficulties encountered in maintaining participant focus, particularly in the context of lengthy narratives and complex questions, underscore the importance of designing studies that not only capture data but also sustain user interest. This insight is critical for future research and applications in personalised content delivery, as it highlights the need for balancing narrative depth with user engagement.

Furthermore, the chapter's discussion on the evolving landscape of large language models and their implications for personalisation systems is highly relevant. The rapid advancements in AI technology, exemplified by models like GPT-4, suggest that the approach to personalisation studied in this chapter is becoming increasingly viable and effective. The chapter's conclusion that future personalisation systems might benefit from adapting language according to the reader's favourite authors, rather than solely based on personality traits, opens new avenues for research and application, aligning with the thesis's exploration of personalised storytelling's potential.

Chapter VI: Text-Based MBTI Recognition

Introduction

In the preceding chapters, it has been established that individuals with distinct personalities exhibit unique preferences in both reading and writing styles. What this chapter asks is how we could discern an individual's reading inclinations from their writing style, particularly by examining their textual contributions on platforms like social media. The chapter's focus on identifying a person's personality type through their writing style builds upon the foundational work established in earlier chapters, where it was demonstrated that personality significantly influences preferences in both reading and writing styles. This chapter extends that inquiry by exploring how these preferences can be predicted and utilised to tailor narratives, thereby enhancing user experience.

The objective here is not to fine-tune a language model to replicate an individual's idiosyncratic writing style – although this too would be an interesting prospect. Instead, the emphasis lies in delving deeper into understanding the person and their personality, using the Myers–Briggs Type Indicator (MBTI), and aiming to leverage this comprehension for broader applications. This is done by using machine learning for personality type classification based on samples of writing in pre-existing data sets, combining the Personality Café MBTI dataset (Keh & Cheng, 2019) and the MBTI9K dataset by Gjurković and Šnajder (2018), both including social media posts and the MBTI type of the user. The Personality Café dataset has often been used for similar studies, as seen in Chapter II.9., but, perhaps peculiarly, the MBTI9K dataset does not appear to have been used for these purposes anywhere near as much, if at all.

This was followed by the creation of an MBTI predictor using as input a passage of text to judge the writer's (or possibly the narrator's or character's) MBTI type. Different machine learning models for the classification were tried out on the Personality Café dataset, with a focus on producing balanced results so that all personality types can get a fair number of recommendations. This was difficult, as the data was heavily imbalanced, and the focus on balanced results was bound to make the overall performance seemingly worse. The same machine learning approaches were then used to create a predictor that was then used on narrator passages of about 100-200 words from free youth novels. The consistency of the predictor was then tested by repeatedly predicting the MBTI type of each passage, and then comparing the results of different narrator passages from the same novel. The aim was to see whether it was possible to figure out the personality of the narrator, so that the reader could be matched with novels where the narrator is similar to them, and see how effective this could be.

In deciding what personality framework to use, despite its less rigorous footing discussed earlier in Chapter II.4, the MBTI was found more interesting for this study than the FFM on grounds of its widespread popularity which would help with finding both data and interested users for any subsequent applications or uses. The available MBTI datasets were also found more helpful than the available FFM datasets. Furthermore, a major disadvantage of the MBTI, the dichotomous nature of the dimensions, which can cause a person whose score in a dimension is around the middle to get opposite results at different times, can actually be helpful in personalisation and recommender systems, as there might only be a number of different versions that can be presented to a user, rather than a continuous range of options. In fact, in our interactive narrative study, we effectively turned the FFM dichotomous, as we judged the users to be either high or low in a given trait, and did the personalisation accordingly to show alternative versions. Therefore, we decided to try out the MBTI this time. Nevertheless, it would be wise to keep expectations low about prediction performance, as text data may offer few cues about the MBTI. It is possible that other types of data, such as behaviour, image or audio could be more useful for predicting the MBTI. Nevertheless, similar problems would be faced when using any personality framework. Indeed, as discussed in Chapter II.9., personality prediction is one of the most difficult author profiling tasks in computational stylometry.

The central inquiry in the chapter revolves around determining a person's personality type through their writing. Though it focuses just on the style of language here, this understanding of personality surpasses the scope of adapting the writing style alone; it extends to personalising elements such as characters and plot, as well. As seen in previous chapters, people may have different preferences for plots and character personality depending on their own personality. Several potential avenues emerge from this exploration. Firstly, there's the prospect of employing adaptable chatbots or interactive narratives, potentially integrated into gaming experiences. Secondly, there is the possibility of personalisation of non-interactive narratives, akin to the previous studies. Lastly, the avenue of recommender systems opens up new possibilities.

In the realm of recommender systems, the incorporation of personality factors could manifest in various ways, concurrently and at different stages of the recommendation process. This multifaceted integration of MBTI personality types into recommender systems presents a captivating research avenue that could be pursued based on the results of the chapter. Such a recommender system could be based on not just the writing style, but some recommendations could be based on human judgement on characters' personality types. The Personality Database[7] has a database of fictional characters with their supposed Myers-Briggs personality types, labelled by fans. Using recommendations based on the writing style of the novel, and user-labelled character personalities, the results would be a multi-approach recommender system with very low initial processing time needed, as everyone with the same MBTI type could get the same results from the beginning, with no other data needed for the recommendations but the MBTI type, effectively skipping the cold-start phase. Once some usage of the recommender system by the users has been recorded, MBTI recommendations can also be complemented by collaborative filtering, as in Nadal et al. (2022). Such an approach could benefit from the wide popularity of the MBTI, drawing in users from the fan base and potentially gaining their interest and trust.

The relevance of this chapter to the overall thesis lies in its potential to deepen the understanding of how personality traits, as captured by the MBTI framework, can be used to create personalised content. By examining social media posts and applying machine learning models to predict MBTI types, the study takes a significant step toward developing systems that can adapt content to better match individual personalities. This approach not only complements previous chapters that explored personalisation through the Five-Factor Model (FFM) but also introduces a different personality framework, demonstrating the versatility and breadth of the thesis's exploration of personalised content.

The decision to utilise the MBTI is justified by its widespread popularity and the availability of extensive datasets, which facilitate the development and testing of machine learning models. This choice aligns with the thesis's objective to explore practical and widely applicable methods of personalisation that can engage a broad audience, leveraging existing interest in the MBTI.

Moreover, the exploration of MBTI recognition through writing not only seeks to refine the understanding of how text can reveal personality but also opens up new possibilities for content personalisation beyond mere text style adaptation. The chapter suggests that recognising personality through writing could lead to personalised narrative elements such

[7] https://www.personality-database.com/profile?pid=2&cid=12

as plot and character development, thus enriching the user experience in more profound ways.

In addition, the chapter introduces the potential application of MBTI recognition in recommender systems. By integrating MBTI personality types into recommender systems, the study proposes a multi-faceted approach that could improve the accuracy, efficiency and relevance of content recommendations. This aligns with the thesis's broader goal of enhancing user engagement and satisfaction through personalised content, offering a practical pathway for the implementation of personalised narratives in various digital environments, including gaming, interactive narratives, and non-interactive storytelling.

Method

Initially, the study was conducted using just the Personality Café dataset by Keh and Cheng (2019), but the results were vastly improved by combining its data with the MBTI9K dataset by Gjurković and Šnajder (2018).

Taking an initial look at the MBTI dataset by Keh and Cheng (2019), collected from Personality Café, it was noted that it includes 8675 users and a set of 50 posts from each of them, making it a total of 422,845 posts, along with their MBTI type, with each post limited to a maximum of 200 characters. The average length of a set of posts was 7235 characters. The longest set was 10,090 characters, while the shortest was just 57 characters. The dataset is freely available on Kaggle[8].

The MBTI9K dataset by Gjurković and Šnajder (2018), collected from Reddit, includes posts from 9252 users, along with their MBTI types. The researchers were contacted to gain access to the dataset, which included multiple files. The file used here was processed by the researchers to include all posts by a user, collated as if it were a single message. Each user's text was on average 204,086 characters long. The shortest text was 5203 characters, while the longest was an enormous 17,191,290 characters. Overall, the MBTI9K data represent 354,996 posts.

The data from both the datasets were combined, creating a dataset of 17,927 users. Like in many such datasets, the personality types in both these datasets are, for reasons unclear,

[8] https://www.kaggle.com/datasets/datasnaek/mbti-type

very heavily skewed towards IN (Introvert and iNtuitive) personalities, with thousands of users each, while ES (Extravert and Sensing) types only have dozens each. The number of users in each personality type was somewhat similar in both the datasets, with some variation.

Table 10: User numbers by personality type. First line Personality Café, second MBTI9K, third total. Colours indicate higher (green) or lower (red) numbers.

ISTJ 205 236 **441**	**ISFJ** 166 134 **300**	**INFJ** 1470 1023 **2493**	**INTJ** 1091 1837 **2928**
ISTP 337 445 **782**	**ISFP** 271 161 **432**	**INFP** 1832 1070 **2902**	**INTP** 1304 2313 **3617**
ESTP 89 88 **177**	**ESFP** 48 65 **113**	**ENFP** 675 605 **1280**	**ENTP** 685 624 **1309**
ESTJ 39 53 **92**	**ESFJ** 42 34 **76**	**ENFJ** 190 206 **396**	**ENTJ** 231 358 **589**

Overall, the proportions of the personality traits were as follows:

Introversion (I) / Extraversion (E):	12,895 (71.9%) / 5032 (28.1%)
Intuition (N) / Sensing (S):	15,514 (86.5%) / 2413 (13.5%)
Thinking (T) / Feeling (F):	9935 (55.4%) / 7992 (44.6%)
Judging (J) / Perceiving (P):	7315 (40.8%) / 10,612 (59.2%)

For the data preprocessing, direct mentions of any MBTI personality type were removed, along with URLs, non-words, punctuation, and words shorter than 4 or longer than 29 characters, and converting everything to lower case. After this, the average post was 9857 characters long. However, as the MBTI9K post sets were much longer, and the individual MBTI9K posts were not separated, it was found that the best way to combine the data was to split all the post sets into sets of a maximum of 4000 characters, or slightly more if the split would otherwise have ended up in the middle of a word. Sets shorter than 3000 characters were excluded. This led to a dataset of 40,877 sets of text. Maximum limits of 500, 1000 and 2000 characters were also attempted but gave worse results in the classification task. Higher limits were not attempted, as that would have ended up cutting out a large proportion of the Personality Café data. Adding another similar dataset[9] was also tried out, but it was found too small, having just 41,700 individual posts, of which just 16,579 were longer than 15 words.

This way of splitting up the text, of course, led to some users ending up being counted multiple times, as they provided multiple text sets, while others ended up being removed. This might skew the results slightly, but it was expected that the effect would be small, and positives would outweigh the negatives, and indeed this seemed to be the case, as the results were improved.

[9] https://git.arts.ac.uk/tbroad/myers-briggs-comments-dataset

Table 11: User numbers by personality type after splitting up the data. First line Personality Café, second MBTI9K, third total. Colours indicate higher (green), lower (red) or medium (peach) overall numbers.

ISTJ 223 822 **1045**	**ISFJ** 145 309 **454**	**INFJ** 1575 2884 **4449**	**INTJ** 1208 6561 **7769**
ISTP 343 1252 **1595**	**ISFP** 246 481 **727**	**INFP** 1812 2411 **4223**	**INTP** 1453 7029 **8482**
ESTP 76 228 **294**	**ESFP** 35 153 **188**	**ENFP** 565 1242 **1807**	**ENTP** 636 1669 **2305**
ESTJ 44 99 **143**	**ESFJ** 48 94 **142**	**ENFJ** 160 488 **648**	**ENTJ** 241 1267 **1508**

Various tools from the scikit-learn (sklearn) library, as detailed by Pedregosa et al. (2011), were used. Firstly, the labels representing the MBTI types underwent a transformation into a numeric format, achieved through the application of the LabelEncoder from scikit-learn. This conversion is essential for compatibility with machine learning algorithms that require numerical input.

The textual content was subjected to preprocessing using the CountVectorizer from scikit-learn, converting the raw text into a numerical representation based on term/token counts. Concurrently, lemmatisation was performed to standardise words by grouping different inflected forms into their common base or root form. This standardisation is a crucial step to enhance the consistency and efficiency of subsequent analyses.

To further refine the data, a dual vectorisation strategy was employed using both CountVectorizer and TF-IDF Vectorizer. The objective was to represent the text in a vectorised form, where words appearing between 10% and 70% of the posts were retained. This selective inclusion aimed to exclude overly common or rare words, focusing on terms with moderate frequency that are more indicative of meaningful patterns. This also makes the data faster to process.

The dataset was subsequently split into two sets of variables, denoted as X and Y. X represented the TF-IDF representation of user posts, with each row corresponding to a user and each column representing a feature. On the other hand, Y encapsulated the binarised MBTI type, signifying the presence or absence of each MBTI factor for every user.

A distinctive approach was adopted in the analysis, wherein each MBTI factor was predicted individually. This strategy allowed for a nuanced examination of linguistic patterns associated with specific personality dimensions, potentially uncovering intricate relationships between language use and distinct aspects of personality.

In order to better interrogate the potential use of the data, we benchmarked a wide range of machine learning algorithms as follows: Random Forest, XGBoost, Stochastic Gradient Descent, Logistic Regression, KNN, Balanced Random Forest, RUSBoost and Easy Ensemble. These are approaches that are generally used for classification tasks such as this, and most of them have been used in previous studies on the same dataset. Extra focus was put on finding balanced approaches because of the imbalanced data and the aim produce diverse recommendations. Over- and undersampling were also tried together with some of the models that did not include it in the first place. This made a small improvement to the performance of some of them. However, for the sake of comparing the models directly, this was not included in the final results.

Table 12: A description of the models used.

Model	Description	Relevance
Random Forest (Breiman 2001)	An ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes for classification tasks.	Known for its robustness and ability to handle imbalanced datasets. It's an ensemble method that combines the predictions of multiple decision trees to improve overall accuracy
XGBoost (Extreme Gradient	A scalable and accurate implementation of gradient	Widely used in machine learning competitions for its efficiency and effectiveness. It sequentially builds

Boosting) (Chen & Guestrin, 2016)	boosting machines, designed for speed and performance.	weak learners (typically decision trees) and combines them to create a strong predictive model.
Stochastic Gradient Descent (Bottou, Curtis & Nocedal 2018)	An optimisation algorithm that updates the model parameters iteratively using a small subset of the training data.	Often used for training large-scale machine learning models. It's particularly useful for large datasets, and it's adaptable to various types of models, including linear models.
Logistic Regression (Hosmer, Lemeshow & Sturdivant 2013)	A linear model used for binary classification. It predicts the probability that an instance belongs to a particular category.	A simple yet effective algorithm for binary classification tasks. It's easy to implement and interpret, making it a common choice for baseline models.
K-Nearest Neighbors (Cover & Hart 1967)	A non-parametric, instance-based learning algorithm. It classifies instances based on the majority class of their k-nearest neighbours in the feature space.	Straightforward and intuitive. It's used when the decision boundary is expected to be irregular.
Balanced Random Forest (Chen, Liaw & Breiman 2004)	An extension of the Random Forest algorithm that incorporates strategies to handle imbalanced datasets.	Specifically designed to address class imbalance by adjusting the weights of individual trees, making it suitable for datasets where one class is underrepresented.
RUSBoost (Seiffert, Khoshgoftaar, Van Hulse & Napolitano 2010)	An ensemble learning method that combines Random Under-Sampling (RUS) with boosting.	Designed to handle imbalanced datasets by under-sampling the majority class and boosting the minority class. It aims to strike a balance between the classes.
Easy Ensemble (Liu, Wu & Zhou 2009)	An ensemble learning method that combines multiple classifiers trained on balanced subsets of the dataset.	Similar to other ensemble methods, is effective for imbalanced classification tasks. It aims to provide diverse classifiers to improve overall model performance.

Results

The outcomes of the analysis were output through a two-step process. Initially, the accuracy scores for each personality type dimension were computed using the capabilities provided by the scikit-learn (sklearn) library (Pedregosa et al., 2011). The accuracy score represents the overall correctness of the model's predictions and is calculated as the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances. It provides a general measure of the model's ability to make correct predictions across all classes. However, in the context of imbalanced datasets, accuracy alone may not be sufficient for a comprehensive evaluation, as it can be influenced by the majority class. Subsequently, the imbalanced-learn library (Lemaitre, Nogueira & Aridas, 2017) was leveraged to generate a comprehensive classification report, offering a detailed assessment for each class within the dimension. The classification report from the imbalanced-learn library incorporated a set of state-of-the-art metrics designed to evaluate the performance of classification models on imbalanced datasets. These metrics include:

1. Precision: Precision, also known as positive predictive value, quantifies the accuracy of positive predictions made by the model. It is calculated as the ratio of true positive predictions to the total number of positive predictions (true positives + false positives). A high precision score indicates a low rate of false positives.

2. Recall (Sensitivity or True Positive Rate): Recall measures the ability of the model to correctly identify all relevant instances, specifically the ratio of true positive predictions to the total number of actual positive instances (true positives + false negatives). A high recall score indicates a low rate of false negatives.

3. Specificity: Specificity, also known as true negative rate, evaluates the ability of the model to correctly identify instances of the negative class. It is calculated as the ratio of true negative predictions to the total number of actual negative instances (true negatives + false positives). A high specificity score signifies a low rate of false positives in the negative class.

4. F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives. It is particularly useful when there is an uneven class distribution.

5. Geometric Mean: The geometric mean is a measure that provides a balanced assessment of a classification model's performance across multiple classes. It is calculated as the square root of the product of the recall scores for each class. The geometric mean is particularly useful for imbalanced datasets as it considers performance in both majority and minority classes.

6. Index Balanced Accuracy of the Geometric Mean: This index is a combination of the geometric mean and balanced accuracy. It accounts for both sensitivity and specificity, providing a holistic measure of a model's ability to perform well across all classes, especially in the presence of imbalanced class distributions. It aims to strike a balance between the recognition of positive instances and the avoidance of misclassifying negative instances.

At the end of the classification report, the “sup” (support) value represents the number of actual occurrences of each class in the specified dimension, providing context for the performance metrics.

By considering these metrics collectively, the classification report provides a nuanced understanding of the model's strengths and weaknesses, facilitating a more informed interpretation of its performance in predicting personality type dimensions, particularly in scenarios involving imbalanced data.

The models specifically meant for imbalanced data – Balanced Random Forest (Chen, Liaw & Breiman, 2004), RUSBoost (Seiffert et al., 2008; 2010), Easy Ensemble (Liu, Wu & Zhou, 2009), and SMOTEENN (Batista, Prati & Monard, 2004) – were among the best for the minority classes, which was found more important than overall performance. Overall, focusing primarily on minority classes, the best results were with Easy Ensemble. It is a hybrid ensemble undersampling algorithm that combines the results from an ensemble of AdaBoost algorithms to make up for the problem that undersampling may discard lots of potentially useful information about the majority class. Nevertheless, the majority class precision was still far better than the minority class precision, though recall levels were similar. Some other models, especially KNN, had terrible performance in the minority classes, appearing to focus solely on the majority. In general, the overall accuracy rates were comparable to the other studies on the datasets used, discussed in Chapter II.9., which vary from ~60% to the less realistic claims of almost 100%, though focusing on this was not our aim. While the other studies did not typically even report the results for the different classes, focusing on this and achieving balanced results was the main point here. The results are shown in detail in Tables 11-18 below.

Table 11: Easy Ensemble performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 57.61%							
I	0.84	0.59	0.54	0.69	0.57	0.32	9438
E	0.25	0.54	0.59	0.34	0.57	0.32	2376
total	0.72	0.58	0.55	0.62	0.57	0.32	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 58.32%							
N	0.91	0.58	0.60	0.70	0.59	0.35	10259
S	0.18	0.60	0.58	0.27	0.59	0.35	1555
total	0.81	0.58	0.60	0.65	0.59	0.35	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 66.48%							
F	0.51	0.63	0.67	0.56	0.65	0.42	4129
T	0.77	0.67	0.63	0.72	0.65	0.43	7685
total	0.68	0.66	0.65	0.66	0.65	0.43	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 61.01%							
J	0.51	0.56	0.55	0.53	0.55	0.31	5345
P	0.60	0.55	0.56	0.57	0.55	0.31	6469
total	0.56	0.55	0.55	0.55	0.55	0.31	11814

Table 12: Balanced random forest performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Test Accuracy: 64.54%							
I	0.82	0.71	0.37	0.76	0.52	0.28	9438
E	0.25	0.37	0.71	0.30	0.52	0.26	2376
total	0.70	0.65	0.44	0.67	0.652	0.27	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 65.39%							
N	0.90	0.68	0.48	0.77	0.57	0.33	10259
S	0.19	0.48	0.68	0.27	0.57	0.32	1555
total	0.80	0.65	0.51	0.71	0.57	0.33	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 66.67%							
F	0.52	0.60	0.70	0.56	0.65	0.42	4129
T	0.77	0.70	0.60	0.73	0.65	0.43	7685
total	0.68	0.67	0.63	0.67	0.65	0.42	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 56.43%							

J	0.52	0.45	0.65	0.49	0.55	0.29	5345
P	0.59	0.65	0.45	0.62	0.55	0.30	6469
total	0.56	0.56	0.55	0.56	0.55	0.30	11814

Table 13: RUSBoost performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 57.24%							
I	0.83	0.59	0.51	0.69	0.55	0.30	9485
E	0.23	0.51	0.59	0.32	0.55	0.30	2329
total	0.71	0.57	0.53	0.62	0.55	0.30	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 49.98%							
N	0.89	0.48	0.60	0.63	0.54	0.29	10304
S	0.15	0.60	0.48	0.24	0.54	0.30	1510
total	0.80	0.50	0.59	0.58	0.54	0.29	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 64.75%							
F	0.51	0.63	0.66	0.56	0.64	0.41	4229
T	0.76	0.66	0.63	0.71	0.64	0.41	7585
total	0.67	0.65	0.64	0.65	0.64	0.41	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 54.10%							
J	0.49	0.55	0.53	0.52	0.54	0.29	5324
P	0.59	0.53	0.55	0.56	0.54	0.29	6490
total	0.55	0.54	0.54	0.54	0.54	0.29	11814

Table 14: KNN performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 39.50%							
I	0.85	0.30	0.78	0.44	0.48	0.22	9485
E	0.22	0.78	0.30	0.34	0.48	0.25	2329
total	0.72	0.40	0.69	0.42	0.48	0.23	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 35.04%							
N	0.91	0.28	0.81	0.43	0.48	0.22	10304
S	0.14	0.81	0.28	0.24	0.48	0.24	1510
total	0.81	0.35	0.74	0.41	0.48	0.22	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 36.18%							
F	0.36	1.00	0.01	0.53	0.08	0.01	4229
T	0.94	0.01	1.00	0.01	0.08	0.01	7585

total	0.73	0.36	0.64	0.20	0.08	0.01	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 45.38%							
J	0.45	1.00	0.01	0.62	0.08	0.01	5324
P	1.00	0.01	1.00	0.01	0.08	0.01	6490
total	0.75	0.36	0.64	0.29	0.08	0.01	11814

Table 15: Logistic regression performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 61.17%							
I	0.84	0.63	0.52	0.72	0.58	0.34	9485
E	0.26	0.52	0.63	0.35	0.58	0.33	2329
total	0.73	0.61	0.55	0.65	0.58	0.33	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 63.90%							
N	0.91	0.65	0.53	0.76	0.59	0.35	10304
S	0.18	0.53	0.65	0.27	0.59	0.35	1510
total	0.81	0.64	0.55	0.70	0.59	0.35	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 67.60%							
F	0.54	0.60	0.72	0.57	0.66	0.43	4229
T	0.76	0.72	0.60	0.74	0.66	0.44	7585
total	0.68	0.68	0.64	0.68	0.66	0.43	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 56.92%							
J	0.53	0.46	0.66	0.49	0.55	0.30	5324
P	0.60	0.66	0.46	0.63	0.55	0.31	6460
total	0.57	0.57	0.55	0.57	0.55	0.30	11814

Table 16: Stochastic Gradient Descent performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 60.12%							
I	0.85	0.61	0.55	0.71	0.58	0.34	9485
E	0.26	0.55	0.61	0.35	0.58	0.33	2329
total	0.73	0.60	0.56	0.64	0.58	0.34	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 54.66%							
N	0.91	0.53	0.66	0.67	0.59	0.35	10304
S	0.17	0.66	0.53	0.27	0.59	0.36	1510
total	0.82	0.55	0.65	0.62	0.59	0.35	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 69.23%							

F	0.58	0.52	0.79	0.55	0.64	0.40	4229
T	0.75	0.79	0.52	0.77	0.64	0.42	7585
total	0.69	0.69	0.62	0.69	0.64	0.41	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 57.10%							
J	0.53	0.50	0.63	0.51	0.56	0.31	5324
P	0.61	0.63	0.50	0.62	0.56	0.32	6490
total	0.57	0.57	0.56	0.57	0.56	0.32	11814

Table 17: XGBoost performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 78.77%							
I	0.81	0.96	0.07	0.88	0.26	0.08	9485
E	0.33	0.07	0.96	0.12	0.26	0.06	2329
total	0.71	0.79	0.25	0.73	0.26	0.07	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 86.41%							
N	0.88	0.98	0.06	0.93	0.24	0.06	10304
S	0.32	0.06	0.98	0.10	0.24	0.05	1510
total	0.81	0.86	0.18	0.82	0.24	0.06	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 68.96%							
F	0.58	0.46	0.82	0.51	0.61	0.36	4229
T	0.73	0.82	0.46	0.77	0.61	0.39	7585
total	0.68	0.69	0.59	0.68	0.61	0.38	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 55.05%							
J	0.50	0.46	0.63	0.48	0.54	0.28	5324
P	0.58	0.63	0.46	0.60	0.54	0.29	6490
total	0.55	0.55	0.53	0.55	0.54	0.29	11814

Table 18: Random forest performance per dimension.

	pre	rec	spe	f1	geo	iba	sup
I/E: Introversion (I) / Extraversion (E) Accuracy: 79.33%							
I	0.80	0.98	0.03	0.88	0.17	0.03	9485
E	0.27	0.03	0.98	0.05	0.17	0.03	2329
total	0.70	0.79	0.22	0.72	0.17	0.03	11814
N/S: Intuition (N) / Sensing (S) Accuracy: 86.73%							
N	0.88	0.99	0.05	0.93	0.22	0.05	10304
S	0.36	0.05	0.99	0.09	0.22	0.04	1510

total	0.81	0.87	0.17	0.82	0.22	0.05	11814
F/T: Feeling (F) / Thinking (T) Accuracy: 68.77%							
F	0.60	0.38	0.86	0.46	0.57	0.31	4229
T	0.71	0.86	0.38	0.78	0.57	0.34	7585
total	0.67	0.69	0.55	0.67	0.57	0.33	11814
J/P: Judging (J) / Perceiving (P) Accuracy: 55.72%							
J	0.51	0.35	0.73	0.41	0.50	0.24	5324
P	0.58	0.73	0.35	0.64	0.50	0.26	6490
total	0.55	0.56	0.52	0.54	0.50	0.25	11814

The Easy Ensemble model was then used to create a predictor that estimates the personality type of the writer of a text, or possibly a character speaking in it. We used it to predict the personality type of the narrator of youth novels. Two narrator passages each from 16 novels were picked, and the predictor was run three times over for each passage with different random seeds to check consistency, for a total of 96 predictions. This is how the personality types were distributed overall for the 96 predictions:

Table 19: The personality type predictions overall.

