

**Humanity Makes Machine, Machine Imitates Humanity: Cognitive Engagement with Artificial  
Intelligence Generated Literature and Origin Differentiation**

Camille R. Archer

Endicott College

### Abstract

The burgeoning fields of artificial intelligence and complex large language models call for investigations into the capabilities of these programs and how humans relate and respond to them. While previous research has investigated human response to literature generated by AI chatbots as well as ability to distinguish between AI and human outputs, this has yet to be done using any narrative outputs. This research utilized ChatGPT 3.5 to generate short narrative fiction texts in response to a prompt and presented 75 participants with a pair of stories, one AI generated and one human written. Cognitive engagement with each text and ability to detect which passage was AI-written were measured. Participants were unable to detect the AI-written text at a level greater than chance ( $\chi^2(1) = 1.946, p = .163$ ), and no variables were associated with accuracy. Engagement between human and AI texts was not significantly different ( $t(72) = -1.432, p = .157, d = -.135$ ), but engagement with the AI-generated texts was negatively associated with ChatGPT familiarity ( $\rho = -.302, p > .001$ ), and reading the human story first was associated with lower engagement with the AI text ( $t(72) = -2.369, p = .021, d = -.561$ ). These findings indicate that individuals are unable to distinguish between AI-generated and human written texts within the form of short narrative fiction, that accuracy in detecting AI-generated material may be very difficult to predict or understand, and that individuals may experience similar levels of engagement with each type of text.

## **Cognitive Engagement with Artificial Intelligence Generated Literature and Origin Differentiation**

The development and dissemination of sophisticated computer programs known as artificial intelligences (AI) in recent decades has far outpaced the true understanding of these developments (Goertzel, 2014). Knowledge about AI's capabilities, applications, and mathematical and theoretical background has flourished, but many of the questions proposed in the paper that coined the term 'AI' (McCarthy et al., 1955) have yet to be fully answered, and have even become more complex and difficult to address (Haenlein & Kaplan, 2019). Now the existence of programs which are intended to duplicate tasks or behavior which have previously belonged only to humans (Butlin et al, 2023) requires developing an understanding of how humans will respond to these novel programs and their outputs, and what their position in human life will be. As language and storytelling is such a central aspect of human cognition and experience (Winston, 2011), and is an area particularly tackled by AI behaviors, one key area for this will be understanding how people react to and interpret AI generated text and narrative content. The original research presented here attempts to investigate whether language, specifically short narrative fiction generated by an online AI 'chatbot' (specifically ChatGPT) may result in the same cognitive engagement and affective response as is experienced in response to many human-generated forms of language and literature. Additionally, the research will explore to what extent detection between human-generated and AI-generated short narrative literature is possible. These inquiries connect to a central characteristic of humanity, that of language, and attempt to provide important insights into the novel interactions between humans and chatbots.

To at least partially address these questions, this report will attempt to synthesize past and present research and scholarly thought in the fields of AI, cognitive science, and psycholinguistics. A large focus in the computer and cognitive science fields has long been evaluating the ‘intelligence’ element of artificial intelligence, therefore a discussion of the different testing methods and how they focus on language as a signifier, and how these tests inspire this report, is also included. As is essential to the question and research primarily at hand in this report, a review of the common uses of chatbots and human reactions to common AI behaviors will be held, along with a review of the research literature regarding the interactions between humans and different forms of language and literature.

### **AI and Ethics**

The model of AI that is the most important in this discussion is ChatGPT, the chatbot released by OpenAI. A chatbot is a Large Language Model (LLM) which is designed to replicate language so it resembles human usage as closely as possible. These models are trained on huge corpuses of text data including websites, books, poetry, informative books and many other forms of text, picking up on the rules of human speech as well as huge amounts of informational data (Naveed et al., 2023). ChatGPT can accept text input prompts in order to write and debug code, answer a large number of questions, summarize or explain large amounts of text, generate creative ideas for stories or names, assist in writing different types of text, explain complex concepts, translate between languages, and much more (OpenAI, n.d). ChatGPT is one of the most powerful LLMs and one of the most powerful AI systems today, with GPT-3.5 and GPT-4 able to pass tests like AP subject tests, the bar exam (GPT-4), the LSAT (GPT-4), and parts of the GRE (Open AI, 2023). Based on the exponential growth of these GPT systems, it is fair to

assume that their abilities will only continue to become broader and more complex, necessitating consistent new evaluation and investigation, as this report attempts to add to.