E 18	I 78
S 37	N 59
T 52	F 44
P 38	J 58

It is possible that the Introversion dominance, with 78 Introverted passages over 18 Extraverted passages, was caused by the fact that novels tend to be written in more formal language than internet comments, and Introverts also tend to write more formally. The imbalance of the dataset could also be showing through here. Nevertheless, this did not appear to happen with the other dimensions, as the imbalance in the dataset for N/S was even larger than for I/E, but smaller here, and the split in the other two dimensions was actually reversed here.

As noted above, the predictor was run three times over for each passage, and different random seeds were used on each round. As there are 4 MBTI dimensions, and 32 passages were used, a total of 128 dimension predictions were made on each round. On the second prediction run, 30/128, or 23.4%, of the predictions turned out different from the first round.

After the third run, 44/128, or 34.3%, of the predictions had not been the same through all three runs, meaning there is just a decent amount of consistency with the results, with 65.7% of them staying the same throughout the three runs. This suggests that the model is somewhat sensitive to randomness, and its predictions are not fully stable.

Next, the results of the different passages from the same novels were compared to see whether the results were consistent within the novels. All three prediction rounds, discussed above, were considered as if they were from different triples of passages. In other words, each round of three predictions for three passages was considered as if they were from the same novel, while the other rounds were treated as if they were from a different novel. This means that we were nominally analysing 48 novels, not 16. However, the passages were only compared within novels, not between novels. As a result, it was determined that I/E matched in 34/48 (70.8%) novels, S/N in 29/48 (60.4%), T/F in 22/48 (45.8%), and P/J in 25/48 (52%). Overall, this means a consistency of 110/192=57.3%. Should the predictions be completely random, we could expect a consistency of 50%. The p-value of the hypothesis test is 0.026, indicating statistical significance at $p<0.05$. This shows that the predictions are relatively consistent, often giving the same personality results for different passages from the novels.

Nevertheless, the consistency does not seem high enough to be particularly helpful in recommending novels, given what a complicated task it is in the first place, and personality and style of language are factors among many relating to what people would like to read. Therefore, clearly the results would have to be improved to consider this avenue of research, as it would be difficult to demonstrate useful results in a recommender system on this basis. The results for I/E were promising however, having a high consistency at 70.8%; the use of Extraversion levels just by themselves could end up being useful, especially given how Chapter IV demonstrated the usefulness of personalising language according to Extraversion levels.

Discussion

This chapter aims to understand users from their writing style, using the Myers–Briggs Type Indicator (MBTI) as a guiding framework. The study employs machine learning for personality type classification based on user-generated content from social media, combining datasets from the Personality Café MBTI dataset (Keh & Cheng, 2019) and the MBTI9K dataset by Gjurković and Šnajder (2018).

Despite the acknowledged limitations of the MBTI discussed earlier, such as its less rigorous foundation, the choice to employ it stems from its widespread popularity, availability of relevant datasets, and potential applicability in recommender systems. The bipolar nature of the MBTI dimensions, often considered a disadvantage, is viewed as advantageous in personalisation and recommendation scenarios.

The study combines datasets, showcasing a focus on addressing the imbalance in personality types. The integration of the Personality Café and MBTI9K datasets amplifies the richness and diversity of the data, providing a more comprehensive basis for analysis. The methodology involves preprocessing text data, vectorisation using tools from the scikit-learn library, and the application of various machine learning algorithms, emphasising a focus on balanced results.

The results, presented through accuracy scores and a comprehensive classification report, reveal the performance of different machine learning models. Notably, models tailored for imbalanced data, such as Easy Ensemble, demonstrate effectiveness in handling minority classes. The discussion acknowledges the challenges posed by imbalanced datasets and the importance of achieving balanced results, particularly for personalised recommendations.

Ultimately, the purpose of the study was to explore the viability of recognising MBTI personality from text for the potential personalisation purposes such as a recommender system for novels. It was found that it does appear possible, but that expectations of performance must be limited, with accuracy of ~60% per dimension in our attempts. However, the overall performance was limited by the fact that we focused on performance in minority classes in an imbalanced dataset. The imbalanced datasets was also the reason for why the models meant for imbalanced datasets were expected to be the best in the first place; with more balanced data, different approaches might work better. Nevertheless, given the data and the purposes, the approach was rather necessary; it was found important that the predictions would be balanced so that there would be roughly similar amounts of novels to recommend for each personality type. In considering the consistency of the predictions, it must be remembered that standard MBTI questionnaires also give varying results for large proportions of people at different times (Pittenger, 1993).

Previously in the landscape of text-based MBTI recognition, diverse methodologies and models have been explored by researchers, each contributing unique insights and approaches to the challenging task of predicting personality types from written content. The literature review in Chapter II.9. reveals a wealth of approaches, ranging from boosting, bagging, and stacking ensemble methods (Das & Prajapati, 2020) to the utilisation of XGBoost (Khan et al., 2020a), Convolutional Neural Networks (Sugihdharma & Bachtiar, 2022), and Long Short-Term Memory (LSTM) algorithms (Maulidah & Pardede, 2021), to mention just a few.

Gjurković and Šnajder (2018), in their own study, opted for multi-layer perceptron (MLP) classifiers, incorporating a variety of alternative text features such as word and character n-grams. Keh and Cheng (2019) used their dataset to introduce pre-trained language models, specifically BERT, to the task of MBTI recognition. The finding of BERT's accuracy above 70% in the task sparked further exploration by Santos and Paraboni (2022) and others. These studies underscore the effectiveness of leveraging contextual embeddings and pre-trained language models for enhanced performance.

While the studies contribute significantly to the field, concerns arise regarding the reproducibility, detail and realism of some of the reported results, particularly in highly imbalanced datasets. There is a need for more detailed statistics and replication information, understanding the importance of rigorous evaluation in imbalanced datasets, which means performance in all classes should be reported.

The diverse approaches and models discussed in the literature underscore the complexity of the task and the multitude of factors influencing predictive performance. The present study, with its focus on balanced results and integration of multiple datasets, adds a unique perspective to the ongoing exploration of text-based MBTI recognition. Future research could benefit from a collaborative effort to establish benchmark datasets, share detailed statistics, and foster a more transparent and replicable research environment in the domain of personality prediction from text.

It is notable that Easy Ensemble was the only model that had accuracy above 60% in the J/P dimension, confirming the old finding that this dimension is particularly hard to predict from text (Plank & Hovy, 2015; Lukito et al., 2016; Verhoeven et al., 2018; Choong & Varathan, 2021).

The exploration extends beyond mere classification accuracy, contemplating potential applications of text-based MBTI recognition. Adaptable chatbots, interactive narratives, personalised non-interactive narratives, and recommender systems open up avenues for future research. The incorporation of personality factors in recommender systems, combining writing style recommendations with human-labelled character personalities, would be a novel approach with potential real-world applications.

Such approaches could involve applying an MBTI predictor to narrator passages from youth novels, as demonstrated in the chapter. The results, while showing statistically significant consistency, prompt reflection on the complexities of predicting personality solely based on text. There are challenges in achieving high consistency, and further refinement is required to make this approach practically useful in recommender systems.

How could better predictions be achieved? One answer seems quite straightforward: with more data. There is oddly little data in the datasets used on Extravert and Sensing types, and MBTI datasets often seem to have such imbalance, not reflecting what the distribution of the different types is supposed to be, according to the MBTI assessments. Though the phenomenon seems ubiquitous, the cause of this doesn't appear to have been studied much, though it has been found that Extraverts tend to prefer offline modes of communication, and Introverts are more likely to prefer online communication (Goby, 2006). Perhaps communities revolving around the MBTI draw in more people of specific types due to specific types having specific interests, or perhaps some types are more respected within the groups, causing them to engage more with them, and perhaps encourage people to claim to be of those types. Whatever the reason, scraping personality boards could focus on the less represented groups. As noted by Gjurković and Šnajder (2018), there are very active Subreddits dedicated to different personality types; these could be scraped to get more data on personality types that feature less in the data.

With some better prediction results, it would be interesting to proceed with building an MBTI-based recommendation system for novels. A user study could be a useful further approach, though it might be even better to use pre-existing reader data. If an available database should exist including reader preferences, it would be great to complement it with MBTI profiles, perhaps based on their own writings. Something like a literature discussion board could be suitable for this. Then there could be analysis on the reading preferences associated with each MBTI personality type. For example, certain personality types may prefer fiction over non-fiction, mystery over romance, or classic literature over contemporary works. This would open up multiple avenues for an MBTI-based recommendation system for novels, featuring recommendations based on language style, personality of characters, plot, genre, themes, any sort of tags used in such a database – potentially any feature of a book, which could then be enhanced with collaborative filtering. The potential of such a system would be vast.

Conclusions

This chapter explored text-based MBTI recognition, aiming to understand the connection between language use and personality dimensions. The study focused on identifying patterns that could predict MBTI personality types using machine learning techniques. The study highlights the prominence of balanced modelling approaches and the importance of thorough reporting in imbalanced datasets. The main finding, that there is a statistically significant consistency in predicting narrator personalities, especially Extraversion, has implications for current research on recommender systems in literature, where leveraging personality types could enhance personalisation in book recommendations.

Considering the acknowledged limitations of the MBTI, such as its less rigorous foundation, the study justified its choice based on the instrument's widespread popularity, availability of relevant datasets, and potential applicability in recommender systems and other personalisation. The dichotomous nature of the MBTI dimensions, often considered a challenge, emerged as advantageous in personalisation and recommendation scenarios. the Personality Café and the MBTI9K datasets emerged as a strategic response to the imbalance in personality types, fostering a more comprehensive basis for analysis. Through preprocessing, text data vectorisation using scikit-learn tools, and application of various machine learning algorithms, the study emphasised the pursuit of balanced results.

The comparative analysis of machine learning models shed light on their performance, with models tailored for imbalanced data, particularly Easy Ensemble, demonstrating effectiveness in handling minority classes, which was found a priority for the sake of generating recommendations for everyone. The results, conveyed through accuracy scores and a comprehensive classification report, add more to the ongoing discourse in the field, particularly in its search for balanced results and reporting.

The discussion section further contextualised the findings within the broader landscape of text-based MBTI recognition, drawing on diverse methodologies explored by previous researchers. Concerns regarding the reproducibility and realism of reported results, especially in highly imbalanced datasets, prompts a call for more detailed statistics and replication information.

In contemplating the viability of recognising MBTI personality from text for personalisation purposes, this study acknowledges the achievable accuracy (~60% per dimension) and the pivotal role of models designed for imbalanced data. However, it cautiously noted that the performance might be further refined with more balanced data, pointing to an avenue for future research.

The implications of this research are twofold. Firstly, the potential of predicting narrator personalities opens avenues for personalisation in literature. Though it focuses just on the style of language here, this understanding of personality surpasses the scope of adapting the writing style alone; it extends to encompassing elements such as characters and plot. This could be used with adaptable chatbots or interactive narratives, personalisation of non-interactive narratives, or recommender systems for literature. Secondly, future research could explore the integration of personality factors into machine learning models, not just for MBTI prediction but also for enhancing user experiences in various applications. Similar approaches could be used with the FFM, or any other factors datasets used could involve. Furthermore, as Extraversion could be predicted quite reliably in this study, and as it is arguably the most important dimension in both the MBTI and the FFM, it alone could be enough for personalisation; the efficiency of personalising language was already demonstrated in Chapter IV, and, additionally, it could just as well be used for personalising characters, as well.

Summary

This chapter focused on the integration of text-based Myers-Briggs Type Indicator (MBTI) recognition with the broader aims of the thesis, specifically in exploring the relationship between language use and personality dimensions to enhance personalised content delivery. By employing machine learning techniques to analyse social media content from the Personality Café and MBTI9K datasets, the chapter sought to uncover patterns that could predict MBTI personality types based on writing style.

A key contribution of this chapter is its focus on addressing the challenge of imbalanced datasets, a common issue in personality prediction studies. The study's approach, which prioritised models tailored for imbalanced data, such as the Easy Ensemble, proved effective in achieving balanced results, particularly in predicting the Extraversion dimension. This emphasis on balanced modelling is crucial to the thesis's overarching goal of developing reliable and robust systems for personalising content based on personality traits.

The findings from this chapter are directly relevant to the thesis's broader objectives in several ways. Firstly, the consistent prediction of narrator personalities, especially Extraversion, lays the groundwork for enhancing personalised experiences in literature. By accurately matching readers with narratives that align with their personality traits, this research contributes to the development of recommender systems that can offer more tailored and engaging content. This ties back to earlier chapters, where the focus was on adapting narrative elements such as language style, plot, and character development to suit individual preferences.

Secondly, the chapter's exploration of MBTI recognition extends beyond mere text style adaptation, highlighting the potential for personalising a wider range of narrative elements, including characters and plot. This aligns with the thesis's aim of using AI-driven personalisation to enhance user engagement by creating more relatable and immersive experiences. The study's insights into the correlation between language use and personality also open up avenues for future research, particularly in integrating personality factors into machine learning models for various applications, from interactive narratives to chatbots.

Overall, Chapter VI contributes to the thesis by advancing the understanding of how personality traits, specifically as defined by the MBTI, can be leveraged for personalised content delivery. It reinforces the thesis's overarching aim of exploring AI-driven personalisation by demonstrating how personality recognition can be integrated into narrative and recommendation systems, ultimately enhancing user engagement and satisfaction.

Chapter VII: Discussion

This chapter engages in a comprehensive discussion of the research findings, addressing both the strengths and limitations of the studies while also exploring potential avenues for future research and development in the realm of interactive narrative personalisation and personality recognition.

1. Interactive Narrative Future Prospects

The interactive narrative in Chapter IV served as a valuable tool for approximating users' Extraversion and Emotional Stability, although it seemingly struggled to assess other personality traits, at least in comparison to the personality test results. Nevertheless, it was actually better at personalising the short story than the personality test, giving some clue it may have done even a better job at estimating personality than the personality test. It is crucial to highlight that the way the interactive narrative measured Need for Affect (NFA) differed from traditional testing methods but aligned with the emotional intensity preferences described by the NFA authors. This suggests that the common assumption of high NFA correlating with emotionally intense preferences might not hold true universally.

Similar studies were being conducted simultaneously, notably McCord, Harman & Purl (2019), who attempted three interactive narrative personality tests, which got similar results to our study, achieving convergent validity for some but not all traits. The performance for Extraversion was particularly consistent through their studies, and ours as well. They adopted slightly different approaches to their three interactive narrative personality tests, with some measuring willingness to choose a course of action associated with one trait over another, and some measuring willingness to choose a course of action associated with high levels of a trait over lower levels.

Overall, they found that several personality factors were consistently correlated, and while some factors weren't, they argued there is fairly strong evidence that a game-like measure could be used to accurately assess personality. Moreover, they find that game-like measure might be better than personality tests, as they increase engagement and implicit interest, preventing “faking and careless responding”, thus questioning whether their validity should be tested by comparing them to traditional measures, which may even be worse. Later, Harman and Purl (2022) confirmed that the results were indeed more repeatable than with personality tests. In Harman & Brown (2022), they also added illustrations, but found that this appeared to make no difference.

The accumulated evidence strongly supports the notion that interactive narratives emerge as a highly promising method for gauging personality traits. This approach holds considerable potential for many purposes, but especially in the realm of tailoring narratives to individual preferences. The inherent interactivity of these narratives makes them particularly engaging, providing a more captivating alternative to traditional personality tests. This becomes especially pertinent for individuals who already possess an inclination towards enjoying narratives. By incorporating gamification elements into interactive narratives, the engagement factor is further heightened. This not only makes the process of personality assessment more enjoyable but also aligns with the preferences of individuals inclined towards interactive and game-like content.

2. Personality Test Limitations

The exploration of personality traits within the interactive narrative study has prompted a deeper examination of the limitations associated with traditional personality tests. While personality tests have long been a standard tool for assessing individual traits, this thesis has shed light on several intriguing aspects that warrant consideration.

One of the primary challenges encountered in the user studies was the possibility of participants rushing through the personality test section. This issue points to a fundamental concern within the realm of personality assessment – the potential for participants to approach these tests with varying levels of engagement, care, or even self-awareness. In some cases, participants may not invest the necessary time and thought to provide accurate responses, leading to results that may not fully represent their true personality.

This phenomenon opens up a thought-provoking avenue of inquiry. Could interactive narratives, like the one used in the first study, potentially capture facets of an individual's personality more effectively than traditional personality tests? This question arises from the fundamental difference in how these two methods engage with participants.

Traditional personality tests typically present respondents with a series of abstract and sometimes ambiguous questions. These questions often require individuals to reflect on their own behaviours, preferences, and tendencies and provide responses that align with predefined personality dimensions. However, this approach has inherent limitations. Respondents may interpret questions differently, leading to varied responses. Additionally, the abstract nature of these questions can make it challenging for individuals to discern how their answers will be translated into personality trait assessments.

In contrast, interactive narratives immerse users in specific, concrete situations. These narratives guide participants through scenarios that elicit natural reactions, behaviours, and choices. As users navigate these interactive experiences, their responses are more likely to reflect their genuine inclinations and tendencies. The narratives provide a context that mirrors real-life decision-making, potentially offering a more authentic glimpse into an individual's personality.

Furthermore, interactive narratives possess a unique advantage in their ability to adapt dynamically to user choices. This adaptability allows for a personalised and tailored experience that can reveal nuanced aspects of personality. For example, a narrative may present a character with various moral dilemmas, and the user's choices in these situations can provide insights into their ethical values and decision-making processes.

As technology continues to advance and interactive storytelling becomes more sophisticated, the potential for using these narratives as a tool for personality assessment becomes increasingly intriguing. Such narratives could be designed to probe specific personality dimensions or traits in a nuanced and contextually rich manner, potentially surpassing the limitations of traditional personality tests.

In essence, the study in Chapter IV has opened the door to an exploration of how the dynamic, immersive, and adaptable nature of interactive narratives might offer new avenues for understanding and assessing the intricate facets of human personality. As researchers continue to investigate this area, we may witness the emergence of innovative methods that redefine how we approach the measurement and understanding of personality traits.

3. Personality and Fictional Behaviour

The use of interactive narratives emerged as a promising avenue for gauging aspects of users' personalities, particularly Extraversion and Emotional Stability. These digital storytelling experiences demonstrate an ability to capture and approximate these facets of an individual's personality reasonably well. However, when it comes to assessing other personality traits, the interactive narrative encountered limitations. Other studies found similar results, but with some variance in what traits could be assessed.

This intriguing intersection of personality and fictional behaviour raises compelling questions about the human psyche and our engagement with fictional worlds. One such question is whether individuals might be inclined to seek in fiction what they wouldn't necessarily pursue in their real lives. For example, could highly Agreeable individuals find themselves drawn to narratives that explore rudeness or confrontational behaviour within the safe and controlled context of a story?

The safe distance provided by fiction allows individuals to engage with challenging or morally ambiguous themes without personal consequences. It offers a form of catharsis, enabling readers to vicariously experience situations and emotions they might never encounter or express in reality. In this way, fiction becomes a playground for the imagination, a space where individuals can experiment with facets of their personality that remain dormant or unexpressed in their day-to-day lives.

In essence, the relationship between personality and fiction reveals that our engagement with fiction is not solely a passive act but rather a dynamic interplay between our intrinsic traits and the limitless possibilities offered by the world of storytelling. As interactive narratives and digital technologies continue to evolve, they may provide even more profound insights into this intricate relationship between personality and the fictional realms we explore.

The findings imply that the personality traits revealed through interaction with fictional narratives might differ from those displayed in real life. This raises the intriguing possibility that recommender systems should consider offering content that allows users to explore aspects of their personality that are not overtly expressed in daily life. This suggests the need for recommender systems to create dynamic user profiles that account for both real-world preferences and fictional inclinations, which could be discovered through interactive narratives. Such an approach would allow recommender systems to evolve alongside the user, adapting to changes in engagement patterns over time and providing more sophisticated and adaptable recommendations. For example, a highly Agreeable individual might be drawn to narratives that explore confrontational or morally ambiguous behaviour, providing a safe and controlled environment to engage with such themes. Recommender systems, therefore, could benefit from recognising this duality in user preferences, offering content that satisfies both the expressed and latent dimensions of the user's personality. Such an approach could result in a richer and more fulfilling user experience, as it caters to a broader spectrum of psychological needs.

The concept of fiction as a space for cathartic and exploratory experiences further enriches the implications for recommender systems. The ability of fiction to offer a safe distance for engaging with challenging or emotionally charged themes suggests that recommender systems should be attuned to the emotional landscapes users seek within narratives. By identifying when a user may benefit from such cathartic experiences – based on their interaction history with certain types of narratives – the system could recommend stories that offer emotional release or introspective exploration. This could significantly enhance the emotional resonance of the content suggested, making it more impactful and personally relevant to the user.

Relatedly, games offer diverse environments and scenarios, and player behaviour within these virtual spaces is influenced by a myriad of factors, including game mechanics, narrative structures, and social interactions. Additionally, players often adopt different personas or playstyles based on the game genre, leading to a dynamic and context-dependent manifestation of their personality. One of the central challenges is the dynamic nature of player behaviour. Individuals may exhibit different traits when engaging in various gaming experiences. For instance, a player might demonstrate strategic thinking and leadership skills in a cooperative multiplayer game while adopting a more explorative and adventurous approach in a single-player narrative-driven game. Furthermore, players may wish to exhibit a side of them within the safe environment of games and interactive narratives that they would not do in real life. Recognising and accounting for this variability is essential in developing accurate models that truly reflect the richness of player personality within the gaming context, without mistakenly assuming they would exhibit similar behaviour in real life.

Future research endeavours could harness the power of machine learning to refine models mapping in-game behaviour to personality. Machine learning algorithms can analyse vast datasets of player interactions, decisions, and gameplay patterns to identify nuanced patterns and correlations. This approach enables the creation of more sophisticated models that adapt to the diverse ways individuals express their personality within different gaming scenarios. However, machine learning models are often limited in explainability, which could severely affect their usefulness, especially in a research context.

In conjunction with machine learning, advanced data analysis techniques could play a crucial role in unravelling the complexities of in-game behaviour and personality mapping. Descriptive analytics can uncover patterns and trends in player behaviour, while predictive analytics can anticipate how certain traits might manifest in response to specific in-game stimuli. Prescriptive analytics can guide the development of personalised gaming experiences that resonate with individual players.

To enhance the accuracy of personality mapping, future research should also consider the impact of specific game design elements on player behaviour. Elements such as narrative choices, character interactions, and in-game challenges can elicit different facets of personality. Integrating these design elements into the mapping models can provide a more holistic understanding of the dynamic interplay between in-game experiences and real-life personality traits.

Incorporating user feedback into the research and development process is instrumental in refining models over time. Understanding how players perceive the alignment (or misalignment) of in-game behaviour with their actual personality traits adds a qualitative dimension to quantitative data. This iterative approach ensures that models continuously evolve to capture the evolving nuances of player behaviour and personality.

Bridging the gap between in-game behaviour and real-life personality requires a multidimensional and adaptive approach. Leveraging machine learning, advanced data analysis techniques, and a nuanced understanding of game design elements can pave the way for more precise models. As the gaming landscape continues to evolve, research endeavours in this direction contribute not only to personalised gaming experiences but also to a deeper understanding of the intricate relationship between virtual interactions and individual personality traits.

As interactive narratives and digital storytelling technologies continue to evolve, they could provide even more profound insights into the intricate relationship between personality and fiction. Recommender systems that integrate these insights could extend their application to various domains, such as product recommendations, social networking, or educational content. By tailoring suggestions to both the user's typical preferences and their exploratory interests, these systems could significantly enhance user engagement and satisfaction.

4. Text-Based Personality Assessment

Recognising personality from text emerged as a possible application for a novel recommender system for novels. However, it's important to temper expectations, as the study yielded approximately 60% accuracy per dimension in attempts to predict MBTI personality types. While models designed for imbalanced data, particularly Easy Ensemble, showed promise, the study focused primarily on minority classes within an imbalanced dataset. Consistency in predictions also presented challenges, echoing the variability observed in standard MBTI questionnaires.

While different kinds of studies based on personality predictions could be valuable, it might be prudent to await more robust results. Achieving improved accuracy might hinge on acquiring more data, particularly for personality types that were underrepresented in the dataset. This underscores the need to explore the causes of data imbalances within the MBTI datasets.

Scraping personality-related boards and forums, such as active Subreddits dedicated to different personality types, could be a strategic approach to obtaining more data on less represented personality types. Furthermore, considering the MBTI as a continuous spectrum rather than a binary classification might enhance accuracy. However, this approach introduces challenges, such as rendering old test results unusable and requiring participants to retake tests. Using the FFM might be a better option, especially given its more empirically laid foundations. However, it would likely require gathering a lot of data that might be harder to come by than MBTI data, given the FFM's lower popularity in wider audiences.

Integrating natural language processing techniques, sentiment analysis, and semantic understanding could contribute to more nuanced personality assessments from text. Advances in machine learning algorithms and the availability of large-scale datasets could potentially address the challenges observed in the study, leading to more accurate and reliable predictions. Exploring alternative data sources and refining methodologies in text-based personality recognition may unlock the full potential of this approach in developing personalised recommender systems for novels.

5. User Profiles and Recommender Systems

The concept of user profiles developed within the context of interactive narratives presents an intriguing opportunity for broader applications, particularly within the realm of recommender systems. While the study in Chapter IV primarily explored the use of the Five-Factor Model for personalising interactive narratives, the insights gained here could have far-reaching implications for enhancing the recommendation of narrative content across diverse domains.

Recommender systems play a pivotal role in modern digital platforms, assisting users in discovering content that aligns with their preferences and interests. These systems traditionally rely on user data, such as browsing history and explicit ratings, to make content recommendations. However, the incorporation of personality-based user profiles, as exemplified by this study, offers a novel approach to enhancing recommendation diversity and accuracy, especially when conventional user data is limited or lacks context.

One notable advantage of integrating personality-based user profiles into recommender systems is the potential to provide more nuanced and tailored content recommendations. Instead of relying solely on past behaviour, these systems could leverage users' personality traits, reading motivations, and narrative preferences to generate personalised recommendations. For instance, a user profile indicating a preference for introspective and emotionally charged narratives could lead to recommendations of novels that align with these traits.

Furthermore, the user profiles generated through interactive narratives could encompass a wide array of factors beyond the Big Five personality traits. Reading motivations, genre preferences, emotional responses to narratives, and even historical reading patterns could be integrated into these profiles, creating a holistic understanding of each user's narrative tastes. This comprehensive user profiling could contribute to a more fine-grained and accurate recommendation process.

The potential applications of these user profiles extend beyond the realm of literature and storytelling. While narrative content, such as books, movies, and interactive fiction, stands to benefit significantly from personalised recommendations, the principles underlying these user profiles could also be adapted for various domains. For instance, personalised recommendations could enhance user experiences in e-learning platforms by tailoring educational content to individual learning styles and preferences. In the realm of e-commerce, product recommendations could align with users' unique tastes and values, leading to more satisfying shopping experiences.

Addressing the issue of users seeking traits in narratives that don't align with their own personality presents a complex challenge for recommender systems. This proposed phenomenon, where individuals are drawn to narratives that allow them to explore facets of their personality they may not typically express, reveals a nuanced relationship between user preferences and content consumption. To effectively address this, recommender systems need to incorporate mechanisms that can identify and accommodate these divergent interests without reducing the accuracy or relevance of recommendations.

One approach to managing this issue would be to design recommender systems that could differentiate between stable, intrinsic personality traits and situational or exploratory preferences. This would involve developing a dual-layered profiling system where the first layer captures the user's consistent personality characteristics, such as those derived from models like the FFM, and the second layer captures more transient or exploratory behaviours. For instance, while a user might generally prefer content that aligns with their high Agreeableness, the system should recognise patterns where they occasionally engage with more confrontational or morally ambiguous narratives. To achieve this, recommender systems could leverage machine learning algorithms that analyse not just the content consumed but the context in which it was consumed. By examining factors such as time of day, mood indicators (possibly inferred from interaction patterns), or even the user's historical engagement with different content types, the system could better understand when and why a user might diverge from their usual preferences. This context-aware recommendation engine would enable the system to suggest content that aligns with both the user's core personality traits and their situational explorations, thereby offering a more holistic and satisfying user experience.

Moreover, the system could incorporate adaptive learning techniques that adjust recommendations based on feedback from the user. For example, if a user consistently interacts with and enjoys content that does not align with their typical personality profile, the system could recalibrate to offer a more balanced mix of recommendations, incorporating both familiar and exploratory content. This adaptive approach not only enhances user engagement but also respects the complexity of human interests, which are not always neatly aligned with stable personality traits. The system could also benefit from integrating explicit user controls that allow individuals to indicate their current mood or specific themes they are interested in exploring at a given time. This feature would give users more agency in guiding the recommendation process, enabling them to seek out content that either aligns with their current state of mind or challenges their usual preferences. Such user-driven inputs could be particularly valuable in situations where the system's algorithms might otherwise default to safer, more predictable recommendations based on past behaviour.

Lastly, considering the potential for narrative experiences to serve as a form of self-exploration or catharsis, recommender systems should be designed to accommodate and even encourage this aspect of content engagement. By recognising that users may turn to fiction as a means of exploring unexpressed or repressed traits, systems can introduce content that provides a safe and meaningful way to engage with these themes. This could be particularly relevant in therapeutic or educational contexts, where narrative content is used to facilitate personal growth or emotional resilience.

6. Personality-Based Personalisation Findings

In Chapter IV, significant findings emerged from the exploration of personalised narratives, providing valuable insights into user preferences, personality traits, and narrative experiences. Personalising the protagonist based on users' FFM personality proved to be highly effective, regardless of whether the FFM score was derived from the interactive narrative or a traditional personality test. In fact, personalisation based on the interactive narrative seemed to produce stronger correlations with enjoyment. This suggests that the interactive narrative might provide a more accurate reflection of users' personalities compared to traditional tests. Additionally, Openness appeared to align with a preference for happier, less emotionally charged endings, suggesting its potential suitability for personalising story conclusions. The results underscore the utility of personalisation in narrative design and its implications for tailoring storytelling experiences to individual characteristics and preferences.

Extraverted individuals were found to prefer narratives with less formal language, while Introverts exhibited a preference for more formal language, aligning with the typical writing styles associated with each personality type. The effectiveness of adjusting language based on Extraversion levels was evident in improved user liking for the language and enhanced relatability to the protagonist, who also served as the narrator. This approach presents a promising avenue for future research to refine interactive narratives for enhanced personality capture and tailor narrative experiences to individual preferences. To separate the effects of personalising the language and personalising the protagonist, the protagonist's personality was not adjusted according to Extraversion, but given the other four factors had successful results with adjusting the protagonist, it is almost unimaginable that Extraversion couldn't do the same successfully, so this too would be a promising approach.

However, in Chapter V, the results on adjusting language presented a more complicated picture. While the study focused on adapting language rather than characters, as explored in the previous chapter, the results were less conclusive, possibly influenced by the varying quality of different versions and challenges in discerning formality. The study's findings regarding the correlation between Openness to Experience and story enjoyment were unsurprising, suggesting that individuals with higher Openness scores may be more receptive to creative and linguistically complex narratives. This is consistent with prior research that links Openness to a preference for novel and diverse experiences, including in literature and art (McCrae & Costa, 1989).

However, one significant limitation of the study lies in the variability of the text versions used to assess user preferences. The different versions of text, particularly those altered to match specific personality traits, exhibited varying quality, which may have confounded the results. For instance, the unexpected preference of Extraverted individuals for Shakespearean style – typically characterised by its formal and archaic language – was surprising given that Extraverts are generally associated with a preference for more accessible and less formal communication styles. This outcome raises questions about whether the preference was genuinely linked to personality traits or if it was influenced by the presence of errors or inconsistencies in the text, which tend to bother Introverted individuals more (Mairesse, 2007). Indeed, it has been noted that maintaining the quality and coherence of text while altering style is a significant challenge for natural language processing models (Hovy & Spruit, 2016).

Another limitation of the study is the reliance on personality traits as the primary basis for language adaptation, rather than considering users' favourite authors or genres, which might offer a more intuitive and effective approach to personalisation. The assumption that personality traits directly translate into preferences for specific writing styles may oversimplify the relationship between personality and literary taste. Research in media psychology suggests that personal preferences for authors or genres often reflect a complex interplay of factors, including past experiences, cultural background, and individual mood at the time of consumption, which may not always align neatly with personality dimensions (Oliver & Raney, 2011).

Reflecting on the study's outcomes, the importance of participant engagement and attention throughout the research process emerges. A larger and more diverse participant pool, spanning demographics and cultural backgrounds as well as different personalities, could provide richer insights. Enhanced participant engagement would contribute to a more thorough examination of the interplay between personality, language styles, and narrative experiences. Methods to achieve this could include using shorter stories, increased gamification, or doing it on location.

Finally, the study's reliance on NLP models that have now been surpassed is another limitation. The recent advancements in large language models suggest that more advanced models could offer significant improvements in this area, enabling more accurate and nuanced style transfer that better aligns with user personalities. Future research could explore the use of these advanced models to refine the personalisation process, potentially leading to more effective and satisfying narrative experiences for users.

Looking forward, the study can help with forming hypotheses for future research endeavours. Investigating how personality traits intersect with writing styles and exploring evolving reader perceptions of AI-driven language adaptation over time offer promising avenues for deeper insights into personalised storytelling dynamics and the integration of advanced technologies in literary experiences.

7. Alternative Frameworks

The attempt to combine the NFA with the FFM was not a perfect success but could be considered a first step to the creation of new frameworks for the study of preferences in fiction. Building new frameworks to study preferences in literary fiction, whether based on the FFM or alternative approaches, involves considering a wide range of factors that can influence individuals' reading choices and enjoyment of fictional content. These factors should encompass both individual psychological traits and external contextual elements.

In the personalised narrative chapter, the emphasis is on creating dynamic and engaging narratives that respond to individual personality traits. The Five-Factor Model and the Need for Affect are given a score from 0 to 1 like factors or vectors. The way that the FFM was combined here with the NFA was effectively a factorial or vectorial model approach to user profiling. Meanwhile, using the MBTI resembles player type models or the stereotype approach, as there is a specific set of personality types. The user is initially assigned to a stereotype, and appropriate responses can be determined based on this categorisation. If the approach is used in gaming, during gameplay, the model can be adjusted to fit the individual player, moving beyond the initial stereotype. Examples of this approach relate the stereotypes to players' gaming profiles rather than their real-life characteristics, as seen in works by Yannakakis & Hallam (2007) and Thue et al. (2007). An MBTI-based recommender system that also uses collaborative filtering would also adapt to the use, but the difference is in how it is based on real-life characteristics.

The MBTI might not be as useful for studying reader preferences as the FFM, but considering its popularity, the most interesting prospect could be reader communities and discussion groups. Online communities and discussion groups that cater to specific MBTI personality types can serve as platforms for readers to share book recommendations, reviews, and insights based on their personality-driven preferences. It would be interesting to build frameworks that facilitate the formation of reader communities centred around the MBTI types, encouraging meaningful literary discussions.

One promising direction for future studies lies in the exploration of personalisation based on alternative personality frameworks or traits. While this work primarily focused on personalisation techniques derived from the FFM or the MBTI personality traits, there are numerous other personality models and dimensions that could be leveraged to enhance user engagement and immersion. For instance, researchers could investigate the potential of personalising narratives based on the HEXACO model, which includes Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) (Ashton et al., 2004). The HEXACO model is quite similar to the FFM, but not as widely used. It is possible this approach could provide a more comprehensive understanding of how personality influences narrative preferences and reactions, but differences to using the FFM would not be huge given the similarity of the trait models.

The Dark Triad traits (Machiavellianism, Narcissism, Psychopathy) could shed light on preferences for antiheroes or morally ambiguous characters. However, collecting such data on readers would likely be highly problematic ethically. The Dark Triad traits are inherently associated with socially undesirable behaviours, such as manipulation, self-centredness, and a lack of empathy. Consequently, the collection of data related to these traits involves probing into deeply personal and potentially distressing aspects of an individual's personality. Individuals identified with Dark Triad traits might experience distress upon learning that they are judged to possess these characteristics. Furthermore, personalisation based on Dark Triad traits could inadvertently reinforce negative stereotypes, biases or behaviour. For example, individuals identified as having high levels of Machiavellianism might be given large amounts of content that portrays manipulation or deceit, making them seem normal, thereby reinforcing these traits.

Other options could include player typologies or frameworks derived from them, such as Brainhex (Nacke, Bateman & Mandryk, 2011), discussed in II.5.3. However, pre-existing studies, such as those by Rogers, Kamm, and Weber (2016) and Busch et al. (2016), have highlighted the limitations of Brainhex in terms of empirical validation. These studies found that the typology, while conceptually robust, struggled to consistently predict player behaviours across diverse populations. This suggests that further refinement of the Brainhex model would likely be necessary before it could be effectively applied to personalisation. One potential avenue for improving its predictive power could involve combining it with other personality models or integrating real-time behavioural analytics to create more dynamic, adaptive profiles. Additionally, exploring hybrid frameworks that draw from both personality traits and player typologies could provide a richer, multi-dimensional understanding of user engagement. For example, integrating Brainhex with the FFM might reveal how personality traits interact with specific gaming motivations. This could lead to more nuanced personalisation systems capable of tailoring not only the content itself but also the mechanics of how players interact with and experience narrative-driven environments.

Player type models, typically used in the context of game design and player psychology, can be adapted and applied to the analysis of literature. These models categorise individuals based on their preferences, behaviours, and motivations in interactive experiences, such as video games. When applied to literature, they could offer valuable insights into how readers engage with and interpret narratives. Existing player type models could be modified to suit literary experiences. For example, as in the Hearts, Clubs, Diamonds and Spades model by Bartle (1996), readers could be categorised as explorers, achievers, socialisers, and killers based on their reading habits, preferences, and motivations. This could be used to analyse the types of literature that different player or reader types are drawn to, and investigate whether certain genres, themes, or narrative structures are more appealing to specific player/reader types. For example, explorers may prefer complex, open-ended narratives, while achievers may seek clear goals and resolutions, and killers engage with dark and morally ambiguous narratives. One could also study how different player types engage with literature and what motivations and goals they have, exploring whether explorers tend to read multiple books simultaneously, whether achievers aim to gain knowledge or finish a certain number of books per month, or whether socialisers participate in book clubs or discussions. One could also examine how player/reader types relate to literary characters and analyse whether certain player types are more likely to identify with protagonists, antagonists, or secondary characters.

However, as seen in Chapter II, player typologies might not be ideal for even games, let alone literature. The existing player typologies, although insightful, are subject to several limitations that hinder their comprehensive applicability and precision. First and foremost, a significant drawback lies in the lack of empirical validation for most of these models. Many player preference models, following Bartle's seminal work, have not undergone rigorous empirical testing, raising concerns about the generalisability of their findings beyond specific gaming contexts. This limitation undermines the robustness and reliability of the typologies, hindering their effectiveness as universal frameworks. Furthermore, the absence of a standardised assessment tool poses a challenge. The diversity in assessment methodologies across different studies makes it difficult to compare findings and draw overarching conclusions about player motivations. A standardised tool would enhance the validity and reliability of the typologies, ensuring consistency in research outcomes. Another critical limitation is the narrow scope of many typologies, often tailored to specific game genres or platforms. The evolving landscape of gaming, with the emergence of new elements like body movement-controlled games, electronic sports, streaming, and casual games, demands a more inclusive approach. Existing typologies may struggle to accommodate these diverse play styles and experiences, limiting their relevance in contemporary gaming contexts.

Hamari and Tuunanen (2014) rightly emphasise the constraints shared by most player typologies, particularly in their reliance on Bartle's work. Bartle's model, originating in the context of MUDs, may not capture the intricacies of modern gaming preferences. The typologies often categorise players into a few broad types, oversimplifying the rich diversity of player traits and interests. An inherent flaw in player typologies is the post hoc consideration of personality and motivation. Tondello et al. (2016) highlight this issue, underscoring the need to integrate these factors from the inception of typology development. Incorporating personality and motivational factors early in the typology creation process ensures a more holistic understanding of players. Bateman, Lowenhaupt, and Nacke (2011) shed light on the inadequacies of type theories, calling for a shift towards a new trait theory of playing preferences. This suggests the necessity of developing typologies based on a solid foundation of a novel trait theory tailored explicitly for gaming preferences. Such an approach would depart from conventional psychological models and better capture the nuanced dimensions of gaming behaviour.

Building on these criticisms, the proposal for reader typologies involves overcoming the identified shortcomings. A reader typology should be characterised by empirical validation, utilising standardised assessment tools to enhance reliability. It should adopt a broad and inclusive approach, accommodating various reading styles, preferences, and emerging trends in literary engagement. The incorporation of personality and motivation considerations should be integral to the typology's development, reflecting a more nuanced understanding of readers. Ultimately, a reader typology should strive for adaptability and relevance in the dynamic landscape of contemporary reading experiences. The development of reader type models should involve a combination of qualitative and quantitative research methods. Qualitative methods, such as in-depth interviews and focus groups, could be employed to explore readers' motivations, preferences, and emotional responses to different types of narratives. These insights would be essential for identifying the key dimensions along which readers differ and for constructing preliminary reader type categories. Quantitative methods, including surveys and factor analysis, could then be used to validate these categories and refine the model. By analysing data on readers' genre preferences, reading habits, and engagement with various literary themes, researchers could identify distinct clusters of reader types. For instance, cluster analysis could reveal groups of readers who consistently seek out complex, multilayered narratives (Explorers) versus those who prefer clear, goal-oriented stories (Achievers).

In fact, personalisation techniques could extend beyond traditional personality models and delve into other individual differences that shape reading experiences. For example, individuals with diverse reading motivations, such as those seeking escapism, intellectual stimulation, or emotional catharsis, might benefit from tailored narratives that align with their specific desires. Future studies could explore how these reading motivations intersect with personality traits and inform the development of personalised narratives that cater to readers' unique preferences.

One way to device a framework specifically for literature could be to make a comprehensive list of different emotions people can seek from literature, and create reader profiles based on how important each emotion is for them. For example, and in order to expand on these ideas discursively, below we have presented a non-authoritative list of some different emotional states that readers may often seek or experience when engaging with fiction:

Table 13: Emotional states sought from fiction.

1. Joy and Happiness	Many readers seek stories that evoke feelings of joy, happiness, and contentment. These may include heartwarming tales, romantic comedies, or stories with uplifting and hopeful themes.
2. Excitement and Thrill	Readers often enjoy narratives that provide excitement, suspense, and thrills. Thrillers, mysteries, and action-adventure stories are known for eliciting feelings of excitement and anticipation.
3. Fear or Anxiety	Horror and suspense fiction aim to evoke fear or anxiety in readers. Some individuals seek the adrenaline rush of being scared while safely immersed in a fictional world.
4. Sadness and Empathy	Stories with poignant and emotional content can bring about feelings of sadness and empathy. These narratives often explore themes of loss, grief, and human struggles.
5. Love and Romance	Romance fiction is designed to evoke feelings of love, affection, and passion. Readers enjoy experiencing the emotional ups and downs of romantic relationships.
6. Surprise and Shock	Plot twists, unexpected revelations, and surprises can elicit shock and surprise from readers. These elements keep readers engaged and guessing.
7. Empowerment and Inspiration	Inspirational stories and tales of personal growth can leave readers feeling empowered and motivated to overcome challenges in their own lives.
8. Curiosity and Intrigue	Mysteries and puzzles within a narrative can pique readers' curiosity and drive them to uncover hidden truths and secrets.
9. Awe and Wonder	Fantasy and science fiction often create worlds filled with wonder and awe. Readers may seek these genres to experience a sense of awe-inspiring imagination.
10. Nostalgia and Sentimentality	Nostalgic fiction can transport readers to a different time or evoke feelings of sentimentality, reminiscing about the past.
11. Anger and Indignation	Some narratives tackle themes that provoke anger, frustration, or a sense of injustice. These stories may prompt readers to reflect on societal issues.

12. Catharsis and Healing	Reading emotional narratives can offer catharsis, allowing readers to release pent-up emotions and find a sense of healing and closure.
13. Compassion and Understanding	Narratives featuring diverse characters and experiences can foster compassion and a deeper understanding of different perspectives.
14. Satisfaction and Closure	Many readers seek resolutions and closure in stories, providing a sense of satisfaction and fulfilment.
15. Arousal and Sensuality	Erotic fiction and sensual romance novels aim to arouse readers' desires and passions.
16. Hope and Optimism	Stories with themes of hope and optimism can uplift readers and leave them with a positive outlook on life.
17. Melancholy and Reflection	Some readers appreciate narratives that evoke feelings of melancholy and lead to introspection and self-reflection.
18. Bittersweetness	Narratives that combine elements of joy and sadness can evoke bittersweet emotions, leaving readers with mixed feelings.
19. Amusement and Laughter	Comedy and humorous fiction are designed to make readers laugh and experience amusement.
20. Awe and Admiration	Biographies, historical fiction, and tales of heroism can inspire awe and admiration for real or fictional figures.