Investigation into the capabilities and uses of AI chatbots is essential in a world where they are used so pervasively, developing so swiftly, and where large numbers of people are already concerned over their proper use and regulation. Varied groups of people have ethical and practical concerns over the development and use of AI chatbots in fields such as media and film (Bedingfield, 2023; Coyle, 2023; Child, 2023; del Barco, 2023; Maddaus, 2023), literature (The Author's Guild, 2023), the labor market (Rainie et al., 2022; Eloundou et al., 2023), and their overall use (Future of Life, 2023). Researchers have suggested that more studies into HumanComputer Interaction research with powerful Conversational Agents (aka chatbots such as ChatGPT) need to be conducted to determine how these novel interactions will proceed and possible risks that may arise (Weidinger et al., 2021), exactly what this current paper attempts to provide more evidence for. While the present paper does not create a solid framework used for evaluating risk, it does attempt to evaluate risk in some areas of human-computer interactions. If people are unlikely to be able to differentiate between text origins, and if narrative literature generated by chatbots are engaging and enticing to readers, it is more likely that texts generated by an AI, which may be more inaccurate, offensive, or misleading may be treated as human written, and perhaps viewed as more trustworthy and reliable. An additional risk that may be made clear through this report is the probability that individuals or companies may be able to use ChatGPT to generate media content such as novels, short stories, poetry, scripts, or journalism, thereby undercutting the jobs and finances of writers who use writing as their full-time job. Clearly, the concerns that common people and experts hold regarding AIs and chatbots in particular are not entirely unfounded and require investigation into the ways in which humans

respond to AIs in order to determine the level and type of threat posed by the literary and linguistic capabilities of these programs.

### **Tests for AI Intelligence and Creativity**

Now that machines and programs which are capable of language production nearing a humans exist, understanding, evaluating, and theorizing on methods through which to distinguish between humans and these machines is more essential than ever before. Scholars do not agree on whether any AI at present or in the future does, or ever can, possess intelligence, knowledge, mind, consciousness, or anything of the case (Butlin et al 2023; McClelland, n.d). However, there are still a number of tests that have been created with the intention of testing artificial programs for the existence of these concepts as well as testing human-computer interactions. As will be seen, many of these tests see language as a lynchpin in distinguishing between humans and machines, as language can be seen as distinctively human and very difficult to directly copy. This makes them ideal for beginning investigations into human responses to the language and literature used and generated by chatbots.

Perhaps the first of these tests, and certainly the most famous, is Alan Turing's Imitation Game, which draws largely on the concept of language as a main method of relaying thought (Turing, 1950). In the test that Turing proposed, which is now more widely known simply as 'the Turing test', a judge is asked to distinguish between two parties, which Turing labels A and B, where A is a digital computer and B is a human agent. The goal of B is to help the judge determine which of them is the true human, while A wants to remain undetectable as a machine and pass for human. Meanwhile, the judge is aware that one of the participants in the game is a machine. This test has long been hailed as the hallmark of evaluating intelligent machines, and has had an impact on computer science, cognitive science, and the scientific community as a

whole as only a few concepts have ever had (French, 2000). It has generated attempts at creating programs which can communicate like humans, arguments over the meaning of intelligence, who can possess it, and how humans can understand themselves, and has generated a competition, the Loebner Prize, designed to evaluate computer programs through Turing's paradigm (Mauldin, 1994). Despite the impact of Turing's propositions, critiques of this test abound, for many reasons (LaCurts, 2011; French, 2000; Landgrebe & Smith, 2019; French, 1990; Floridi et al, 2008; Russell & Norvig, 2009; Riedl, 2011; Hayes & Ford, 1995; Gunderson, 1964). However, variations on the Turing Test remain in use in some academic and theoretical settings, and very clearly places an emphasis on the use of language in understanding the functioning of AI programs as well as the ways in which humans and AIs will interact with each other.