These kinds of emotional descriptors could be used for a factorial model, for instance, giving a score for how important each of these is for a given user, perhaps based simply on what they have liked. Users could also be asked to create such a profile for themselves, or asked, when looking for something new, what sort of emotions they would currently like to seek. This could involve a tagging system for books and stories, where each piece of content is tagged with the emotions it elicits, done by either users, staff or sentiment analysis. The user profile could also be complemented with the factors of literature seen above in the previous list, such as preference for different genres or narrative structures. The distinction between familiarity and novelty could be particularly important, and a wholly different recommendation style might be needed for people who prefer novelty, as recommendation systems can be all about building bubbles of familiarity.

The exploration of alternative frameworks opens avenues for a more nuanced understanding of reader preferences. By incorporating diverse psychological models, fostering reader communities based on personality types, and adapting player type models to literature, researchers can embark on a journey to unravel the intricate dynamics of individual interactions with fictional content. Further research and experimentation are imperative to refine these frameworks and tailor them to the unique landscape of literature. The integration of emotional frameworks offers a promising direction, allowing for a personalised approach that goes beyond traditional personality classifications.

8. Advancing NLP for Personalisation

The exploration of Natural Language Processing (NLP) in the second study marks the initial strides in a transformative journey toward more sophisticated personalisation. Future research in this domain holds the key to refining NLP techniques, enabling automation, and augmenting the personalisation process. However, this journey is not without its challenges, and addressing specific hurdles will be crucial for unlocking the full potential of NLP-driven personalisation.

One significant hurdle in advancing NLP for personalisation lies in effectively adapting language to individual preferences. Language is a nuanced and dynamic aspect of communication, shaped by various factors such as cultural background, regional differences, and individual idiosyncrasies. NLP models must evolve to grasp these subtleties and tailor language outputs in a manner that resonates with the user. Achieving this level of sophistication requires a nuanced understanding of the intricate interplay between language nuances and personality traits.

A key ethical issue in advancing NLP-driven personalisation is the risk of reducing literary diversity. By continually adapting language and content to suit a reader's preferences, there is the potential for readers to become insulated within a narrow range of styles, themes, and genres that align with their immediate tastes. This risks limiting exposure to challenging or unfamiliar works that might foster intellectual and emotional growth. Personalisation, if not carefully managed, could reinforce echo chambers, leading to a homogenised reading experience where diverse voices and unconventional narratives are less likely to be encountered. Writers, particularly those who seek to push the boundaries of literary convention, may find their work increasingly filtered out by algorithms designed to cater to prevailing reader preferences.

Moreover, the focus on reader preferences could alter the creative process for writers. In a personalised literary environment, there may be commercial pressures for authors to write in ways that align with algorithmic recommendations or to adjust their style to maximise engagement with target audiences. This raises concerns about the commodification of literature and the extent to which personalisation might influence authors' creative autonomy. Writers may feel compelled to produce content that is more readily adaptable to personalised algorithms, potentially prioritising marketability over artistic integrity. The fear is that literature may become more formulaic, with creative risks being sacrificed in favour of predictability and widespread appeal.

Data privacy is another significant ethical concern in this context. Personalisation relies heavily on collecting and analysing large amounts of personal data, including readers' behavioural patterns, preferences, and potentially sensitive psychological traits. While NLP models can be designed to adapt language outputs to align with users' personalities, this raises questions about consent, data security, and the potential misuse of personal information. Readers may be unaware of the extent to which their data is being used to shape their experience, and there is a risk that such data could be exploited for commercial purposes beyond personalisation, such as targeted advertising or manipulation of consumer behaviour.

NLP technologies have the potential to automate and streamline the process of altering writing styles, allowing for more efficient and nuanced personalisation. By harnessing NLP algorithms, narratives could dynamically adapt in real-time to match users' personality traits, reading motivations, or emotional states. This level of responsiveness could significantly elevate the user experience by creating a seamless and immersive narrative journey.

The real-time adaptation capabilities of NLP algorithms could also be integrated into interactive storytelling platforms, where the narrative evolves in response to the user's input during key decision points. In this context, NLP can be used to not only change the direction of the story but also to fine-tune the details, such as the specific wording used in dialogue or the descriptive language in scene settings, ensuring that each narrative element contributes to a cohesive and personalised experience. For instance, a narrative could adjust its dialogue to match the preferred communication styles of Introverted or Extraverted users. Additionally, the pacing of the story could be fine-tuned to cater to users' preferences for either fast-paced action or contemplative reflection.

NLP algorithms can be harnessed to dynamically adapt narratives in real-time by employing a combination of techniques, including sentiment analysis, topic modelling, and context-aware text generation. These methods allow the system to continuously monitor and respond to the user's interactions, ensuring that the narrative remains engaging and aligned with the user’s evolving preferences. For example, sentiment analysis could be used to assess the user's emotional responses to different narrative elements, allowing the story to adjust its tone and content accordingly. If the system detects that a user is particularly engaged by emotionally intense scenes, it could introduce more such elements, deepening the narrative’s emotional impact. Conversely, if a user exhibits discomfort or disinterest during distressing scenes, the system could shift the focus to more uplifting or neutral content, thereby maintaining the user’s engagement.

Context-aware text generation, powered by advanced language models, could be used to tailor the narrative in real-time based on the user's past choices and current emotional state. This approach could enable the creation of dialogue and narrative descriptions that feel highly personalised, as if the story were being co-authored by the user. For example, a character's response to the user’s actions could be dynamically generated to reflect a deep understanding of the user’s personality traits, leading to a more authentic and immersive interaction.

Furthermore, topic modelling can be applied to dynamically alter the themes and subplots within a story. By analysing the user’s previous interactions and stated preferences, the system could introduce new themes that resonate with the user’s interests, or shift the focus of the narrative to explore topics that align with their curiosity. For instance, a user with a high Openness to Experience might see the introduction of more abstract, philosophical discussions within the narrative, while someone with a preference for action might encounter more suspenseful and plot-driven sequences.

Moreover, the scalability of NLP-driven personalisation allows for the continuous refinement of the narrative as more data is gathered from user interactions. Machine learning models can be trained to recognise patterns in user behaviour and preferences, enabling the system to anticipate the user’s needs and pre-emptively adjust the narrative in a way that feels both intuitive and natural. However, while the potential for real-time dynamic adaptation through NLP is vast, it is not without challenges. Ensuring that these adaptations do not disrupt the coherence of the story or result in jarring shifts in tone or style is crucial. Additionally, maintaining a balance between user-driven personalisation and authorial intent is essential to preserve the integrity of the narrative.

The approach of using NLP algorithms to dynamically adapt narratives in real-time can be compared to the mechanisms employed in AI Dungeon, a well-known interactive text-based game that uses artificial intelligence to generate and adapt stories based on user input. Both approaches leverage AI to create personalised and responsive narrative experiences, yet they differ in their underlying methodologies and the depth of personalisation offered.

AI Dungeon utilises a model based on OpenAI's GPT-3, which generates text in response to user commands, allowing for a highly flexible and open-ended storytelling experience. The system is designed to respond to any input, generating narrative developments that align with the user’s commands, thereby providing a broad scope for user agency. However, AI Dungeon’s adaptability is primarily reactive rather than proactive. While it responds creatively to user input, it does not inherently adjust the narrative based on a detailed understanding of the user's personality, preferences, or emotional state unless explicitly directed by the user's commands.

In contrast, the proposed approach of harnessing NLP algorithms for real-time adaptation goes beyond simply reacting to user inputs. It aims to create a more sophisticated form of personalisation by dynamically adjusting the narrative to align with deeper aspects of the user’s identity, such as personality traits, emotional states, and reading motivations. This could involve tailoring narrative elements such as tone, pacing, and dialogue to create a more immersive and resonant experience for the user. For instance, the narrative might subtly shift in tone to match an Introverted user's preference for introspective storytelling or adjust the pacing to cater to a user’s desire for either rapid or gradual plot development.

Furthermore, while AI Dungeon excels in providing an open-ended, sandbox-style environment where the narrative can take almost any direction, the NLP-driven approach could potentially offer a more guided and coherent narrative structure, with personalisation enhancing rather than overwhelming the story's internal consistency. This approach could ensure that the narrative remains engaging and cohesive while still being deeply personalised, addressing one of the potential limitations of open-ended AI-generated content, which can sometimes become disjointed or lose focus.

From a practical perspective, the quality of the data used to train NLP models for personalisation must be considered. If datasets are not diverse or representative, there is a risk that personalisation algorithms could perpetuate biases and exclude marginalised voices. Ensuring that NLP models are trained on a wide range of linguistic styles, cultural contexts, and demographic backgrounds is essential to prevent the reinforcement of stereotypes or the marginalisation of non-mainstream literature. Writers from underrepresented communities may find their work sidelined if personalisation models favour content that conforms to majority tastes.

The effectiveness of NLP-driven personalisation relies heavily on the quality of the underlying data. Robust datasets are essential for training models that accurately capture individual preferences and linguistic nuances. Data curation requires structured and controlled processes to ensure datasets are of high quality, representative of diverse user experiences, and relevant to the specific goals of the NLP model. Future research should prioritise developing datasets that encompass a wide range of linguistic styles, user demographics, and reading preferences. Rigorous curation and annotation are imperative to ensure models are trained on reliable and representative data.

Data collection must be conducted ethically, with proper user consent and privacy safeguards. For instance, user interaction data from reading platforms should be anonymised to protect identities and comply with regulations like GDPR. Annotations should be carried out by experts familiar with both the literary domain and language nuances. Human annotators trained to categorise text elements are often used, although semi-automated or crowdsourced methods may be considered, provided quality and consistency are maintained.

Ensuring the dataset reflects the diversity of user experiences and preferences is crucial. This involves including a variety of genres, authors, and linguistic styles, while ensuring representation across age, gender, ethnicity, and cultural background. Such diversity prevents the reinforcement of biases and the marginalisation of certain groups. To achieve this, curators might stratify the dataset or actively seek out underrepresented voices to enrich it. Additionally, addressing underrepresentation in MBTI types may require broadening the search for textual data, engaging with participants via surveys, collaborating with institutions, employing advanced NLP techniques, and fostering community discussions to achieve a more balanced dataset.

In the context of personalisation, especially in interactive and dynamic narrative environments, real-time processing capabilities are paramount. Users engage with content in the moment, and NLP models must swiftly adapt to user inputs, preferences, and evolving contexts. Future research should focus on enhancing the real-time processing speed of NLP algorithms to deliver seamless and responsive personalisation experiences. This involves striking a balance between computational efficiency and the complexity of linguistic analyses.

The iterative nature of personalisation requires a continuous feedback loop. Incorporating user feedback into the NLP-driven personalisation process is crucial for refining and adapting recommendations over time. Future research should explore methods to seamlessly integrate user feedback into NLP models, enabling dynamic adjustments based on user responses and evolving preferences. This iterative feedback loop contributes to the adaptive nature of personalisation systems.

The feedback loop inherent in personalisation also poses ethical challenges. While integrating user feedback into NLP models is essential for refining recommendations, it can lead to self-reinforcing cycles where users are only exposed to content that aligns with their established preferences. This risks stifling curiosity and intellectual diversity, as readers may be less inclined to explore genres or authors outside their comfort zones. Moreover, over-reliance on personalisation could devalue serendipitous discovery, a fundamental part of the literary experience where readers encounter unexpected works that challenge their worldview or introduce them to new ideas.

The landscape of narrative personalisation has been significantly reshaped by advancements in large language models (LLMs), such as OpenAI's GPT-4 and Google's Bard. These models have revolutionised natural language processing through their ability to generate human-like text, understand complex linguistic patterns, and adapt to various contexts. One of the most impactful innovations in LLMs is their capability to process and generate contextually appropriate narratives in real time. This development has particular relevance to personalised storytelling, where LLMs can craft narratives that evolve based on user input, personality traits, or behavioural patterns. For instance, by analysing a user's past interactions with content, LLMs can generate storylines that align with the player's emotional states, personality, and preferences, thereby offering a more personalised and immersive experience.

Moreover, advancements in reinforcement learning techniques, combined with LLMs, allow models to adapt content based on real-time user feedback. This iterative, self-improving process could be applied to personalised music curation or even interactive films, where a viewer's decisions or reactions shape the unfolding narrative. Such real-time adaptability not only enhances user engagement but also opens up avenues for more emotionally resonant content.

Recent developments also highlight the role of LLMs in improving the personalisation of recommendations. Netflix, Spotify, and other content platforms already use machine learning algorithms to suggest films or music. However, LLMs could refine these systems by incorporating more nuanced user data, such as mood or cognitive preferences, rather than relying solely on past consumption patterns. By providing more tailored content recommendations, these platforms can offer users a deeper emotional connection to the material, improving satisfaction and engagement.

Finally, the integration of LLMs with multimodal systems—those that combine text, images, video, and sound—offers further potential for personalisation across creative domains. LLMs, in conjunction with visual models like DALL·E, can generate personalised multimedia content, such as bespoke interactive films or music videos, where both narrative and visual elements are tailored to the user's personality and engagement patterns. This multidimensional personalisation holds immense potential for redefining user interaction with content across various artistic and entertainment forms. As LLMs continue to evolve, their integration into personalised content creation and curation will likely become more seamless, enabling more intimate, emotionally resonant experiences across creative domains.

The trajectory of NLP for personalisation is poised for significant advancements, with the potential to revolutionise how individuals interact with literary content. Future research endeavours should be dedicated to overcoming challenges related to language adaptation, ensuring data quality, addressing real-time processing demands, integrating user feedback seamlessly, and navigating the ethical dimensions of personalisation. By surmounting these challenges, NLP-driven personalisation can evolve into a powerful tool, enhancing the richness and relevance of literary experiences for diverse audiences.

9. Pathways for Personalisation

While this research has made progress in understanding how personalisation techniques based on personality traits can enhance user experiences in interactive narratives, it also highlights the vast potential for further exploration and innovation in this field. The success of certain personalisation methods, as well as the inherent adaptability of interactive narratives, paves the way for exciting avenues of research and development in the realm of personalisation techniques.

The exploration of alternative personalisation techniques in interactive narratives is a rich area for future research and innovation, having deep potential in creating more and better content for different users. As technology continues to advance and our understanding of user preferences deepens, we can anticipate increasingly sophisticated and tailored narrative experiences that cater to a diverse range of individual traits, motivations, and preferences. This evolving landscape holds the potential to revolutionise how we engage with and appreciate the art of storytelling.

To consider new avenues for personalisation, we should isolate the different factors in literature that are relevant to it, and what the appeal of literature is in the first place and to different individuals.

Based on what has been discussed throughout the thesis, the following list of different factors can be derived:

Table 14: Factors relevant to literature.

Genre Preferences	People often have genre-specific preferences, such as a love for science fiction, fantasy, romance, mystery, or historical fiction. Understanding how individual traits align with genre choices can be insightful.
Plot Complexity	Some readers may enjoy intricate, multi-layered plots, while others prefer straightforward narratives. Assessing how individuals' cognitive styles relate to their plot complexity preferences can be valuable.
Narrative Structure	Books with specific narrative structures, such as linear, nonlinear, or experimental storytelling, based on the reader's preference.
Character Development	The depth of character development, including relatability and complexity, can significantly impact reading enjoyment. Studying how personality traits affect preferences for well-developed characters can provide insights.
Narrative Style	Providing books with specific narrative styles, such as first-person, third-person, present tense, or past tense, depending on the reader's preference.

Emotional Engagement	Some readers seek emotionally charged narratives, while others prefer a more detached reading experience. Neuroticism and Need for Affect could be relevant traits to explore here.
Writing Style and Language	Preferences for writing style, tone, and language complexity can vary wildly. It was found that Extraversion might relate to preferences for informal language; what else could be found?
Familiarity and Novelty	Some might prefer predictability and familiar characters and stories, others surprises and something they haven't seen before.
Story Length	Narratives of varying lengths, from short stories to epic novels, based on the reader's available time, reading pace, preferences and characteristics such as patience.
Emotional Tone	Books with emotional tones (e.g. humour, drama, suspense) that align with the reader's mood, emotional state, or general preferences.
Temporal Setting	Preferences for stories set in different time periods, such as historical, contemporary, or futuristic settings, can be influenced by individual characteristics and experiences.
Moral and Ethical Themes	Individuals' moral values and ethical beliefs might align with their enjoyment of stories with specific moral themes or dilemmas.
Cultural and Social Factors	The impact of cultural background, social context, and personal experiences on reading preferences. Cultural sensitivity and diversity in reading preferences should be considered.
Interactivity and Engagement	In the context of interactive narratives, individuals' need for agency, control, and immersion in the story influences their choices and satisfaction with the narrative.
Demographics	Demographic factors like age, gender, educational background, and socioeconomic status can also play a role in reading preferences such as having protagonists whose gender, ethnicity, or background aligns with the reader's identity or interests. These factors may intersect with personality traits and genre preferences.
Technology and Medium	The medium through which fiction is consumed (e.g. physical books, e-books, audiobooks, interactive apps) can impact reading preferences and experiences.

Visual Enhancements	Some may prefer a reading experience with illustrations, maps, or multimedia elements.
Reading Speed and Pace	Adjusting the pacing of e-books or audiobooks to match the reader's reading speed or listening preferences. This would also be particularly important in gaming.
Psychological States	Temporary psychological states, such as mood, stress levels, and current life circumstances, can influence reading choices. These states can interact with personality traits to shape preferences.
Reading Goals	How individuals' goals for reading (e.g. escapism, learning, entertainment, self-improvement) align with their choices of reading material and narrative preferences.
Neurocognitive Factors	Neurocognitive processes involved in reading comprehension, including attention, memory, and cognitive load. How these processes interact with personality traits can shed light on reading preferences.
Feedback and Recommendations	User feedback, peer recommendations, and algorithmic book recommendations influence reading choices. This factor intersects with individual preferences and external influences.
Personalised Content	Personalisation itself, including content recommendations and adaptive narratives, could have an impact on the experience, particularly if the reader is aware of it, which might give them a different set of expectations and receptiveness.

What can be done with these? Each FFM trait could be associated with distinct reading preferences. Whilst some preferences would be well-suited for personalisation, some might be less so, and recommendations would be more helpful. For instance, similarly to Hirsh, Kang and Bodenhausen (2012), Extraverts may seek excitement and social rewards, Agreeable people themes relating to family and community, Conscientious people may seek efficiency and goals, Neurotic people safety and security, and those high in Openness creativity and intellectual stimulation, preferring imaginative and unconventional narratives, something avant-garde or intellectually challenging. Additionally, perhaps those high in Conscientiousness might gravitate towards structured and well-organised plots, and perhaps self-help books. Readers high in Neuroticism might seek emotionally intense stories with complex conflicts. Those with Emotional Stability may prefer lighter, feel-good stories. Or this could even turn out to be the other way around. These are just some suggestions that could be explored. Genre preferences of people with different personalities have often been studied before, as seen in Chapter II.4.1.

Character development, on the other hand, can work with both recommender systems and personalisation. Understanding how the FFM traits relate to preferences for character depth and complexity can help personalisation create more relatable and engaging characters. Characters with relatable flaws and vulnerabilities may resonate with individuals high in Neuroticism, allowing for emotional connection. Characters who display kindness and cooperation can appeal to those high in Agreeableness.

Language style, as already explored, could be the easiest aspect to personalise automatically. Readers with high Extraversion appear to appreciate conversational and engaging writing styles. On the other hand, individuals high in Conscientiousness might favour precise and structured language. More research would be needed to confirm more such speculative preferences.

By associating these factors with individual personality traits, it is possible to tailor the reading experience in a way that aligns more closely with a reader's intrinsic preferences and psychological makeup. This approach could represent a significant enhancement over existing tools such as Storygraph, Storywise, and Goodreads, which offer personalised recommendations but do so with varying levels of sophistication and focus.

Storygraph is one of the most advanced tools in terms of personalisation, allowing users to track their reading preferences with fine detail, including mood, pace, and genre preferences. The platform enables readers to discover books that match their preferences based on past behaviour and user input. Storywise, meanwhile, takes a narrative-driven approach, focusing on the type of stories readers want to experience. It categorises stories based on themes, tropes, and narrative elements, offering recommendations that align with these preferences. However, like Storygraph, Storywise is primarily driven by explicit user input and genre preferences. It does not integrate psychological models or neurocognitive factors that could more precisely tailor the reading experience to an individual's psychological profile. On the other hand, Goodreads offers a more traditional recommendation system based on social networking. It leverages user reviews, ratings, and reading history to suggest books. While Goodreads benefits from a vast database and user community, its recommendation system is less sophisticated in terms of personalisation.

The proposed system, drawing on the factors listed in Table 13, would take personalisation to a new level by incorporating psychological models such as the FFM. This would enable the system to go beyond surface-level preferences and delve into the deeper motivations and emotional needs of readers. For example, by identifying a reader as high in Openness to Experience, the system could recommend avant-garde or intellectually stimulating books, rather than merely suggesting books based on past genre preferences. This system could also dynamically adapt to the reader's current psychological state or mood, a feature that is largely unexplored by existing tools. For instance, if a reader is feeling stressed, the system could suggest lighter, more comforting reads that align with a preference for emotional stability. Conversely, if a reader is in the mood for a challenge, it could recommend something more complex or emotionally intense. Another key enhancement would be the integration of neurocognitive factors, such as attention span, cognitive load, and memory. The system could adjust recommendations based on how much mental effort a reader is willing or able to invest at a given time. For instance, during a period of cognitive fatigue, the system might suggest shorter, more straightforward narratives rather than complex, multi-layered plots.

Moreover, by understanding the reader’s goals – whether for escapism, learning, or self-improvement – the system could tailor recommendations that not only match their preferences but also their desired outcomes from reading. For example, a reader seeking self-improvement might be guided towards structured, goal-oriented literature, while someone seeking escapism might be directed to immersive, imaginative narratives.

The potential for integrating this system with interactive narratives offers another avenue for differentiation. Unlike static recommendations, this system could dynamically adapt the narrative content in real-time, responding to changes in the reader’s mood, psychological state, or preferences. This approach could transform the reading experience from a passive activity into an interactive, personalised journey, where the narrative evolves in response to the reader’s engagement and emotional responses.

Such insights are not relevant to just personalisation and recommender systems, but also marketing and helping with who to advertise to. Publishers and marketers can gather data on readers' personality traits through surveys, online interactions, or purchase histories. This data can inform marketing decisions, content creation, and book recommendations. They can then create genre-specific marketing campaigns designed to appeal to readers with particular personality traits. They can also use FFM insights to identify specific personality traits that align with particular books or genres. Once target audiences are identified, marketing strategies can be tailored to appeal to those with specific personality traits. For instance, marketing campaigns aimed at Extraverts might emphasise community engagement, social events, and interactive book clubs to attract readers who thrive on social interactions. Bookstores and digital platforms can curate book selections based on the dominant personality traits of their customer base. For example, an online community with predominantly Conscientious individuals might highlight well-organised, self-help, and productivity-oriented books. The design and aesthetics of book covers could also be aligned with the preferences of target personality traits: For example, books targeting individuals high in Openness might feature unique, artistic, or abstract cover designs. Structured and well-organised cover layouts can appeal to those with high Conscientiousness.

It's important to note that while the FFM provides valuable insights into personality, individuals are complex and multifaceted. People exhibit a range of traits, and preferences can be influenced by various factors beyond personality. Therefore, while the FFM can offer valuable guidance, a holistic approach that considers additional factors, such as cultural background, mood, and life experiences, is essential for a comprehensive understanding of reading preferences. Combining FFM insights with data analytics and machine learning techniques can lead to more effective personalised reading experiences and recommendations.

The detailed exploration of factors influencing personalisation in literature highlights the intricate interplay between individual traits, preferences, and contextual elements. The list provides a comprehensive overview of potential dimensions for personalisation, ranging from narrative elements to psychological states. The association of the FFM traits with specific preferences offers a structured framework for tailoring reading experiences. However, the nuanced and multifaceted nature of individual preferences necessitates a cautious approach. The proposed pathways for personalisation open up opportunities for future research and practical implementations, particularly in the realms of recommendation systems, character development, and marketing strategies.

Further analysis could focus on the potential challenges and ethical considerations associated with implementing personalisation strategies. Additionally, empirical studies and user feedback could validate the proposed associations between the FFM traits and reading preferences. Exploring the dynamics of reader communities and their role in shaping preferences could provide deeper insights into the social aspects of reading. While the pathways for personalisation lay a foundation for enhancing reader experiences, continual refinement and adaptation based on user feedback and evolving preferences are essential.

The intersection of psychological traits, narrative elements, and external influences forms a dynamic landscape that requires ongoing exploration and innovation.

10. Cross-Domain Personalisation

The insights gained from this thesis extend beyond literature and interactive storytelling. They have the potential to find application in various creative domains, including gaming, film, music, and beyond. Exploring the transferability of personalisation techniques to these domains can unlock new possibilities for tailoring content and recommendations. The intricate understanding of user preferences, behaviour patterns, and the interplay between personality traits and content experiences can be seamlessly applied to other creative domains.

In the domain of literature, the insights gained from personalised storytelling and language adaptation can revolutionise the way narratives are crafted and presented. Personalised narrative experiences, shaped by individual traits, can elevate the reader's connection with the story, potentially leading to a more immersive and satisfying reading experience. Personalised recommendations, aligned with readers' personalities and preferences, can redefine how books are suggested and marketed.

By incorporating personalisation techniques, game developers can tailor in-game narratives, character interactions, and gameplay mechanics to align with players' personalities and preferences. This approach could lead to more immersive gaming experiences, where players feel a deeper connection to the game world and its characters. Additionally, personalised recommendations for games based on players' personality traits could enhance the discovery of titles that resonate with their gaming style and preferences. As gaming continues to evolve as a form of interactive entertainment, the integration of personalisation techniques can further blur the lines between storytelling and gameplay, offering players unique and engaging experiences tailored to their individual traits and preferences.

Applying personalisation techniques to the film industry holds immense potential for reshaping the cinematic landscape. Tailoring movie recommendations based on viewers' personalities can enhance the discovery of films aligned with their tastes. Furthermore, exploring personalised narrative structures within films, akin to interactive storytelling in gaming, could pave the way for a new era of viewer engagement and cinematic storytelling.

Black Mirror: Bandersnatch is one of the most visible examples of a "choose-your-own-adventure" format applied to a mainstream media production. The film allows viewers to determine the protagonist's actions, leading to multiple endings and narrative branches. Another example is Steven Soderbergh's *Mosaic* (2017), a narrative that unfolds across multiple media platforms, including an app that allows viewers to explore different perspectives and timelines. However, as AI and machine learning continue to develop, future films may move beyond simple branching structures. AI could enable films to adapt in real time based on a viewer's past decisions or even analyse individual preferences, psychological traits, or emotional states, leading to highly personalised and unique narrative experiences. As technology continues to advance, future personalised narratives may integrate more sophisticated forms of interaction, such as adaptive AI or even biometric feedback, to create truly immersive and responsive storytelling experiences.

The realm of music stands to benefit significantly from cross-domain personalisation. Understanding how personality traits influence musical preferences can inform personalised music recommendations. Algorithmic systems could curate playlists that resonate with users on a deeper level, considering not just genre preferences but also emotional tones and thematic elements aligned with individual traits.

The implications of cross-domain personalisation extend to unconventional creative forms, opening doors to innovative applications in areas such as virtual reality experiences, augmented reality art installations, and interactive multimedia projects. As technology continues to evolve, exploring personalisation across these novel creative landscapes could redefine how users interact with and consume content.

For example, *Rain Room* by Random International, first exhibited at the Barbican in London (2012), provides an immersive experience where visitors walk through a simulated rainstorm, with motion sensors ensuring that no raindrops fall on them. While not personalised in the sense of individual traits, the installation adapts to the visitor's movements, creating a unique experience for each participant. Expanding on this concept, future iterations could use biometric data or emotional response trackers to adjust the sound, lighting, or intensity of the rain to match the user's psychological state, offering a deeply personalised interaction with the environment.

Similarly, *The Weather Project* (2003) by Olafur Eliasson at Tate Modern used atmospheric elements like mist, mirrors, and lighting to evoke different emotional reactions from visitors. In a more personalised version of such installations, VR or AR technology could be incorporated to modify the environment based on user input or preferences, altering the mood or setting according to their emotional state or sensory preferences. Personalisation in such works could extend beyond interaction to the adaptation of the experience in real-time based on user-specific data, such as mood or movement patterns.

In the realm of augmented reality, *Thresholds* (2017), a VR installation by Mat Collishaw, reimagined a 19th-century photography exhibition using immersive technology. While the original work is designed for the general audience, AR technologies could tailor individual experiences by adapting the visual content based on user preferences for historical or contemporary themes.

Personalisation is prominently featured in the work of Meow Wolf, an art collective renowned for its large-scale immersive environments, such as *The House of Eternal Return* (2016). In future iterations of such installations, personalisation could enable each visitor to experience unique narratives or interactions tailored to their interests and previous interactions within the space. For example, AI algorithms could dynamically adjust the narrative threads, characters, and interactions a user encounters based on their engagement with the environment. This would create a highly individualised experience, allowing each visitor to navigate a personalised journey that reflects their preferences and history, enhancing their overall engagement and connection to the artwork.

These examples illustrate how cross-domain personalisation can transform conventional and digital art spaces, allowing for highly tailored experiences that respond in real-time to the individual needs and preferences of users. As these technologies evolve, they promise to make interactive art installations more engaging and personalised, fundamentally changing how audiences experience and interact with creative works.

While the prospect of cross-domain personalisation holds promise, it comes with its set of challenges. Each creative domain possesses its unique characteristics, audience expectations, and content delivery mechanisms. Adapting personalisation techniques to suit these diverse contexts requires a nuanced understanding of each domain's intricacies.

The exploration of cross-domain personalisation presents an exciting opportunity for interdisciplinary collaboration. Researchers, practitioners, and creatives from various domains can come together to share insights, methodologies, and best practices. Collaborative efforts can lead to the development of standardised frameworks that balance personalisation with ethical considerations. As the exploration of cross-domain personalisation progresses, it has the power to redefine creative expression and audience engagement, ushering in a new era of personalised and enriching content experiences.

11. User Empowerment and Privacy

One of the central challenges in personalisation is the issue of data privacy. Personalisation algorithms rely on the continuous acquisition of user data to improve their accuracy, but this data is often sensitive, particularly when it includes personal identifiers or psychological profiles. If not managed correctly, this raises significant privacy risks, including data breaches or misuse by third parties. The more personalised a service becomes, the more data it requires, and this can often feel intrusive to users. Therefore, it is essential for companies to implement clear privacy frameworks that outline exactly how user data is collected, stored, and applied.

Moreover, the notion of user consent is crucial. Many personalisation systems operate passively, collecting data without explicit user input beyond initial terms and conditions. This creates an ethical grey area, where users may not fully understand the extent of the data collection. For example, the use of machine learning algorithms that predict behaviour based on past interactions may lead to a form of personalisation that users did not explicitly request. As such, transparency is vital: users should be able to access, manage, and modify their personalisation settings easily. This includes giving them the ability to opt in or out of data collection practices, as well as providing insight into how their data influences the recommendations they receive.

The balance between convenience and autonomy is another ethical consideration. While personalisation algorithms can provide value by simplifying decision-making and presenting content that is closely aligned with a user’s preferences, there is a risk of over-reliance on these systems. If personalisation becomes too pervasive, it can limit a user's exposure to diverse content, creating so-called “filter bubbles” where individuals are only exposed to information that reinforces their existing beliefs or preferences. This is particularly problematic in domains such as news media or social networks, where the constant tailoring of content may contribute to political polarisation or limit a user's ability to critically engage with opposing viewpoints.

In artistic and creative contexts, such as literature or film, algorithmic determinism could stifle serendipity, leading to homogenised content experiences. If personalisation systems are too finely tuned, users may miss out on diverse or challenging works that fall outside their typical preferences. This raises a broader cultural question about the role of personalisation in shaping taste and limiting the breadth of cultural consumption.

Personalisation's effectiveness hinges on its ability to cater to diverse audiences with varying preferences and behaviours. Creating systems that can adapt to this diversity is a complex endeavour, requiring careful consideration of user segmentation and content adaptation. Additionally, the unpredictable nature of dynamic systems presents challenges in delivering consistent and reliable personalised experiences.

However, content adaptation becomes increasingly complex when applied across diverse and global audiences. Users from different cultural or linguistic backgrounds may respond differently to the same piece of content. What might resonate emotionally or intellectually with one group could be alienating or irrelevant to another. This makes it crucial for personalisation systems to incorporate a deep understanding of cultural context, language variation, and even historical or social factors when adapting content. For instance, music recommendation algorithms must understand how cultural factors influence musical preferences and not simply rely on universal genre preferences.

The nature of personalised systems is dynamic and continuously evolving, particularly when powered by machine learning algorithms. Unlike static systems, where content or services remain fixed, dynamic personalised systems must adapt in real-time, recalibrating recommendations or experiences based on continuous user interactions and feedback. While this dynamism offers flexibility and adaptability, it also introduces a degree of unpredictability that poses unique challenges. As personalisation systems respond to real-time data, users might experience variability in how content or recommendations are presented. A film recommendation that appears one day might disappear the next, or the way a website adapts to user preferences could shift unpredictably, leading to confusion or frustration. Maintaining a balance between responsiveness and consistency is key to ensuring that users still feel a sense of control and continuity in their interactions with a personalised system.

The unpredictable nature of dynamic systems is particularly challenging in fields such as entertainment, where personalisation could distort or limit the types of content users are exposed to. In music, for instance, recommendation systems driven by dynamic algorithms could over-prioritise certain artists or genres, leading to homogenisation and a lack of discovery of new or experimental works. In gaming or interactive media, personalisation might inadvertently steer users down repetitive narrative paths or towards gameplay options that reduce the richness of their experience.

One potential solution is the development of hybrid systems that combine rule-based personalisation with machine learning models. This approach could ensure that while algorithms continue to evolve based on user data, certain foundational rules or guidelines remain in place to provide a more consistent and reliable experience. Another area of innovation is in explainable AI where personalisation systems are designed to offer transparency in their recommendations, giving users insight into why certain content is being suggested and allowing them to adjust their preferences accordingly.

Empowering users to take control of their personalisation preferences and privacy settings is becoming an increasingly critical area of development as personalised systems proliferate across digital platforms. The success of these systems hinges not only on their ability to deliver tailored content but also on their transparency, ethical practices, and respect for user autonomy. As personalisation technologies evolve, fostering user trust through transparency and control becomes fundamental to their long-term acceptance and success.

A key aspect of user empowerment is providing granular control over privacy settings, allowing individuals to dictate how their data is collected, processed, and used for personalisation. Current systems often provide limited, binary options – users can either opt in or opt out of data collection entirely. However, more sophisticated systems, such as those employed by platforms like Google, are beginning to offer more refined control mechanisms. These allow users to manage what specific data points, such as location, search history or behavioural tracking, are used for personalisation. This approach to customisation ensures that users can make informed decisions about the trade-offs between personalisation and privacy. In streaming platforms like Netflix, users can adjust settings related to viewing history and preferences. Although Netflix does not offer detailed privacy customisation in terms of data control, it provides the ability to create separate user profiles, which can prevent individual preferences from influencing recommendations for shared accounts. Expanding such options to include data consent at a more granular level would further empower users to determine how much of their personal behaviour they wish to share.

To address this, personalisation systems should incorporate clear, user-friendly language in privacy policies and actively communicate how personalisation algorithms function. For example, Spotify's recommendation algorithm, which analyses user listening habits, could benefit from providing a transparent overview of how it curates playlists, allowing users to see which listening behaviours influenced their recommendations. This would foster a better understanding of how user data feeds into the system's decisions, demystifying the often opaque process of algorithmic personalisation.

Providing this transparency not only builds trust but also empowers users to make informed choices. By demystifying the role of data in personalisation, platforms can create environments where users feel a sense of control and ownership over their digital experiences. This is particularly crucial in a time where users are becoming increasingly concerned about privacy breaches and data misuse.

Ethics plays a central role in ensuring that personalisation systems respect user autonomy and privacy. Personalisation algorithms are not value-neutral; they can inadvertently perpetuate biases, amplify stereotypes, or exclude marginalised groups if not designed thoughtfully. For example, social media algorithms have been criticised for reinforcing echo chambers by only presenting users with content that aligns with their pre-existing beliefs, thereby limiting exposure to diverse perspectives.

Central to the concept of user empowerment is the idea of user agency, where individuals not only have control over their privacy but also understand and influence how personalisation systems operate. In many current systems, users are passive recipients of personalised content, with limited ability to interact with or adjust the recommendations they receive. However, as personalisation technologies evolve, there is a growing recognition of the need for feedback loops that allow users to fine-tune the system based on their evolving preferences.

Platforms like YouTube and Netflix already provide basic feedback mechanisms, such as "like" or "dislike" buttons, allowing users to indicate whether a particular recommendation aligns with their interests. These interactions help the algorithm adjust future suggestions. However, more sophisticated systems could offer deeper insights, such as explaining why specific recommendations were made and allowing users to adjust the underlying factors influencing those suggestions. For example, a music streaming platform could allow users to not only like or dislike a song but also indicate whether the tempo, genre, or mood of the song was the most relevant factor, thereby giving the system more detailed guidance.

Providing these kinds of feedback mechanisms enhances user agency by creating a more dynamic and responsive personalisation system. Instead of a one-way flow of recommendations, personalisation becomes a collaborative process between the user and the system, ensuring that the experience evolves in line with the user's preferences over time.

User empowerment and privacy should not be seen as optional features of personalisation systems but as fundamental principles. By placing user control, transparency, and ethical practices at the heart of personalisation strategies, companies can create more trustworthy and effective systems. Emerging technologies such as AI-driven personalisation offer exciting possibilities for more nuanced, user-centric personalisation, where systems can learn from users in a manner that is transparent, respectful, and adaptive.

In the future, we may see more privacy-first personalisation systems, where users have full control over what data they share and how it is used. For example, companies could adopt blockchain technology to decentralise data storage, giving users greater control over their personal data and ensuring that it is only used with their explicit consent. Similarly, advances in explainable AI will enable personalisation systems to offer clearer insights into their decision-making processes, giving users more confidence in how their data is being used.

By prioritising user empowerment and privacy, personalisation systems can strike a balance between delivering highly tailored experiences and respecting individual autonomy. This approach will be essential in fostering the continued growth and acceptance of personalisation technologies, as users become increasingly aware of the importance of data privacy and ethical AI practices.

Lastly, from a regulatory standpoint, legal frameworks such as the General Data Protection Regulation (GDPR) in Europe have established the importance of data protection, mandating that users have control over their personal data. However, personalisation strategies often outpace regulatory development, particularly in industries like advertising, entertainment, and social media. Future regulations will need to address the specific challenges posed by AI-driven personalisation, ensuring that ethical guidelines are developed in tandem with technological advances.

Chapter VIII: Conclusions

In this concluding chapter, we briefly review the essential themes, key findings, and implications of our exploration into the realms of personalisation in narratives. We also consider the broader impact and future directions of personalisation, particularly in leveraging artificial intelligence and personality frameworks to create tailored and engaging user experiences.

Personalisation has emerged as a versatile and indispensable tool across various domains, including gaming, marketing, education, and persuasive systems. Its capacity to cater to individual preferences and behaviours has made it a valuable asset for creators and organisations seeking to connect with their target audience in a meaningful way. Whether it's crafting personalised gaming experiences, recommending products based on consumer preferences, or delivering tailored educational content, personalisation has consistently demonstrated its potential to enhance user engagement and satisfaction.

While personalisation has proven to be a valuable tool, it is not without its challenges and considerations. Balancing convenience and privacy remains a critical aspect of any personalisation strategy. Users must have control over their personalisation preferences, including privacy settings, to ensure a comfortable and ethical experience.

The incorporation of psychological personality models such as the FFM holds significant promise in personalisation. It enables content creators to gain deep insights into user preferences, allowing for the fine-tuning of narratives, games, and recommendations. The ability to adapt characters, language, and thematic elements based on a user's personality profile has the potential to revolutionise content creation across creative industries.

Furthermore, the recognition of the Need for Affect (NFA) as a key measure of emotional and thematic preferences in media introduces a novel dimension to personalisation. Understanding how users engage with and respond to the emotional and thematic aspects of content can inform the creation of more captivating and resonant narratives and experiences. One of the outcomes of our exploration is the development of a method for creating interactive narratives that capture user profiles featuring the FFM and the NFA. This method has demonstrated its effectiveness in personalising narratives, offering users tailored storytelling experiences. Nevertheless, this was not a complete, straightforward success, as the NFA personality test results had very little to do with seeking emotionally intense storylines, as the NFA framework would expect it to. It might be better to create a wholly new measure for seeking such intensity, perhaps to be called Preference for Tragedy.

Adapting characters and language based on user profiles has proven particularly impactful, with recent advancements in large language models simplifying the implementation of these adaptations. This user profile creation approach holds promise not only for interactive narratives but also for recommender systems. The ability to leverage user-generated text data further enhances the personalisation process. Users' own words can provide valuable insights into their preferences, allowing for more accurate and meaningful recommendations.

Despite the evident efficacy of the Five-Factor Model in understanding and predicting user preferences, the exploration of additional personality frameworks is imperative for advancing the field of personalised recommendations. Diversifying the repertoire of personality models, and potentially exploring hybrid approaches, holds the promise of achieving greater accuracy and nuance in personality-based recommendations. Furthermore, the integration of real-time personality assessments stands as a potential frontier for enhancing the depth and immediacy of personalisation efforts.

While the FFM has demonstrated its utility, personality is inherently complex and multidimensional. Exploring alternative personality frameworks allows for a more comprehensive understanding of individual differences. Different frameworks may capture unique facets of personality that complement or extend beyond the FFM. For instance, frameworks like the HEXACO model or even novel approaches specific to literature engagement could provide valuable insights into nuanced aspects of user preferences.

The prospect of combining multiple personality frameworks into hybrid models presents an intriguing avenue for enhancing the accuracy and granularity of personality assessments. By leveraging the strengths of various frameworks, hybrid approaches could offer a more holistic representation of an individual's personality profile. Integrating dimensions that might be overlooked by a single framework could contribute to a more nuanced and refined understanding of user preferences in the context of literary experiences.

The evolution of personalisation strategies could benefit significantly from the integration of real-time personality assessments. Traditional personality assessments often rely on static traits, providing a snapshot of an individual's personality at a specific point in time. Real-time assessments, on the other hand, have the potential to capture dynamic and evolving aspects of personality. Continuous monitoring of user interactions, preferences, and emotional states could contribute to a more adaptive and responsive personalisation system.

The challenge of accurately mapping behaviour within interactive narratives or games to real-life personality traits persists, as users might behave differently in different environments. Future research should aim to develop more precise and nuanced models that consider the dynamic nature of user or player behaviour. Leveraging advances in machine learning and data analysis can help bridge this gap effectively.

The exploration of new personality frameworks and real-time assessments is not without its challenges. Validating the effectiveness of alternative frameworks requires rigorous empirical research. Additionally, considerations regarding user privacy, consent, and ethical implications must be at the forefront of such endeavours. Striking a balance between the depth of personality insights and the responsible use of personal data is crucial for the ethical advancement of personalisation technologies.

Future research in this domain should embark on systematic investigations into alternative personality frameworks, evaluating their relevance and applicability to the domain of literature preferences. Comparative studies, pitting different frameworks against each other, can provide valuable insights into their respective strengths and limitations. Moreover, research efforts should be directed towards understanding the dynamics of real-time personality assessments in the context of literary engagement, exploring how these assessments can be seamlessly integrated into personalisation algorithms.

Finally, future avenues should include a more thorough assessment of all the different factors in literature that are relevant to it, and seeking ways to personalise all of these. One way to do this would be considering how important different factors are to different users, and what they would seek in each aspect, extending to issues such as genre preferences, narrative structure and reading goals. The emotional states sought should be another consideration, much depending on not just the individual's personality, but on their mood and context, as well. This type of personalisation could be used either in conjunction with or separately from personality-based personalisation.

Contributions

This thesis has provided an extensive literature review as a wide introduction to the topic of personalising narratives and conducted studies to progress on the topic, with findings including, but not limited to, the following topics:

1. **Interactive Narratives as Personality Tests:** By demonstrating the efficacy of capturing personality traits, particularly Extraversion and Emotional Stability, within interactive narratives, the research highlighted the potential of using such narratives as tools for personality assessment. It was highlighted that this is especially suitable for studying preferences within fiction.
2. **New Narrative Personalisation Approaches:** The studies presented new approaches in the field of narrative personalisation based on personality traits. It was notable that making the protagonist resemble the reader increased satisfaction in the story across the board.
3. **Insights into Reader Preferences and Writing Styles:** It was noted that Extraverted people enjoyed the story more if the language was less formal, and Introverted people enjoyed more formal language. People with different personalities also had different preferences with different writing styles, such as Shakespearean language. Finally, it was demonstrated that it is possible to recognise a person's personality from their own writing style.
4. **Addressing Challenges in Textual MBTI Recognition:** Nevertheless, the research also addressed challenges and limitations in the field of textual MBTI recognition, particularly regarding over-optimistic outcomes and inadequate data reporting. Moreover, the integration of diverse datasets and the application of machine learning models tailored for imbalanced data contribute to enhancing the comprehensiveness and accuracy of analyses.

Closing Thoughts

In concluding this thesis, we do so with a sense of anticipation for the next chapter in the evolving narrative of personalisation. It is a narrative driven by innovation, ethics, and user empowerment. The field of personalisation is not static; rather, it is a dynamic and ever-evolving terrain. Embracing this dynamism with open minds and a commitment to adaptability is paramount. The interdisciplinary nature of personalisation invites researchers, practitioners, and creatives from various domains to converge and share their expertise.

The potential impact of the outcomes derived from this exploration extends far beyond the confines of academia. It spans across various industries, including literature, film, music, and gaming. The ability to harness AI-driven personalisation offers unprecedented opportunities for creators to engage their audiences on a deeper level. Personalisation, in its true essence, is about creating connections and resonating with the unique essence of each individual. Empowering users to navigate and customise their experiences ensures that personalisation is a tool that enhances rather than diminishes the user's agency and autonomy. Whether it be in literature, gaming, film, or other creative domains, the quest is to enrich the user experience by tailoring content that aligns seamlessly with personal preferences and resonates on an emotional and cognitive level. The fusion of technology, data analytics, and user insights opens doors to possibilities previously unimagined. The potential to refine and elevate digital experiences, driven by a nuanced understanding of individual traits, preferences and behaviours promises a future where each interaction is imbued with personal significance.

BIBLIOGRAPHY

Abidin, N. H. Z., Remli, M. A., Ali, N. M., Phon, D. N. E., Yusoff, N., Adli, H. K. & A. H. B. (2020). Improving intelligent personality prediction using Myers-Briggs type indicator and random forest classifier. *International Journal of Advanced Computer Science and Applications, 11*(11).

Adomavičius, G. & Tuzhilin, A. (2011). Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender systems handbook* (pp. 217–253).

Al Masum Shaikh, M., Prendinger, H. & Ishizuka, M. (2010). Emotion Sensitive News Agent (ESNA): A system for user-centric emotion sensing from the news. *Web Intelligence and Agent Systems, 8*(4), 377–396.

Alharthi, H. (2015). The Use of Items Personality Profiles in Recommender Systems (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W. & Jacobs, A. M. (2012). The power of emotional valence—from cognitive to affective processes in reading. *Frontiers in Human Neuroscience, 6*, 192.