The Turing test is not the only test that has been generated in order to test the intelligence and capabilities of an artificial program. The Lovelace Test 2.0 (Riedl, 2011) was generated with the intention of making changes to the original test to allow an agent to pass, while still making it rigorous enough to demand some level of creative 'thought' or action. In this test, an artificial agent is charged with creating an artifact of a certain type (e.g. drawing, poem, novel, story, email). A human evaluator may set a limit of constraints which the artifact must fit within, and must be satisfied that the artifact, when created, is a valid example of the type of artifact requested and that the constraints have been satisfied. If all these conditions are met, one could conclude that the computer may possess some amount of understanding and creativity, as it must manipulate its output according to complicated instructions. This test, as did Turing's, places the bulk of the emphasis on language. It is not necessary that the artifact be of a literary quality, it could indeed be anything that a human could be asked to create, but it is explicitly mentioned

that evaluating an AI's ability to generate linguistic properties would be an excellent use of the test.

These tests demonstrate that it is highly desirable to investigate the behaviors and characteristics of the AI programs in order to understand how they and humans may interact and how they may be used effectively and properly. They are also important for the current question as they illustrate the importance of the conjunction between language, AI capabilities, and human reactions to them, and provide opportunities to measure those elements. While the present study does not attempt to judge the intelligence of the AI machine, as both of these tests do, it uses the paradigm of the Turing Test (asking a judge to label output as either AI or human generated) combined with the specific type of output as the basis of judging (a creative artifact, as in the Lovelace 2.0 test). It is not important for this question in this context whether a program must be intelligent to produce these outputs, but rather how people respond to the output of these programs, which can still be looked at through manipulation of these tests.

### **Human and AI Interaction**

The uses of AI in human life would be a difficult topic to cover comprehensively, but the functions of chatbots is of particular concern in this report, as they are the most commonly accessible and familiar to the layperson, and operate almost solely in text outputs. Some of their uses include answering content questions for classes, explaining concepts, generating, explaining, and assisting to debug code, generating emails, social media posts, or other short passages, summarizing long documents, synthesizing data, generating review questions for test preparation, and much more (Ray, 2023; Kasneci et al., 2023). Companies also use AIs as customer service chatbots, taking in customer or patient information, making appointments, analyzing market trends and company data, and answering inquiries (Ray, 2023). Large areas of



chatbot use are within the fields of education (e.g Kasneci et al., 2023; Eysenbach, 2023; Berse et al., 2023; Zawacki-Richter et al., 2019) and language learning (e.g Kohnke et al., 2023; Haristiani, 2019; Pikhart, 2020). All of these applications require the production of language to be understandable and interesting to readers, and for there to exist a solid understanding of the capabilities of these programs, their capacity to produce this type of language, and how humans react to them in these and other capacities. They must be trusted by consumers and producers to produce content that will be accurate, meaningful, and aligned with the goals of the user. Further research on the ways in which chatbots are used and interpreted by consumers is necessary to demonstrate the suitability of chatbots for the purposes in which they are currently being used as well as determine their potential for other uses.

Objectively, AI-generated texts can be identified relatively easily using analysis of dimensions of texts (Sardinha, 2024), including abstraction, narrative style, informational content, and use and frequency of word types. However, this is not always the case when humans attempt to distinguish between texts of different origins. A study (Hassoulas et al., 2023) used student written and ChatGPT generated essays to test distinction accuracy by markers, professional test graders experienced at reading and evaluating essay writings. They discovered that even experienced markers were unable to consistently distinguish between texts written by students and those written by NLP tools like ChatGPT. When asked about how they made their determinations, references to the content of the texts as a clue was associated with more accurate labeling, but references to writing style as a clue did not correlate with accuracy in origin detection. This indicates that ‘writing style’ may not be a suitable method of distinguishing between chatbot generated and human generated texts, as these programs may be able to mimic human writing styles quite well. However, chatbots seemed to be unable to master “reflective

nature” of writing in essays that called for personal details, suggesting that they have yet to be able to mimic the thoughts and experiences of a human in writing. It appears that there are many limits to the capabilities of ChatGPT and other chatbots, when can be demonstrated through assessing human reactions to AI generated texts.