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W. & Jacobs, A. M. (2014). Fact vs fiction—how paratextual information shapes our reading processes. *Social Cognitive and Affective Neuroscience, 9*(1), 22–29.

Amirhosseini, M. H. & Kazemian, H. (2020). Machine learning approach to personality type prediction based on the Myers-Briggs type indicator. *Multimodal Technologies and Interaction, 4*(1).

Anderson, C. (2007). *The long tail: How endless choice is creating unlimited demand*.

André, E., Rist, T., Van Mulken, S., Klesen, M. & Baldes, S. (2000). The automated design of believable dialogues for animated presentation teams. In Embodied conversational agents (pp. 220–255).

Annalyn, N., Bos, M. W., Sigal, L. & Li, B. (2018). Predicting personality from book preferences with user-generated content labels. *IEEE Transactions on Affective Computing, 11*(3), 482–492.

Appel, M. (2008). Manche mögen's heiß. *Diagnostica, 54*(1), 2–15.

Appel, M., Gnambs, T. & Maio, G. R. (2012). A short measure of the need for affect. *Journal of Personality Assessment, 94*(4), 418–426.

Ardissono, L., Goy, A., Petrone, G., Segnan, M. & Torasso, P. (2003). Intrigue: Personalized recommendation of tourist attractions for desktop and handset devices. In *Applied Artificial Intelligence* (pp. 687–714).

Argamon, S., Dhawle, S., Koppel, M. & Pennebaker, J. W. (2005). Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America* (pp. 1-16).

Asabere, N. Y. & Acakpovi, A. (2020). ROPPSA: TV program recommendation based on personality and social awareness. *Mathematical Problems in Engineering, 2020*.

Asabere, N. Y., Acakpovi, A. & Michael, M. B. (2018). Improving socially-aware recommendation accuracy through personality. *IEEE Transactions on Affective Computing, 9*(3), 351–361.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., ... & De Raad, B. (2004). A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions From Psycholexical Studies in Seven Languages. *Journal of Personality and Social Psychology, 86*(2), 356–366.

Asmundson, G. J. G., Norton, G. R. & Jackobson, S. J. (1996). Social, blood/injury, and agoraphobic fears in patients with physically unexplained chronic pain: Are they clinically significant? *Anxiety, 2*, 28–33.

Aylett, R. S., Louchart, S., Dias, J., Paiva, A. & Vala, M. (2005). FearNot!–an experiment in emergent narrative. In *International Workshop on Intelligent Virtual Agents* (pp. 305-316). Springer.

Back, M. D., Schmukle, S. C. & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*(4), 1116-1122.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B. & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*(3), 372-374.

Bakkes, S. C. J., Spronck, P. H. M. & van Lankveld, G. (2012). Player Behavioural Modelling for Video Games. *Entertainment Computing, 3*(3), 71-79.

Bakkes, S., Tan, C. T. & Pisan, Y. (2012). Personalised gaming: A motivation and overview of literature. In *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System* (p. 4).

Balabanovic, M. & Shoham, Y. (1997). Content-based, collaborative recommendation. *Communications of the ACM, 40*(3), 66–72.

Balakrishnan, V. & Arabi, H. (2018). HyPeRM: A hybrid personality-aware recommender for movies. *Malaysian Journal of Computer Science, 31*(1), 48–62.

Barber, H. & Kudenko, D. (2007). A user model for the generation of dilemma-based interactive narratives. In *Workshop on Optimizing Player Satisfaction at AIIDE (Vol. 7)*.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research, 1*(1), 19.

Bartle, R. (2004). *Designing virtual worlds*.

Bartsch, A., Appel, M. & Storch, D. (2010). Predicting emotions and meta-emotions at the movies: The role of the need for affect in audiences' experience of horror and drama. *Communication Research, 37*(2), 167-190.

Bateman, C. & Boon, R. (2005). *21st Century Game Design (Game Development Series)*.

Bateman, C., Lowenhaupt, R. & Nacke, L. E. (2011). Player typology in theory and practice. In *DiGRA Conference*.

Batista, G. E., Prati, R. C. & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter, 6*(1), 20-29.

Batista, G. B., Bazzan, A. L. & Monard, M. C. (2003). Balancing training data for automated annotation of keywords: A case study. In *WOB* (pp. 10-18).

Batson, C. D., Early, S. & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin, 23*, 751–758.

Baumeister, R. F., Dale, K. & Sommer, K. L. (1998). Freudian defense mechanisms and empirical findings in modern social psychology: Reaction formation, projection, displacement, undoing, isolation, sublimation, and denial. *Journal of Personality, 66*, 1081–1124.

Beilin, L. & Yi, S. (2013). Survey of personalized recommendation based on society networks analysis. In *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering* (pp. 337-340). IEEE.

Beliakov, G., Calvo, T. & James, S. (2011). Aggregation of preferences in recommender systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 705–734). Springer.

Berkovsky, S., Kuflik, T. & Ricci, F. (2007). Cross-Domain Mediation in Collaborative Filtering. In C. Conati, K. McCoy & G. Paliouras (Eds.), *UM 2007* (pp. 355–359).

Berlyne, D. E. (1971). *Aesthetics and psychobiology*.

Bernardo, F., Zbyszynski, M., Fiebrink, R. & Grierson, M. (2017). Interactive machine learning for end-user innovation. *2017 AAAI Spring Symposium Series*.

Bernhaupt, R., Wilfinger, D., Weiss, A. & Tscheligi, M. (2008). An ethnographic study on recommendations in the living room: Implications for the design of itv recommender systems. In *EUROITV'08: Proceedings of the 6th European Conference on Changing Television Environments* (pp. 92–101). Springer-Verlag.

Bialas, M., Tekofsky, S. & Spronck, P. (2014). Cultural influences on play style. In *2014 IEEE Conference on Computational Intelligence and Games* (pp. 1-7). IEEE.

Birk, M. V., Toker, D., Mandryk, R. L. & Conati, C. (2015). Modeling motivation in a social network game using player-centric traits and personality traits. In *Int. Conf. on user modeling, adaptation, and personalization*. Springer, pp. 18–30.

Blizzard Entertainment. (2004). *World of Warcraft*. [Computer game].

Blom, P. M., Bakkes, S., Tan, C. T., Whiteson, S., Roijers, D., Valenti, R. & Gevers, T. (2014). Towards personalised gaming via facial expression recognition. *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Bolger, N., Davis, A. & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1), 579–616.

Bonnefoy, D., Bouzid, M., Lhuillier, N. & Mercer, K. (2007). More like this or not for me: Delivering personalised recommendations in multi-user environments. In *UM'07: Proceedings of the 11th international conference on User Modeling* (pp. 87–96).

Bontchev, B. & Georgieva, O. (2018). Playing style recognition through an adaptive video game. *Computers in Human Behavior, 82*, 136-147.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968-1987. *Psychological Bulletin, 106*(2), 265-289.

Bortolussi, M. & Dixon, P. (2003). Psychonarratology: foundations for the empirical study of literary response. Cambridge.

Bostan B. (2009). Player motivations: A psychological perspective. In *Computers in Entertainment, 7*(2), 26.

Bostan, B. & Marsh, T. (2010). The 'interactive' of interactive storytelling: customizing the gaming experience. In *International Conference on Entertainment Computing* (pp. 472-475).

Bottou, L., Curtis, F. E. & Nocedal, J. (2018). Optimization methods for large-scale machine learning. arXiv preprint arXiv:1606.04838.

Boyd, R. L. (2017). Psychological text analysis in the digital humanities. *Data analytics in digital humanities*, 161-189.

Breault, V., Ouellet, S. & Davies, J. (2021). Let CONAN tell you a story: Procedural quest generation. *Entertainment Computing, 38*, 100422.

Breese, J., Heckerman, D. & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Bross, I. D. (1958). How to use ridit analysis. *Biometrics, 14*, 18-38.

Brown, P. J., Bovey, J. D. & Chen, X. (1997). Context-aware applications: from the laboratory to the marketplace. *IEEE Personal Communications, 4*, 58–64.

Browne, C. B. (2008). Automatic generation and evaluation of recombination games (Doctoral dissertation, Queensland University of Technology).

Bullough, E. (1912). " 'Psychical distance' as a factor in art and an aesthetic principle". *British Journal of Psychology, 5*, 87-118.

Burke, R. (2000). Knowledge-Based Recommender Systems. *Encyclopedia of Library and Information Science, 69*(32), 180–200.

Busch, M., Mattheiss, E. E., Hochleitner, W., Hochleitner, C., Lankes, M., Fröhlich, P., Orji, R. & Tscheligi, M. (2016). Using Player Type Models for Personalized Game Design-An Empirical Investigation. In *IxD&A, 28*, 145-163.

Callison-Burch, C., Osborne, M. & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. In *11th conference of the european chapter of the association for computational linguistics* (pp. 249-256).

Canossa, A., Martinez, J. B. & Togelius, J. (2013). Give me a reason to dig Minecraft and psychology of motivation. In *2013 IEEE Conference on Computational Intelligence in Games* (pp. 1-8). IEEE.

Cantador, I. (2008). Exploiting the conceptual space in hybrid recommender systems: a semanticbased approach. Ph.D. thesis, Universidad Autonoma de Madrid (UAM), Spain.

Cantador, I., Fernández-Tobías, I., Bellogín, A., Kosinski, M. & Stillwell, D. (2013). Relating personality types with user preferences in multiple entertainment domains. In *UMAP Workshops*, 997.

Carlson, K., Riddell, A. & Rockmore, D. (2018). Evaluating prose style transfer with the bible. *Royal Society open science, 5*(10), 171920.

Cassell, J. & Bickmore, T. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User modeling and user-adapted interaction, 13*, 89-132.

Cattell, R. B. (1946). *Description and measurement of personality.*

CD Projekt Red. (2015). *The Witcher 3: Wild Hunt*. [Video game].

Celma, O. & Herrera, P. (2008). A new approach to evaluating novel recommendations. In *RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems* (pp. 179–186).

Chamorro-Premuzic, T., Kallias, A. & Hsu, A. (2013). Understanding individual differences in film preferences and uses: a psychographic approach. In *The social science of cinema*, 87.

Charles, D., McNeill, M., McAlister, M., Black, M., Moore, A., Stringer, K., Kücklich, J. & Kerr, A. (2005). Player-centred game design: Player modelling and adaptive digital games.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM, 50*(4), 31- 34.

Chen, C., Liaw, A. & Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley, 110, 1-12.

Chen, D. & Dolan, W. B. (2011). Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 190-200).

Chen, L., Chen, G. & Wang, F. (2015). Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction, 25*, 99–154.

Chen, L., Wu, W. & He, L. (2013). How personality influences users' needs for recommendation diversity? In *Proceeding of CHI'13 EA*.

Chen, L., Zeng, W., Yuan, Q. (2013). A unified framework for recommending items, groups and friends in social media environment via mutual resource fusion. *Expert Systems with Applications, 40*(8), 2889–2903.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

Cheng, Y., Chenyi, C., Lin, C-P., Chou, K-H. and Decety, J. (2010). Love hurts: an fMRI study. *NeuroImage, 51*, 923–929.

Cheng, Y., Lin, C-P., Liu, H-L., Hsu, Y-Y., Lim, K-E. and Hung, D. (2007). Expertise modulates the perception of pain in others. *Current Biology, 17*, 1708–1713.

Choong, E. J. & Varathan, K. D. (2021). Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum. *PeerJ, 9*, e11382.

Christodoulou, E., Gregoriades, A., Herodotou, H. & Pampaka, M. (2022). Extracting user preferences and personality from text for restaurant recommendation. *CEUR Workshop Proceedings*.

Celli, F. & Lepri, B. (2018). Is big five better than MBTI? A personality computing challenge using Twitter data. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it* (Vol. 2018, pp. 93-98).

Celli, F. & Rossi, L. (2012). The role of Emotional Stability in Twitter conversations. In *Proceedings of the workshop on semantic analysis in social media* (pp. 10-17).

Chung, C. K. & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of research in personality, 42*(1), 96-132.

Ciechanowski, P. S., Walker, E. A., Katon, W. J. and Russo, J. E. (2002). Attachment theory: a model for health care utilization and somatization. *Psychosomatic Medicine, 64*, 660–667.

Cikara, M., Botvinick, M. M. and Fiske, S. T. (2011). Us versus them: social identity shapes responses to intergroup competition and harm. *Psychological Science, 22*, 306–313.

Collishaw, M. (2017). *Thresholds* [Virtual reality installation].

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Costa, P. T. & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4*(1), 5-13.

Costa, P. T. & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment, 2*(2), 179-198.

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21-27.

Cowley, B., Charles, D., Black, M. & Hickey, R. (2006). User-system-experience model for user centered design in computer games. In *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 419–424).

Cowley, B., Charles, D., Black, M. & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment (CIE), 6*(2), 20.

Cowley, B., Charles, D., Black, M. & Hickey, R. (2013). Real-time rule-based classification of player types in computer games. *User Modeling and User-Adapted Interaction, 23*(5), 489-526.

Cui, B. & Qi, C. (2017). Survey analysis of machine learning methods for natural language processing for MBTI Personality Type Prediction. Retrieved from http://cs229.stanford.edu/proj2017/finalreports/5242471.pdf

Cupchik, G. C. (2002). The evolution of psychical distance as an aesthetic concept. *Culture and Psychology, 8*(2), 155-187.

Darley, W. K. & Lim, J. S. (1992). The effect of consumers' emotional reactions on behavioral intention: The moderating role of personal relevance and self-monitoring. *Psychology and Marketing,9*(4), 329–346.

Das, K. & Prajapati, H. (2020). Personality identification based on MBTI dimensions using natural language processing. *International Journal of Creative research Thoughts, 8*(6), 1653–1657.

Dasborough, M.T., Sinclair, M., Russell-Bennett, R. & Tombs, A. (2008). Measuring emotion - Methodological issues and alternatives, in Ashkanasy, N. M. & Cooper, C. L. (eds.), *Research companion to emotions in organizations*, 197–208.

de Castell, S., Taylor, N., Weiler, M. & Jenson, J. (2011). Theoretical and Methodological Challenges in Virtual Worlds Research. In *Videogame Cultures Conference (Vol. 3)*.

de Lima, E., Feijó, B. & Furtado, A. L. (2014). Hierarchical generation of dynamic and nondeterministic quests in games. In *Proceedings of the 11th conference on advances in computer entertainment technology* (pp. 1-10).

de Lima, E. S., Feijó, B. & Furtado, A. L. (2018). Player behavior and personality modeling for interactive storytelling in games. *Entertainment Computing, 28*, 32-48.

de Vette, A. F. A., Tabak, M., Dekker-van Weering, M. G. H. & Vollenbroek-Hutten, M. M. (2016). Exploring Personality and Game Preferences in the Younger and Older Population: A Pilot Study. In *ICT4AgeingWell* (pp. 99-106).

De Wied, M., Zillmann, D. & Ordman, V. (1994). The role of empathic distress in the enjoyment of cinematic tragedy. *Poetics, 23*, 91-106.

Decety, J., Echols, S. C. & Correll, J. (2009). The blame game: The effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience, 22*, 985–997.

Denden, M., Tlili, A., Essalmi, F. & Jemni, M. (2017). Educational gamification based on personality. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1399-1405).

Denisova, A. (2016). Adaptive technologies in digital games: The influence of perception of adaptivity on immersion (Doctoral dissertation, University of York).

Denisova, A. & Cairns, P. (2015). Adaptation in digital games: The effect of challenge adjustment on player performance and experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 97-101). ACM.

Derrick, J. L. (2012). Energized by television: Familiar fictional worlds restore self-control. *Social Psychological and Personality Science, 4*(3), 299-307.

Dewaele, J-M. (2007). “The effect of multilingualism, sociobiographical, and situational factors on communicative anxiety and foreign language anxiety of mature language learners.” *International journal of bilingualism* 11(4), 391-409.

Dhelim, S., Aung, N., Bouras, M. A., Ning, H. & Cambria, E. (2022). A survey on personality-aware recommendation systems. *Artificial Intelligence Review, 1-46*.

Dhelim, S., Chen, L. L., Aung, N., Zhang, W. & Ning, H. (2021). Big-Five, MPTI, Eysenck or HEXACO: The ideal personality model for personality-aware recommendation systems. *arXiv preprint arXiv:2106.03060*.

Dias, R. & Martinho, C. (2011). Adapting content presentation and control to player personality in videogames. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE '11)*.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*: 417–40.

Dijkstra, A. (2014). The persuasive effects of personalization through name mentioning in a smoking cessation message. *User Modeling and User-Adapted Interaction, 24*(5), 393-411.

Donnellan, M. B. & Lucas, R. E. (2008). Age differences in the Big Five across the life span: evidence from two national samples. *Psychology and aging*, *23*(3), 558.

Dormans, J. (2006). The Hacker: New mythical content of Narrative games. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 104-115).

Ducheneaut, N. (2010). Massively multiplayer online games as living laboratories: Opportunities and pitfalls. In *Online worlds: Convergence of the real and the virtual* (pp. 135-145).

Dunn, G., Wiersema, J., Ham, J. & Aroyo, L. (2009). Evaluating interface variants on personality acquisition for recommender systems. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 259-270).

Duttweiler, P. (1984). The internal control index: A newly developed measure of locus of control. *Educational and Psychological Measurement, 44*, 209–221.

Ehnvall, A., Mitchell, P. B., Hadzi-Palovic, D., Malhi, G. S. & Parker, G. (2009). Pain during depression and relationship to rejection sensitivity. *Acta Psychiatrica Scandinavica, 119*, 375–382.

Eisenberger, N. I. (2006). Social connection and rejection across the lifespan: A social cognitive neuroscience approach to developmental processes. *Annual Review of Psychology, 58*, 259-289.

Eisenberger, N. I. (2012). The pain of social disconnection: Examining the shared neural underpinnings of physical and social pain. *Nature Reviews Neuroscience, 13*(6), 421-34.

El-Nasr, M. S. (2007). Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories. *Interaction Studies, 8*(2), 209-240.

El-Nasr, M. S., Vasilakos, A., Rao, C. & Zupko, J. (2009). Dynamic intelligent lighting for directing visual attention in interactive 3-D scenes. *IEEE Transactions on Computational Intelligence and AI in Games, 1*(2), 145-153.

Elahi, M., Braunhofer, M., Ricci, F. & Tkalcic, M. (2013). Personality-based active learning for collaborative filtering recommender systems. In *AIIA 2013: Advances in Artificial Intelligence* (pp. 360–371).

Eliasson, O. (2003). *The Weather Project* [Art installation, Tate Modern].

Etheredge, M., Lopes, R. & Bidarra, R. (2013). A generic method for classification of player behavior. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Fang, Q., Giachanou, A., Bagheri, A., Boeschoten, L., van Kesteren, E. J., Kamalabad, M. S. & Oberski, D. L. (2022). On Text-based Personality Computing: Challenges and Future Directions. *arXiv preprint arXiv:2212.06711*.

Farseev, A., Yang, Q., Filchenkov, A., Lepikhin, K., ChuFarseeva, Y.-Y. & Loo, D.-B. (2020). *SoMin.ai: Personality-driven content generation platform.*

Felfernig, A. & Gula, B. (2006). An empirical study on consumer behavior in the interaction with knowledge-based recommender applications. *Proceedings of the 8th IEEE International Conference on E-Commerce Technology (CEC 2006) / Third IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (EEE 2006)*, 37.

Feng, H. & Qian, X. (2013). Recommendation via user's personality and social contextual. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management (CIKM '13)* (pp. 1521–1524).

Fennis, B. M. & Pruyn, A. T. H. (2007). You are what you wear: Brand personality influences on consumer impression formation. *Journal of Business Research, 60*(6), 634-639.

Fernández-Tobías, I. & Cantador, I. (2014). Personality-aware collaborative filtering: An empirical study in multiple domains with Facebook data. In *E-Commerce and Web Technologies* (pp. 125–137).

Fernández-Tobías, I. & Cantador, I. (2016). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction, 26*(2–3), 221–255.

Fernau, D., Hillmann, S., Feldhus, N. & Polzehl, T. (2022). Towards automated dialog personalization using MBTI personality indicators. In *Proc. Interspeech, 1968-1972*.

Ferwerda, B., Graus, M. P., Vall, A., Tkalčič, M. & Schedl, M. (2016). The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In *EMPIRE@ RecSys* (pp. 43-47).

Ferwerda, B., Yang, E., Schedl, M. & Tkalčič, M. (2019). Personality and taxonomy preferences, and the influence of category choice on the user experience for music streaming services. *Multimedia Tools and Applications*.

Fischer, P., Kastenmüller, A. & Greitemeyer, T. (2010). Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior. *Journal of Experimental Social Psychology, 46*(1), 192– 195.

Fu, Z., Tan, X., Peng, N., Zhao, D. & Yan, R. (2018). Style transfer in text: Exploration and evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1)*.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science, 21*(3), 1–18.

Gadiyar, A. R. (2014). Gamification 3.0: The power of personalization. White paper. Cognizant's Global Technology.

Ganu, G., Kakodkar, Y. & Marian, A. (2013). Improving the quality of predictions using textual information in online user reviews. *Information Systems, 38*(1), 1–15.

Garcia Esparza, S., O'Mahony, M. P. & Smyth, B. (2010). Effective product recommendation using the real-time web. In *Bramer, M., Petridis, M. & Hopgood, A. (Eds.)* Proceedings of the 30th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (pp. 5-18). Cambridge, UK, Springer.

Garrido, S. & Schubert, E. (2011). Negative emotion in music: What is the attraction? A qualitative study. *Empir. Musicol. Rev., 6*, 214–230.

Gerrig, R. J. (1990). The construction of literary character: A view from cognitive psychology. *Style 24*, 380–92.

Giannatos, S., Cheong, Y. G., Nelson, M. & Yannakakis, G. N. (2012). Generating narrative action schemas for suspense. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 8, No. 2, pp. 8-13).

Giannatos, S., Nelson, M., Cheong, Y. G. & Yannakakis, G. (2011). Suggesting new plot elements for an interactive story. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 7, No. 2, pp. 25-30).

Gladis, S. D. (1993). Writetype: Personality types and writing styles. *Human Resource Development*.

Golbeck, J. (2006). Generating predictive movie recommendations from trust in social networks. In *International Conference on Trust Management* (pp. 93-104).

Golbeck, J. & Norris, E. (2013). Personality, movie preferences, and recommendations. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1414-1415).

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several Five-Factor Models. *Personality psychology in Europe, 7*(1), 7-28.

Gjurković, M. & Šnajder, J. (2018). Reddit: A gold mine for personality prediction. In *Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media* (pp. 87-97).

Gómez-Gauchía, H. & Peinado, F. (2006). Automatic customization of non-player characters using players temperament. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 241-252).

Gonzalez, G., Lopez, B. & Rosa, J.L.D.L. (2002). The Emotional Factor: An Innovative Approach to User Modelling for Recommender Systems, in *AH 2002, Workshop on Recommendation and Personalization in e-Commerce*, 90–99.

Gorgoglione, M., Panniello, U. & Tuzhilin, A. (2019). Recommendation strategies in personalization applications. *Information & Management*.

Grant, A. (2013). Goodbye to MBTI, the fad that won't die. *Psychology Today*, 9.

Grasbon, D. & Braun, N. (2001). A morphological approach to interactive storytelling. In *Proc. CAST01, Living in Mixed Realities*. Special issue of Netzspannung. org/journal, the Magazine for Media Production and Inter-media Research (pp. 337-340).

Grasso, A. & Bergholz, A. (2007). U.S. Patent No. 7,240,055. Washington, DC: U.S. Patent and Trademark Office.

Green, D. (2016). *Procedural content generation for C++ game development*.

Gretzel, U. & Fesenmaier, D. R. (2007). Persuasion in Recommender Systems, *International Journal of Electronic Commerce 11*(2), 81-100.

Groh, G. & Ehmig, C. (2007). Recommendations in taste related domains: collaborative filtering vs. social filtering, *GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work*, 127–136. ACM, New York, NY, USA.

Gunter, B. (1985). *The dimensions of television violence*.

Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S. & Ofek-Koifman, S. (2009). Personalized recommendation of social software items based on social relations, in *RecSys '09: Proceedings of the third ACM conference on Recommender systems*, 53–60.

Hallifax, S., Serna, A., Marty, J. C., Lavoué, G. & Lavoué, E. (2019). Factors to Consider for Tailored Gamification. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 559-572).

Hamari, J. & Tuunanen, J. (2014). Player types: A meta-synthesis. In *Transactions of the Digital Games Research* 1, 2.

Hamburger, Y. A. & Ben-Artzi, E. (2000). The relationship between Extraversion and neuroticism and the different uses of the Internet. *Computers in human behavior, 16*(4), 441-449.

Harman, J. L. & Purl, J. (2022). Advances in Gamified Personality Assessment. *Trends in Psychology*.

Harman, J. L. & Brown*, K. (2022). Illustrating a narrative: A test of game elements in game-like personality assessment. *International Journal of Selection and Assessment*.

Hastings, E. J., Guha, R. K. & Stanley, K. O. (2009). Automatic content generation in the galactic arms race video game. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4), 245-263.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D. & Singer, T. (2010) "Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping", *Neuron 68*: 149–160.

Herlocker, J.L., Konstan, J.A., Terveen, L.G. & Riedl, J.T. (2004). Evaluating collaborative filtering recommender systems, *ACM Transaction on Information Systems 22*(1), 5–53.

Hernández del Olmo, F. & Gaudioso, E. (2008). Evaluation of recommender systems: A new approach, *Expert Systems with Applications 35*, 790-804.

Hirsh, J. B., Kang, S. K. & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological science, 23*(6), 578-581.

Hoffner, C. A. & Levine, K. J. (2005). Enjoyment of mediated fright and violence: A meta-analysis. *Media Psychology, 7*(2), 207-237.

Hogan, P. C. (2001). The epilogue of suffering: heroism, empathy, ethics. *SubStance, 30*, 119–143.

Hollingdale, J. & Greitemeyer, T. (2013). The changing face of aggression: The effect of personalized avatars in a violent video game on levels of aggressive behavior. *Journal of Applied Social Psychology, 43*(9), 1862-1868.

Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied logistic regression.*

Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 591-598.

Hsu, C-T., Jacobs A. M. & Conrad M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex, 63*, 282–295.

Hu, R. & Pu, P. (2009). Acceptance issues of personality-based recommender systems. In RecSys '09: Proceedings of the third ACM conference on Recommender systems (pp. 221–224). ACM, New York, NY, USA.

Hu, R. & Pu, P. (2010). A study on user perception of personality-based recommender systems. In International Conference on User Modeling, Adaptation, and Personalization (pp. 291-302).

Hu, R. & Pu, P. (2011). Enhancing collaborative filtering systems with personality information. In Proceedings of the fifth ACM conference on Recommender systems (pp. 197-204).

Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R. & Xing, E. P. (2017). Toward controlled generation of text. In International conference on machine learning (pp. 1587-1596).

Huang, S. L. (2011). Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods. *Electronic Commerce Research and Applications, 10*(4), 398-407.

Huron, D. (2008). *Sweet anticipation: music and the psychology of expectation.*

Iacobelli, F., Gill, A. J., Nowson, S. & Oberlander, J. (2011). Large scale personality classification of bloggers. In Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part II (pp. 568-577).

Jain, D., Kumar, A. & Beniwal, R. (2022). Personality BERT: A Transformer-Based Model for Personality Detection from Textual Data. In Proceedings of International Conference on Computing and Communication Networks: ICCCN 2021 (pp. 515-522).

Jannach, D., Zanker, M., Felfernig, A. & Friedrich, G. (2011). *Recommender systems: an introduction.*

Jeong, W., Kim, J., Park, D. & Kwak, J. (2013). Performance Improvement of a Movie Recommendation System based on Personal Propensity and Secure Collaborative Filtering. *Journal of Information Processing Systems, 9*(1).

Jhamtani, H., Gangal, V., Hovy, E. & Nyberg, E. (2017). Shakespearizing modern language using copy-enriched sequence-to-sequence models. arXiv preprint arXiv:1707.01161.

Jiang, H., Zhang, X. & Choi, J. D. (2020). Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract). In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 10, pp. 13821-13822).

Jin, Z., Jin, D., Mueller, J., Matthews, N. & Santus, E. (2019). IMaT: Unsupervised text attribute transfer via iterative matching and translation. arXiv preprint arXiv:1901.11333.

John, V., Mou, L., Bahuleyan, H. & Vechtomova, O. (2018). Disentangled representation learning for non-parallel text style transfer. arXiv preprint arXiv:1808.04339.

John, O., Donahue, E. & Kentle, R. (1991). The big five inventory: Versions 4a and 54, institute of personality and social research, University of California, Berkeley, CA.

John, O., Naumann, L. & Soto, C. (2008). Paradigm shift to the integrative big five trait taxonomy. In Handbook of personality: Theory and research 3 (pp. 114–158).

Jose, P. E. & Brewer, W. F. (1984). Development of story liking: character identification, suspense, and outcome resolution. *Developmental Psychology, 20*(5), 911-924.

Jung, C. (1921). *Psychological Types*.

Kanjo, E., Al-Husain, L. & Chamberlain, A. (2015). Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing, 19*, 1197-1212.

Kaptein, M., De Ruyter, B., Markopoulos, P. & Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(2), 10.

Karumur, R. P., Nguyen, T. T. & Konstan, J. A. (2018). Personality, user preferences and behavior in recommender systems. *Information Systems Frontiers, 20*(6), 1241-1265.

Kaya, H. & Alpaslan, F. N. (2010). Using social networks to solve data sparsity problem in one-class collaborative filtering. In *2010 Seventh International Conference on Information Technology: New Generations* (pp. 249-252).

Keen, S. (2006). A theory of narrative empathy. *Narrative, 14*(3), 207–236.

Keh, S. S. & Cheng, I. (2019). Myers-Briggs personality classification and personality-specific language generation using pre-trained language models. arXiv preprint arXiv:1907.06333.

Keirsey, D. (1998). *Please understand me II: Temperament, character, intelligence.*

Keirsey, D. & Bates, M. M. (1984). *Please understand me.*

Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arif, A. & Khalid, H. A. (2020a). Personality classification from online text using a machine learning approach. *International Journal of Advanced Computer Science and Applications, 11*(3), 460–476.

Khan, E. M., Mukta, M. S. H., Ali, M. E. & Mahmud, J. (2020b). Predicting users' movie preference and rating behavior from personality and values. *ACM Transactions on Interactive Intelligent Systems (TiiS), 10*(3), 1-25.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.*

Kline, S., Dyer-Witheford, N. & De Peuter, G. (2003). *Digital play: The interaction of technology, culture, and marketing.*

Kobsa, A. (1993). User modeling: Recent work, prospects and hazards. *Human Factors in Information Technology*, *10*, 111-111.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31.

Koenigstein, N. & Koren, Y. (2013). Towards scalable and accurate item-oriented recommendations. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 419-422).

Koopman, E. M. (2016). *Reading Suffering: an empirical inquiry into empathic and reflective responses to literary narratives* (PhD thesis). Erasmus University Rotterdam.

Koopman, E. M., Hilscher, M. & Cupchik, G. C. (2012). Reader responses to literary depictions of rape. *Psychology of Aesthetics, Creativity, and the Arts, 6*(1), 66-73.

Komiak, S. Y. X., Wang, W. & Benbasat, I. (2005). Trust Building in Virtual Salespersons Versus in Human Salespersons: Similarities and Differences. *e-Service Journal, 3*(3), 49-63.

Koren, Y., Bell, R. M. & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer, 42*(8), 30–37.

Košir, A., Odić, A., Kunaver, M., Tkalčič, M. & Tasič, J. (2011). Database for contextual personalization. *Elektrotehniški vestnik, 78*(5), 270–274.

Kozak, M. N., Marsh, A. A. & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology, 90*, 543–555.

Krath, J., Altmeyer, M., Tondello, G. F. & Nacke, L. E. (2023). Hexad-12: Developing and Validating a Short Version of the Gamification User Types Hexad Scale. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

Krishna, K., Wieting, J. & Iyyer, M. (2020). *Reformulating unsupervised style transfer as paraphrase generation.* arXiv preprint arXiv:2010.05700.

Kristaf, A. L. (1996). Person-Organization Fit: An Integrative Review of Its Conceptualization Measurement and Implication. *Personnel Psychology, 49*, 1-49.

Kumar, A., Beniwal, R. & Jain, D. (2023). Personality detection using kernel-based ensemble model for leveraging social psychology in online networks. *ACM Transactions on Asian and Low-Resource Language Information Processing, 22*(5), 1–20.

Kumar, K. N. P. & Gavrilova, M. L. (2022). Latent personality traits assessment from social network activity using contextual language embedding. *IEEE Transactions on Computational Social Systems, 9*(2), 638–649.

Kyatric, I. (2013). Bartle's Taxonomy of Player Types (And Why It Doesn't Apply to Everything). *Game Development Tuts+*. Retrieved from http://gamedevelopment.tutsplus.com/articles/bartles-taxonomy-of-player-types-and-why-it-doesnt-apply-to-everything--gamedev-4173.

Lamm, C., Decety, J. & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*, 2492–2502.

Lample, G., Subramanian, S., Smith, E. M., Denoyer, L., Ranzato, M. & Boureau, Y. (2019). Multiple-attribute text rewriting. In *Proceedings of the 7th International Conference on Learning Representations, ICLR 2019*, New Orleans, LA, May 6–9, 2019.

Landers, R. N., Auer, E. M., Collmus, A. B. & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming, 49*(3), 315–337.

Landers, R. N. & Callan, R. C. (2012). Training evaluation in virtual worlds: Development of a model. *Journal of Virtual Worlds Research, 5*, 1–20.

Lang, P. J., Levin, D. N., Miller, G. A. & Kozak, M. J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective response integration. *Journal of Abnormal Psychology, 92*(3), 276–306.

Lathia, N., Hailes, S., Capra, L. & Amatriain, X. (2010). Temporal Diversity in Recommender Systems. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 210–217.

Laursen, R. & Nielsen, D. (2005). *Investigating small scale combat situations in real-time-strategy computer games* (Master's thesis). Department of Computer Science, University of Aarhus, Aarhus, Denmark.

Lazarus, R. S. (1991). *Emotion and adaptation*.

Lazarus, R. S. (2001). Relational meaning and discrete emotions. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 37–67).

Lemaitre, G., Nogueira, F. & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research, 18*(17), 1–5.

Leung, C. W. K., Chan, S. C. F. & Chung, F. (2006). Integrating collaborative filtering and sentiment analysis: A rating inference approach. In *Proceedings of the ECAI 2006 Workshop on Recommender Systems*, 62–66.

Lhuillier, N., Bouzid, M., Mercer, K. & Picault, J. (2006). System for content item recommendation. Patent application GB2438645.

Li, C., Hancock, M., Bowles, B., Hancock, O., Perg, L., Brown, P., ... & Wade, R. (2018). Feature extraction from social media posts for psychometric typing of participants. In *Augmented Cognition: Intelligent Technologies: 12th International Conference, AC 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I* (pp. 267-286).

Liapis, A., Yannakakis, G. N. & Togelius, J. (2011). Optimizing visual properties of game content through neuroevolution. In *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*.

Liapis, A., Yannakakis, G. N. & Togelius, J. (2012). Adapting models of visual aesthetics for personalized content creation. *IEEE Transactions on Computational Intelligence and AI in Games, 4*(3), 213–228.

Liebroth, A. (2015). Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture, 10*(3), 229–248.

Lidén, L. (2004). Artificial Stupidity: The Art of Making Intentional Mistakes. In S. Rabin (Ed.), *AI Game Programming Wisdom 2* (pp. 41–48).

Linden, G., Smith, B. & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*, 76–80.

Lindley, C. A. & Sennersten, C. C. (2006). Game play schemas: From player analysis to adaptive game mechanics. In *Proceedings of the 2006 International Conference on Game Research and Development* (pp. 47–53).

Liu, A., Wang, A. & Okazaki, N. (2022). Semi-supervised formality style transfer with consistency training. *arXiv preprint arXiv:2203.13620*.

Liu, X. Y., Wu, J. & Zhou, Z. H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39*(2), 539–550.

Logeswaran, L., Lee, H. & Bengio, S. (2018). Content preserving text generation with attribute controls. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett (Eds.), *Advances in Neural Information Processing Systems, vol. 31*.

Logg, J. M. (2016). *When do people rely on algorithms?* (Doctoral dissertation). UC Berkeley.

Loiacono, D., Cardamone, L. & Lanzi, P. L. (2011). Automatic track generation for high-end racing games using evolutionary computation. *IEEE Transactions on Computational Intelligence and AI in Games, 3*(3), 245–259.

Lops, P., de Gemmis, M. & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 73–105).

Louchart, S., Aylett, R., Kriegel, M., Dias, J., Figueiredo, R. & Paiva, A. (2008). Authoring emergent narrative-based games. *Journal of Game Development, 3*(1), 19–37.

Lucas, K. & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication Research, 31*(5), 499–523.

Lukito, L. C., Erwin, A., Purnama, J. & Danoekoesoemo, W. (2016). Social media user personality classification using computational linguistic. In *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 1–6).

Lury, C. & Day, S. (2019). Algorithmic personalization as a mode of individuation. *Theory, Culture & Society, 36*(2), 17–37.

Luyckx, K. & Daelemans, W. (2008). Authorship attribution and verification with many authors and limited data. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)* (pp. 513–520).

Lynn, V., Balasubramanian, N. & Schwartz, H. A. (2020). Hierarchical modeling for user personality prediction: The role of message-level attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5306–5316).

MacDonald, G. & Kingsbury, R. (2006). Does physical pain augment anxious attachment? *Journal of Social and Personal Relationships, 23*, 291–304.

MacDonald, G. & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Review, 131*, 202–223.

Magerko, B. (2005). Story representation and interactive drama. In *AIIDE* (pp. 87–92).

Magerko, B., Heeter, C., Fitzgerald, J. & Medler, B. (2008). Intelligent adaptation of digital game-based learning. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (pp. 200–203).

Mahmood, T. & Ricci, F. (2007). Learning and adaptivity in interactive recommender systems. In *Proceedings of the 9th International Conference on Electronic Commerce (ICEC'07)* (pp. 75–84).

Maio, G. R. & Esses, V. M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of Personality, 69*, 583–615.

Mairesse, F. & Walker, M. A. (2010). Towards personality-based user adaptation: Psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction, 20*(3), 227–278.

Mairesse, F., Walker, M. A., Mehl, M. R. & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

Mani, I. & Zhang, I. (2003). kNN approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets* (Vol. 126, No. 1, pp. 1-7).

Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G. & Symeonidis, P. (2011). Social tagging recommender systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 615–644).

Markey, P. M. & Markey, C. N. (2010). Vulnerability to violent video games: A review and integration of personality research. *Review of General Psychology, 14*(2), 82–91.

Martin, A., Lim, A., Colton, S. & Browne, C. (2010). Evolving 3D buildings for the prototype video game subversion. In *European Conference on the Applications of Evolutionary Computation* (pp. 111–120).

Martín Sujo, J. C. (2023). BRAIN L: A book recommender system. *arXiv preprint arXiv:2302.00653*.

Martín Sujo, J. C. & Golobardes i Ribé, E. (2022). Personality Profile of Fictional Characters in Books Using Natural Language Processing. *Personality Profile of Fictional Characters in Books Using Natural Language Processing*.

Massa, P. & Avesani, P. (2004). Trust-aware collaborative filtering for recommender systems. In *Proceedings of the International Conference on Cooperative Information Systems, CoopIS* (pp. 492–508).

Mateas, M. & Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference* (Vol. 2, pp. 4-8).

Maulidah, M. & Pardede, H. F. (2021). Prediction of Myers-Briggs type indicator personality using long short-term memory. *Jurnal Elektronika dan Telekomunikasi, 21*(2), 104–111.

McCarthy, K., Salamo, M., Coyle, L., McGinty, L., Smyth, B. & Nixon, P. (2006). Cats: A synchronous approach to collaborative group recommendation. In G. Sutcliffe & R. Goebel (Eds.), *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference* (pp. 86–91).

McCord, J. L., Harman, J. L. & Purl, J. (2019). Game-like personality testing: An emerging mode of personality assessment. *Personality and Individual Differences, 143*, 95–102.

McCrae, R. R. & Costa Jr, P. T. (1989). Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40.

McDonnell, M., Owen, J. E. & Bantum, E. O. C. (2020). Identification of emotional expression with cancer survivors: Validation of linguistic inquiry and word count. *JMIR Formative Research, 4*(10), e18246.

McIntyre, N. & Lapata, M. (2010). Plot induction and evolutionary search for story generation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 1562-1572).

McMahon, N., Wyeth, P. & Johnson, D. (2012). Personality and player types in Fallout New Vegas. In *Proceedings of the 4th International Conference on Fun and Games* (pp. 113-116).

McNee, S. M., Lam, S. K., Konstan, J. A. & Riedl, J. (2003). Interfaces for eliciting new user preferences in recommender systems. *User Modeling 2003, LNCS 2702*, 178-187.

McNee, S.M., Riedl, J. & Konstan, J.A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems* (pp. 1097–1101).

McStay, A. (2018). *Emotional AI: The rise of empathic media*.

Medler, B. (2009). Using recommendation systems to adapt gameplay. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 1(3), 68-80.

Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E. & Eetemadi, S. (2020). Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 1184–1189).

Melamed, I. D., Green, R. & Turian, J. (2003). Precision and recall of machine translation. In *Companion volume of the proceedings of HLT-NAACL 2003-short papers* (pp. 61-63).

Meow Wolf. (2016). *The House of Eternal Return* [Interactive multimedia installation].

Meyer, M. (1903). Experimental studies in the psychology of music. *American Journal of Psychology, 14*, 456-475.

Miall, D. S. (2011). Enacting the other: towards an aesthetics of feeling in literary reading. In P. Goldie & E. Schellekens (Eds.), *The aesthetic mind: philosophy and psychology* (pp. 285-298). Oxford.

Miall, D.S. and Kuiken, D. (1999). What is literariness? Three components of literary reading. *Discourse Processes, 28*(2), 121-138.

Michelson, D. (2014). Personality and the varieties of fictional experience. *Journal of Aesthetic Education, 48*(2), 64-85.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS 2013*.

Mojang (2011). *Minecraft*. [Computer game].

Myers, I. B. (1962). *The Myers-Briggs Type Indicator: Manual*.

Nacke, L. E., Bateman, C. & Mandryk, R. L. (2011). BrainHex: preliminary results from a neurobiological gamer typology survey. In *International conference on entertainment computing* (pp. 288-293).

Nadal, A. J. C., Reyes, A. S., Blanco, M. C. R., Alipio, A. J. & Cortez, D. M. A. (2022). Enhancement of collaborative filtering using Myers-Briggs Type Indicator (MBTI) applied in recommendation system. *South Asian Journal of Engineering and Technology, 12*(1), 104–110.

Nagle, A., Wolf, P. & Riener, R. (2016). Towards a system of customized video game mechanics based on player personality: Relating the Big Five personality traits with difficulty adaptation in a first-person shooter game. *Entertainment computing, 13*, 10-24.

Nairat, M., Dahlstedt, P. & Nordahl, M. G. (2011). Character evolution approach to generative storytelling. In *2011 IEEE Congress of Evolutionary Computation (CEC)* (pp. 1258-1263).

Nairat, M., Dahlstedt, P. & Nordahl, M. G. (2013). Story characterization using interactive evolution in a multi-agent system. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design: Second International Conference, EvoMUSART 2013* (pp. 168-179).

Namaste Entertainment (2012). *Storybricks*. [Computer game].

Neal, M. (2012). Creating and maintaining psychological flow state in augmented reality applications. In *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Nikolov, N. I. & Hahnloser, R. H. (2018). Large-scale hierarchical alignment for data-driven text rewriting. *arXiv preprint arXiv:1810.08237*.

Nguyen, T.T., Hui, P-M, Harper, F.M., Terveen, L. & Konstan, J.A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity, *Proceedings of the 23rd international conference on World wide web*, 677-686.

Nowack, K. (1996). Is the Myers Briggs Type Indicator the right tool to use? Performance in practice. *American Society of Training and Development, Fall*, 6.

Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Human Communication Research, 19*(3), 315-342.

Oliver, M.B. & Raney, A.A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, *61*(5), 984-1004.

Oliver, M. B., Weaver, J. B. & Sargent, S. L. (2000). An examination of factors related to sex differences in enjoyment of sad film. *Journal of Broadcasting and Electronic Media, 44*, 282-300.

Ong, T. & Leggett, J. J. (2004). A genetic algorithm approach to interactive narrative generation. In *Proceedings of the fifteenth ACM conference on Hypertext and hypermedia* (pp. 181-182).

Onori, M., Micarelli, A. & Sansonetti, G. (2016). A Comparative Analysis of Personality-Based Music Recommender Systems. In *Empire@ RecSys* (pp. 55-59).Orestis, N. & Christos, T. (2017).

The 50/50 recommender: A method incorporating personality into movie recommender systems.

Orji, R. (2014). *Design for behaviour change: A model-driven approach for tailoring persuasive technologies* (Doctoral dissertation, University of Saskatchewan).

Orji, R., Mandryk, R. L., Vassileva, J. & Gerling, K. M. (2013). Tailoring persuasive health games to gamer type. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2467-2476).

Orkin, J. & Roy, D. (2007). The restaurant game: Learning social behavior and language from thousands of players online. *Journal of Game Development, 3*(1), 39-60.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., ... & Hall, L. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 194-201).

Pang, R. Y. (2019). Towards actual (not operational) textual style transfer auto-evaluation. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)* (pp. 444-445).

Papineni, K., Roukos, S., Ward, T. & Zhu, W. J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*.

Pazzani, M. J. & Billsus, D. (2007). Content-based recommendation systems. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The Adaptive Web* (pp. 325–341).

Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review, 13*, 393–408.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peinado, F. & Gervás, P. (2004). Transferring game mastering laws to interactive digital storytelling. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 48-54).

Peirce, N. & Wade, V. (2010). Personalised Learning for Casual Games: The 'Language Trap' Online Language Learning Game. In *4th European Conference on Game Based Learning (ECGBL), Copenhagen, Denmark* (pp. 21–22).

Pennebaker, J. W. & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296.

Pennebaker, J.W. & Stone, L.D. (2003). Words of wisdom: language use over the life span. *Journal of personality and social psychology*, *85*(2), 291.

Pereira, R. E. (2000). Optimizing human-computer interaction for the electronic commerce environment. *Journal of Electronic Commerce Research, 1*(1), 23-44.

Pette, C. (2001). *Psychologie des Romanlesens: Lesestrategien zur subjektiven Aneignung eines literarischen Textes*.

Picault, J., Ribiere, M., Bonnefoy, D. & Mercer, K. (2011). How to get the recommender out of the lab? In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 333-365).

Picazo-Vela, S., Chou, S. Y., Melcher, A. J. & Pearson, J. M. (2010). Why provide an online review? An extended theory of planned behavior and the role of Big-Five personality traits. *Computers in Human Behavior, 26*(4), 685-696.

Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*.

Pittenger, D. J. (1993). The utility of the Myers-Briggs type indicator. *Review of Educational Research, 63*(4), 467-488.

Plank, B. & Hovy, D. (2015). Personality traits on Twitter—or—how to get 1,500 personality tests in a week. In *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 92-98).

Plutchik, R. & Conte, H. R. (Eds.). (1997). *Circumplex models of personality and emotions*.

Prabhumoye, S., Tsvetkov, Y., Black, A. W. & Salakhutdinov, R. (2018). Style transfer through multilingual and feedback-based back-translation. *arXiv preprint arXiv:1809.06284*.

Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R. & Black, A. W. (2018). Style transfer through back-translation. *arXiv preprint arXiv:1804.09000.*

Propp, V. (1928). *Morphology of the folktale* (2nd ed.).

Quadrana, M., Cremonesi, P. & Jannach, D. (2018). Sequence-aware recommender systems. *ACM Computing Surveys (CSUR), 51*(4), 66.

Quantic Dream. (2010). *Heavy Rain*.

Quercia, D., Kosinski, M., Stillwell, D. & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 180-185).