Testing abilities of people to distinguish between human and AI generated poetry (Kobis & Mossink, 2021) has also revealed that participants are not reliably able to determine between human and AI generated content of this other literary form. Specifically, when the response from the chatbot to be presented to the human participant was chosen at random, there was a significant preference for human-generated poetry over chatbot generated, but when the researchers picked the best response that the chatbot generated, there was no significant preference for human writings over chatbot writings. Chatbots may produce different texts in response to the same prompt, and these may apparently be of different quality. When the best quality are chosen, they are similar enough to human writings for the preference for human writings to be less significant than when the output is chosen at random. Additionally, participants were able to accurately detect the AI generated poetry when chatbot outcome was chosen at random, but not when humans selected the best one.

Similar findings were detected when presenting AI and human generated text to linguists in the form of research paper abstracts (Casal & Kessler, 2023). Even professionals trained in analyzing and reviewing language and literature could not significantly succeed at accurately detecting AI content. Additionally, the researchers saw that participants were more accurate at identifying human generated texts as human rather than identifying an AI generated text as AI created; that is, some AI texts were determined to be human. This latter finding was replicated in a study (Gao et al., 2023) which presented scientific researchers with abstracts either generated

by AI or humans. While judges were largely able to correctly identify articles as either AI or human generated, they incorrectly identified 32% of generated abstracts as real, and only 14% of human abstracts as being AI generated. This event may be concerning, as it suggests a possibility of writing done by chatbots, which can be inaccurate, outdated, or biased, as being mistaken for human writing, and therefore perhaps more trustworthy.

Only one study (Hulman et al., 2023) found that participants were able to distinguish between AI generated (through ChatGPT) and human generated texts, specifically presenting participants with answers to questions about diabetes and asking them to identify AI generated answers. Participants were able to identify AI generated answers 58.5% of the time, a level higher than chance and against the researchers' hypotheses. This research also demonstrated an interesting finding regarding previous use of ChatGPT, which they found as the most important predictor for accurate identification of text origin. These findings suggests that there may be an element of linguistic features unique to ChatGPT that make those familiar with its patterns more adept at picking out the artificial text. This finding suggests that the ability to identify origins of texts may become easier with time, as more people are exposed to the patterns of chatbot writings. This may lessen the possible ethical violations of chatbot use, but it may also decrease the practical uses that ChatGPT may be used for, if people become able to identify chatbot generated writings in everyday life. In the field of art, Demmer and colleagues (2023) utilized a Random Noise Generator to randomly assign black or white to a 32 x 32 grid of squares and tasked human artists with doing the same. By presenting these different outcomes as stimuli to participants, the researchers measured origin detection, effects of priming, and emotions felt in response to different artworks. This investigation revealed that there was not a significant effect

of art origin on emotions reported by participants, and that guesses on origin were significantly predicted by the actual origin of the artwork.

The literature on the whole suggests that people, even those educated in the field or in analyzing language patterns, are largely unable to consistently distinguish between AI chatbot generated texts and those created by humans. However, there has of yet not been investigations into the possibility of differentiation in any sort of narrative text. The present report thus attempts not only to demonstrate ways in which these programs could be used effectively, but also to add to the scholarly literature by filling in this gap surrounding chatbot writing capabilities and the ways humans respond to and interpret them.

## **Literature and Engagement**

### ***What is Engagement?***

Narrative engagement refers to the feeling of involvement with a form of media such that a new world is created within the mind, and where it may sometimes be difficult to return to the real world, or see aspects of the real world differently (Busselle & Bilandzic, 2008). There are several different theories regarding how engagement in narratives is constructed, including the mental models perspective, situation models, the story world model, and deictic shift theory (Busselle & Bilandzic, 2008). These theories draw largely on the same concepts, but they predict different ways that engagement may be created and what factors may limit or predict it. For the purposes of this study, the structure and functioning of engagement largely follows the theory structured in Busselle & Bilandzic (2009). This theory presents the necessary pillars to create engagement, features of active engagement, and the personal outcomes of becoming engaged in a narrative (full model displayed In Figure 1). It also aligns with the Narrative Engagement Scale, used in this report to measure engagement with narrative texts.