Quijano-Sanchez, L., Recio-Garcia, J. A. & Diaz-Agudo, B. (2011). Happymovie: A Facebook application for recommending movies to groups. In *2011 IEEE 23rd international conference on tools with artificial intelligence* (pp. 239-244).

Random International. (2012). *Rain Room* [Art installation, Barbican Centre].

Rawlings, D. & Ciancarelli, V. (1997). Music preference and the Five-Factor Model of the NEO Personality Inventory. *Psychology of Music, 25*(2), 120-132.

Reid, M. & Zhong, V. (2021). LEWIS: Levenshtein editing for unsupervised text style transfer. arXiv preprint arXiv:2105.08206.

Rentfrow, P. J., Goldberg, L. R. & Zilca, R. (2011). Listening, watching, and reading: The structure and correlates of entertainment preferences. *Journal of Personality, 79*(2), 223-258.

Rentfrow, P. J. & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings ACM Conference on Computer Supported Cooperative Work* (pp. 175–186).

Ricci, F., Rokach, L. & Shapira, B. (2011). Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 1-35).

Riedl, M. O. & Stern, A. (2006). Believable agents and intelligent story adaptation for interactive storytelling. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 1-12).

Riedl, M. O. & Young, R. M. (2006). From linear story generation to branching story graphs. *IEEE Computer Graphics and Applications, 26*(3), 23-31.

Rishes, E., Lukin, S. M., Elson, D. K. & Walker, M. A. (2013). Generating different story tellings from semantic representations of narrative. In *International Conference on Interactive Digital Storytelling* (pp. 192-204).

Ritschel, H., Baur, T. & André, E. (2017). Adapting a Robot's linguistic style based on socially-aware reinforcement learning. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 378-384).

Rockstar Games. (2018). *Red Dead Redemption 2*. [Video game].

Rogers, K., Kamm, C. & Weber, M. (2016). Towards Player-Centric Adaptivity: Interactions of Gameplay Behaviour and Player Traits in a Survival Game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 269- 276).

Roshchina, A., Cardiff, J. & Rosso, P. (2011). User profile construction in the twin personality-based recommender system. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)* (pp. 73-79).

Rubens, N., Elahi, M., Sugiyama, M. & Kaplan, D. (2015). Active learning in recommender systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 809-846).

Ryan, N., Pascoe, J. & Morse, D. (1997). Enhanced reality fieldwork: The context-aware archaeological assistant. In V. Gaffney, M. van Leusen & S. Exxon (Eds.), *Computer Applications in Archaeology*.

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55*(1), 68.

Sacco, O., Liapis, A. & Yannakakis, G. N. (2016). A holistic approach for semantic-based game generation. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)* (pp. 1-8). IEEE.

Sadoski, M., Goetz, E. T. & Rodriguez, M. (2000). Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text types. *Journal of Educational Psychology, 92*(1), 85-95.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*(5), 513–523.

Santos, V. G. D. & Paraboni, I. (2022). Myers-Briggs personality classification from social media text using pre-trained language models. *arXiv preprint arXiv:2207.04476.*

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2000). Application of dimensionality reduction in recommender system-a case study (No. TR-00-043). Minnesota Univ Minneapolis Dept of Computer Science.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *WWW '01: Proceedings of the 10th International Conference on World Wide Web* (pp. 285–295). ACM.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2002). Incremental singular value decomposition algorithms for highly scalable recommender systems. In *Proceedings of the 5th International Conference in Computers and Information Technology* (pp. 27-28).

Schafer, J.B., Konstan, J. & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery, 5*, 115–153.

Schein, A. I., Popescul, A., Ungar, L. H. & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *SIGIR '02: Proc. of the 25th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval* (pp. 253–260).

Schell, J. (2005). Understanding entertainment: Story and gameplay are one. *Computers in Entertainment (CIE), 3*(1), 6-6.

Schiefele, U. (1996). *Motivation und Lernen mit Texten*.

Schraw, G. (1997). Situational interest in literary text. *Contemporary Educational Psychology, 22*, 436-456.

Schraw, G., Bruning, R. & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior, 27*, 1-18.

Schouten, B. A., Tieben, R., van de Ven, A. & Schouten, D. W. (2011). Human behavior analysis in ambient gaming and playful interaction. In *Computer Analysis of Human Behavior* (pp. 387-403).

Scott, M. J., Guntuku, S. C., Lin, W. & Ghinea, G. (2016). Do personality and culture influence perceived video quality and enjoyment?. *IEEE Transactions on Multimedia, 18*(9), 1796-1807.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J. & Napolitano, A. (2010). Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 40*(1), 185-197.

Shaker, N., Togelius, J. & Nelson, M. J. (2016). Procedural content generation in games.

Shang, M., Li, P., Fu, Z., Bing, L., Zhao, D., Shi, S. & Yan, R. (2019). Semi-supervised text style transfer: Cross projection in latent space. *arXiv preprint arXiv:1909.11493*.

Shani, G. & Gunawardana, A. (2011). Evaluating recommendation systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor (Eds.), *Recommender systems handbook* (pp. 257-297). Springer, Boston, MA.

Shardanand, U. & Maes, P. (1995). Social information filtering: algorithms for automating ‘word of mouth’. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI’95)* (pp. 210–217).

Sharma, M., Ontañón, S., Mehta, M. & Ram, A. (2010). Drama management and player modeling for interactive fiction games. *Computational Intelligence, 26*(2), 183-211.

Shen, J., Brdiczka, O., Ducheneaut, N., Yee, N. & Begole, B. (2012). Inferring personality of online gamers by fusing multiple-view predictions. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 261-273).

Shen, Y., Lin, Z., Huang, C. W. & Courville, A. (2017). Neural language modeling by jointly learning syntax and lexicon. *arXiv preprint arXiv:1711.02013*.

Sheth, B. & Maes, P. (1993). Evolving agents for personalized information filtering. In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications* (pp. 345–352).

Silva, M. P., do Nascimento Silva, V. & Chaimowicz, L. (2017). Dynamic difficulty adjustment on MOBA games. *Entertainment Computing, 18*, 103-123.

Silvia, P. J. (2006). *Exploring the psychology of interest*.

Silvia, P. J. (2008). Interest – the curious emotion. *Current Directions in Psychological Science, 17*(1), 57-60.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*, 466–469.

Sinha, R. R. & Swearingen, K. (2001). Comparing recommendations made by online systems and friends. In *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*.

Slade, D. (2018). Black Mirror: Bandersnatch. [Interactive film]. Netflix.

Smith, G., Whitehead, J. & Mateas, M. (2011). Tanagra: Reactive planning and constraint solving for mixed-initiative level design. *IEEE Transactions on Computational Intelligence and AI in Games, 3*(3), 201-215.

Snowdon, J. & Oikonomou, A. (2011). Creating more entertaining and re-playable games by dynamically introducing and manipulating, static gameplay elements and events. In *Computer Games (CGAMES), 2011 16th International Conference on* (pp. 94–100).

Soderbergh, S. (2017). Mosaic. [Interactive TV series]. HBO.

Sofia, G., Marianna, S., George, L. & Panos, K. (2016). Investigating the role of personality traits and influence strategies on the persuasive effect of personalized recommendations. In *4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE)* (p. 9).

Song, M., Namgoong, H., Kim, H. G. & Eune, J. (2009). A proposed movie recommendation method using emotional word selection. In *Online Communities and Social Computing: Third International Conference, OCSC 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings 3* (pp. 525-534).

Soto-Sanfiel, M. T., Aymerich-Franch, L. & Romero, E. (2014). Personality in interaction: How the Big Five relate to interactive narratives reception. *Communication & Society / Comunicación y Sociedad, 27*(3), 151-186.

Square Product. (1997). *Final Fantasy VII*. [Computer game.]

Staewen, R., Trevino, P. & Chang, Y. (2014). Player Characteristics and their Relationship to Goals and Rewards in Video Games. In *2014 IEEE Games Media Entertainment* (pp. 1-8).

Štajner, S. & Yenikent, S. (2021). Why is mbti personality detection from texts a difficult task?. In *Proceedings of the 16th conference of the European chapter of the association for computational linguistics: main volume* (pp. 3580-3589).

Steenberg, E. (2010). *Love*. [Computer game].

Stewart, B. (2011). Personality and play styles: A unified model. *Gamasutra, September*(1).

Straube, T., Preissler, S., Lipka, J., Hewig, J., Mentzel, H. J. & Miltner, W. H. (2010). Neural representation of anxiety and personality during exposure to anxiety-provoking and neutral scenes from scary movies. *Human Brain Mapping, 31*, 36–47.

Straube, T., Sauer, A. & Miltner, W. H. (2011). Brain activation during direct and indirect processing of positive and negative words. *Behavioural Brain Research, 222*(1), 66-72.

Su, X. & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence, 4*.

Sugihdharma, J. A. & Bachtiar, F. A. (2022). Myers-Briggs Type Indicator Personality Model Classification in English Text using Convolutional Neural Network Method. *Jurnal Ilmu Komputer dan Informasi, 15*(2), 93-103.

Sulem, E., Abend, O. & Rappoport, A. (2018). BLEU is not suitable for the evaluation of text simplification. *arXiv preprint arXiv:1810.05995*.

Sullivan, A., Mateas, M. & Wardrip-Fruin, N. (2010). Rules of engagement: moving beyond combat-based quests. In *Proceedings of the Intelligent Narrative Technologies III Workshop* (pp. 1-8).

Sundar, S. S. & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research, 36*(3), 298-322.

Swanson, R. & Gordon, A. S. (2012). Say anything: Using textual case-based reasoning to enable open-domain interactive storytelling. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(3), 1-35.

Swearingen, K. & Sinha, R. (2001). Beyond algorithms: An HCI perspective on recommender systems. In J. L. Herlocker (Ed.), *Recommender Systems, papers from the 2001 ACM SIGIR Workshop*.

Sykes, J. & Federoff, M. (2006). Player-centred game design. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1731–1734).

Szmydt, M. (2021). Contextual personality-aware recommender system versus big data recommender system. In *Business Information Systems* (pp. 163–173).

Tavakolifard, M. & Almeroth, K. C. (2012). Social computing: an intersection of recommender systems, trust/reputation systems, and social networks. *IEEE Network*, *26*(4), 53-58.

Taylor, L. N. (2002). *Video Games: Perspective, Point-of-View, and Immersion*. Gainesville, Florida, USA: Graduate Art School, University of Florida.

Tekofsky, S., Van Den Herik, J., Spronck, P. & Plaat, A. (2013). Psyops: Personality assessment through gaming behavior. In *Proceedings of the International Conference on the Foundations of Digital Games*.

Teng, C. I. (2009). Online game player personality and real-life need fulfillment. *International Journal of Cyber Society and Education, 2*(2), 39-50.

Teng, C. I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior, 26*(6), 1547–1554.

Thompson, W. F., Geeves, A. M. & Olsen, K. N. (2018). Who enjoys listening to violent music and why?. *Psychology of Popular Media Culture*.

Thue, D., Bulitko, V. & Spetch, M. (2008). PaSSAGE: A Demonstration of Player Modeling in Interactive Storytelling. *AIIDE, 8*, 227-228.

Thue, D., Bulitko, V., Spetch, M. & Wasylishen, E. (2007). Interactive Storytelling: A Player Modelling Approach. In *AIIDE* (pp. 43-48).

Tiihonen, J. & Felfernig, A. (2008). Towards recommending configurable offerings. In *Workshop on Configuration Systems (ECAI '08)* (pp. 29–34).

Tkalčič, M. & Chen, L. (2015). Personality and recommender systems. In F. Ricci, L. Rokach & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 715–739).

Tkalčič, M., Kunaver, M., Košir, A. & Tasič, J. (2011). Addressing the new user problem with a personality based user similarity measure. In *The UMAP 2011 Workshops-DEMRA 2011 and UMMS 2011: Decision Making and Recommendation Acceptance Issues in Recommender Systems; User Models for Motivational Systems: The affective and the rational routes to persuasion* (Vol. 740, pp. 106-111).

Tkalčič, M., Kunaver, M., Tasic, J. & Košir, A. (2009). Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges* (pp. 30-37).

Togelius, J., De Nardi, R. & Lucas, S. M. (2007). Towards automatic personalised content creation for racing games. In *2007 IEEE Symposium on Computational Intelligence and Games* (pp. 252-259).

Togelius, J., Preuss, M., Beume, N., Wessing, S., Hagelbäck, J. & Yannakakis, G. N. (2010). Multiobjective exploration of the starcraft map space. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games* (pp. 265-272).

Togelius, J., Yannakakis, G. N., Stanley, K. O. & Browne, C. (2010). Search-based procedural content generation. In *European Conference on the Applications of Evolutionary Computation* (pp. 141-150).

Tognetti, S., Garbarino, M., Bonarini, A. & Matteucci, M. (2010). Modeling enjoyment preference from physiological responses in a car racing game. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games* (pp. 321-328).

Tondello, G. F., Mora, A., Marczewski, A. & Nacke, L. E. (2019). Empirical validation of the gamification user types hexad scale in English and Spanish. *International Journal of Human-Computer Studies, 127*, 95-111.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A. & Nacke, L. E. (2016). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 229-243).

Tondello, G. F., Wehbe, R. R., Orji, R., Ribeiro, G. & Nacke, L. E. (2017). A framework and taxonomy of videogame playing preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 329-340).

Toy, M, Wichman, G., Arnold, K. & Lane, J. (1980). *Rogue*. [Computer game].

Tuedon, N. E. (2020). *Utilizing consumer personality profiles and keywords associated with personality types in information systems for movie recommendations.* (Doctoral dissertation, Manchester Business School).

Turkay, S. & Adinolf, S. (2010). Free to be me: a survey study on customization with World of Warcraft and City Of Heroes/Villains players. *Procedia-Social and Behavioral Sciences, 2*(2), 1840–1845.

Van Lankveld, G., Spronck, P., Van den Herik, J. & Arntz, A. (2011). Games as personality profiling tools. In *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)* (pp. 197-202).

VandenBerghe, J. (2012). The 5 domains of play: Applying psychology's big 5 motivation domains to games. In *Game Developers Conference, GDC Vault*. Retrieved from [URL: http://www.gdcvault.com/play/1015364/The-5-Domains-of-Play] (Archived at: http://www.webcitation.org/6Xva2AWDK)

Vazire, S. & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*(1), 123.

Verhoeven, B., Daelemans, W. & Plank, B. (2016). Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In *Proceedings of the Tenth international conference on language resources and evaluation (LREC'16)* (pp. 1632-1637).

Vicencio-Moreira, R., Mandryk, R. L. & Gutwin, C. (2014). Balancing multiplayer first-person shooter games using aiming assistance. In *2014 IEEE Games Media Entertainment* (pp. 1-8). IEEE.

Vrana, S. R., Cuthbert, B. N. & Lang, P. J. (1986). Fear imagery and text processing. *Psychophysiology, 23*, 247–253.

Vuoskoski, J. K., Thompson, W. F., McIlwain, D. & Eerola, T. (2012). Who enjoys listening to sad music and why?. *Music Perception: An Interdisciplinary Journal, 29*(3), 311-317.

Wade, S. E., Buxton, W. M. & Kelly, M. (1999). Using think-alouds to examine reader-text interest. *Reading Research Quarterly, 34*(2), 194-216.

Waldinger, R. J., Schulz, M. S., Barsky, A. J. & Ahern, D. K. (2006). Mapping the road from childhood trauma to adult somatization: the role of attachment. *Psychosomatic Medicine, 68*, 129–135.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A. & Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *NeuroImage, 58*(3), 963-973.

Walton, N. (2019). *AI Dungeon 2.* [Computer game]. https://www.aidungeon.io/

Wang, J., De Vries, A. P. & Reinders, M. J. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 501-508).

Wang, W. (2005). *Design of Trustworthy Online Recommendation Agents: Explanation Facilities and Decision Strategy Support.* Doctoral Thesis, University of British Columbia, Vancouver.

Wang, Y. (2015). Understanding personality through social media. Retrieved from https://nlp.stanford.edu/courses/cs224n/2015/reports/6.pdf

Wang, Y., Chan, S. C. F. & Ngai, G. (2012). Applicability of demographic recommender system to tourist attractions: a case study on trip advisor. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03* (pp. 97-101).

Wang, Y., Liu, Y. & Yu, X. (2012). Collaborative filtering with aspect-based opinion mining: A tensor factorization approach. In *Proceedings of the IEEE International Conference on Data Mining, Brussels, Belgium, IEEE Computer Society, ICDM'12* (pp. 1152–1157).

Wang, Y., Wu, Y., Mou, L., Li, Z. & Chao, W. (2019). Harnessing pre-trained neural networks with rules for formality style transfer. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3573-3578).

Wang, Y., Wu, Y., Mou, L., Li, Z. & Chao, W. (2020). Formality style transfer with shared latent space. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 2236-2249).

Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 47*, 1063-1070.

Weaver III, J. B. (1991). Exploring the links between personality and media preferences. *Personality and individual differences, 12*(12), 1293-1299.

Weber, R. & Shaw, P. (2009). Player types and quality perceptions: A social cognitive theory based model to predict video game playing. *International journal of gaming and computer-mediated simulations (IJGCMS), 1*(1), 66-89.

Wen, Z., Cao, J., Yang, Y., Wang, H., Yang, R. & Liu, S. (2023). DesPrompt: Personality-descriptive prompt tuning for few-shot personality recognition. *Information Processing & Management, 60*(5), 103422.

West, P. M., Ariely, D., Bellman, S., Bradlow, E., Huber, J., Johnson, E., Kahn, B., Little, J. & Schkade, D. (1999). Agents to the rescue?. *Marketing Letters, 10*(3), 285-300.

Widdeson, K. & Hadžidedić, S. (2022). FFM-SVD: A novel approach for personality-aware recommender systems. In *2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–8).

Winoto, P. & Tang, T. (2008). If you like the Devil Wears Prada the book, will you also enjoy the Devil Wears Prada the movie? A study of cross-domain recommendations. *New Generation Computing, 26*(3), 209–225.

Wong, C. O., Kim, J., Han, E. & Jung, K. (2009). Human-centered modeling for style-based adaptive games. *Journal of Zhejiang University-Science A, 10*(4), 530-534.

Wright, W. R. (2020). Inferring human personality from written media. (Doctoral dissertation, University of Hawai'i at Manoa).

Wu, W. (2017). Implicit acquisition of user personality for augmenting recommender systems. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion (IUI '17 Companion)* (pp. 201–204).

Wu, W., Chen, L. & He, L. (2013). Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM conference on hypertext and social media* (pp. 225-229).

Wu, W., Chen, L. & Zhao, Y. (2018). Personalizing recommendation diversity based on user personality. *User Modeling and User-Adapted Interaction, 28*(3), 237–276.

Xiao, B. & Benbasat, I. (2007). E-Commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly, 31*(1), 137-209.

Xu, X., Zuo, X., Wang, X. & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience, 29*, 8525–8529.

Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., Zhong, X. & Sun, J. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence, 48*(11), 4232–4246.

Yabar, Y., Johnston, L., Miles, L. & Peace, V. (2006). Implicit behavioral mimicry: investigating the impact of group membership. *Journal of Non-verbal Behavior, 30*, 97–113.

Yang, T., Deng, J., Quan, X. & Wang, Q. (2023). Orders are unwanted: dynamic deep graph convolutional network for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 11* (pp. 13896–13904).

Yang, T., Quan, X., Yang, Y. & Yu, J. (2021a). Multi-document transformer for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 16* (pp. 14221–14229).

Yang, X., Steck, H., Guo, Y. & Liu, Y. (2012). On top-k recommendation using social networks. In *Proceedings of the 6th ACM International Conference on Recommender Systems, Dublin, Ireland, ACM, RecSys'12* (pp. 67–74).

Yang, T., Yang, F., Ouyang, H. & Quan, X. (2021b). Psycholinguistic tripartite graph network for personality detection. *arXiv preprint arXiv:2106.04963*.

Yannakakis, G. N. (2012). Game AI revisited. In *Proceedings of the 9th conference on Computing Frontiers* (pp. 285-292).

Yannakakis, G. N. & Hallam, J. (2007). Towards optimizing entertainment in computer games. *Applied Artificial Intelligence, 21*(10), 933-971.

Yannakakis, G. N. & Togelius, J. (2015). Experience-driven procedural content generation. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 519-525).

Yannakakis, G. N. & Togelius, J. (2018a). *Artificial intelligence and games (Vol. 2).*

Yannakakis, G. N. & Togelius, J. (2018b). Modeling players. In *Artificial Intelligence and Games* (pp. 203-255).

Yannakakis, G. N., Spronck, P., Loiacono, D. & André, E. (2013). Player modeling. *Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik*.

Yee, N. (2006a). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and virtual environments, 15*(3), 309-329.

Yee, N. (2006b). Motivations for play in online games. *CyberPsychology & behavior, 9*(6), 772-775.

Yee, N., Ducheneaut, N., Nelson, L. & Likarish, P. (2011). Introverted elves & Conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 753-762).

Yi, M. Y., Lee, O. J. & Jung, J. J. (2016). MBTI-based collaborative recommendation system: a case study of webtoon contents. In *Context-Aware Systems and Applications: 4th International Conference, ICCASA 2015, Vung Tau, Vietnam* (pp. 101–110).

You, F., Wang, J., Zheng, Z., Jung, J., Che, C., Luo, X. & Tam, P. (2007). RAModel: A Taxonomy Model of Player Activities in Mobile MMORPGs. In: *Proc. of the 4th Intl. Conf. on Mobile Technology, Applications and Systems (Mobility 2007)* (pp. 704-707).

Young, R. M., Riedl, M. O., Branly, M., Jhala, A., Martin, R. J. & Saretto, C. J. (2004). An architecture for integrating plan-based behavior generation with interactive game environments. *Journal of Game Development, 1*(1), 51-70).

Youyou, W., Kosinski, M. & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*(4), 1036-1040.

Yu, H. & Riedl, M. O. (2013). Personalized Interactive Narratives via Sequential Recommendation of Plot Points. In *IEEE Transactions on Computational Intelligence and AI in Games, 6*(2), 174–187.

Zad, D. D., Angelides, M. C. & Agius, H. (2012). Personalise your massively multiplayer online game (MMOG) with Artemis. *Multimedia Systems, 1–26*.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt.2), 1–27.

Zammitto, V. L. (2010). *Gamers' personality and their gaming preferences* (Doctoral dissertation, Communication, Art & Technology: School of Interactive Arts and Technology).

Zhang, B. & Sundar, S. S. (2019). Proactive vs. reactive personalization: Can customization of privacy enhance user experience? *International Journal of Human-Computer Studies, 128*, 86-99.

Zhang, W., Ding, G., Chen, L., Li, C. & Zhang, C. (2013). Generating virtual ratings from Chinese reviews to augment online recommendations. *ACM Transactions on Intelligent Systems and Technology, 4*(1), 9.

Zhang, Y., Ding, N. & Soricut, R. (2018). Shaped: Shared-private encoder-decoder for text style adaptation. *arXiv preprint arXiv:1804.04093*.

Zhang, Y., Ge, T. & Sun, X. (2020). Parallel data augmentation for formality style transfer. *arXiv preprint arXiv:2005.07522*.

Zhao, S., Du, N., Nauerz, A., Zhang, X., Yuan, Q. & Fu, R. (2008). Improved recommendation based on collaborative tagging behaviors. In *Proceedings of the 13th International Conference on Intelligent User Interfaces, Gran Canaria, Canary Islands, Spain, ACM, IUI'08* (pp. 413–416).

Zhou, T., Kuscsik, Z., Liu, J., Medo, M., Wakeling, J. & Zhang, Y. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. In *Proceedings of the National Academy of Sciences of the United States of America, 107*, 4511-5.

Ziegler, C. N., McNee, S. M., Konstan, J. A. & Lausen, G. (2005). Improving recommendation lists through topic diversification. *WWW '05: Proceedings of the 14th international conference on World Wide Web, 22–32*.

Zuckerman, M. & Litle, P. (1986). Personality and curiosity about morbid and sexual events. *Personality and Individual Differences, 7*, 49–56.