There are multiple separate theorized constructs that make up the concept of engagement within this definition: transportation, identification, presence, and perceived realism. The concept of transportation refers to a state similar to that of the ‘flow state’ (Csikszentmihalyi, 1997), where the consumer unconsciously creates a mental model of the narrative, which results in some loss of awareness of self and real world and a move into the ‘story world’ represented in the head. The process of identification involves the consumer taking on the perspective of the characters in the story world, so that they understand their actions, motives, or personalities, facilitating a level of stepping into the characters’ minds. The concept of presence suggests a sense of physical transportation: they may feel that space or time have somehow changed and are now following that of the story world (thus a movie may feel over in only a few minutes; Busselle & Bilandzic, 2009).

In order to facilitate engagement in a narrative, the reader should feel that the story world follows consistent rules, and that information is not presented in the construction of the narrative that does not follow the understanding of the story world as previously explained. However, consistency of rules between the real world and the story world does not always serve as a barrier to engagement, as people are quite adept at understanding and accepting the fictional nature of a story (Dijkstra & Fleming, 2023): if the deviation from the real world is explained within the structure of the story world, perceived realism and therefore the immersive experience of the world can be maintained. When engaging with a more fictional narrative, in any form, the user must “exercise a creative faculty” (Worth, 2004) in order to involve themselves in the narrative, in order to feel connected to and part of the fictional world that is being displayed. This use of creativity of mind allows for some small deviations between real and story worlds, as the consumer understands that a fictional world may change rules from that of the real world and

still maintain their sense of transportation, identification, and presence within the story world. Nevertheless, it is still possible for breaks between rules of the real world and those of the story world to negatively interfere with engagement (Busselle & Vierrether, 2022), particularly when the story otherwise follows the rules and knowledge of the real world. The construct of narrative engagement is an important one in this study, as it may demonstrate the extent to which chatbots have the ability to craft language which is engaging to readers, which may then stretch to applications in non-narrative literature such as professional documents or advertisements. Whether or not these applications are deemed to be agreeable on the whole, understanding the possibility of these uses of chatbots is essential to understanding human-computer interactions and generating proper useability and restriction guidelines to ensure that the development and use of these programs is in the best interest of society.

### ***Evidence of Engagement.***

Narrative engagement has been demonstrated in several different forms of literary media: poetry (Johnson-Laird & Oatley, 2022), film (Oliver, 1993; Oliver, Weaver, & Sargent, 2000), short stories (Cupchik et al., 1998; Busselle & Vierrether, 2022), novels (Hogan, 2018) and marketing stories (van Laer et al., 2019; Kang et al., 2018) are all capable of generating some level of internal engagement from a reader, although this may vary in composition between form of media. Levels of transportation have been associated with favorable responses to stories (Kang et al., 2018), while inconsistencies in external realism result in less emotional engagement and sense of presence in the story (Busselle & Vierrether, 2022). Identification with characters has been seen to be driven by trait empathy and tendency to take on the mind of a character ('Theory of Mind'), demonstrating that internal characteristics have an effect on levels of engagement, and not all people will respond to narratives in the same way (Eekhof et al., 2023).

It is theorized that there may be as many as four categories of emotions that can be elicited while engaging with a form of media: emotions of empathy, emotions of sympathy, remembered emotions, emotions of identification, and relieved emotions (Mar et al., 2011; Oatley, 2002; Oatley, 1994). Many of these are elicited through engagement with characters in narratives, through an understanding of their dilemmas and emotions, and sometimes a feeling of similarity with certain characters. However, it is unessential for a character to always be sympathetic or identified with, or even present at all, for many forms of literature or media to produce emotions (Mar et al., 2011). Emotions and engagement in different forms of media can even be studied and proven from within the brain (Altmann et al., 2012; Jacobs, 2015; Satpute & Lundquist, 2020; Schmälzle et al., 2022), making it clear that these are not imagined phenomena, but actually deeply ingrained in humans. It is no surprise then, that literature is an area of human functioning that is rife for emotional experiences and associations.

Even personality traits may be changed simply through exposure to a single narrative short story (Djikic et al., 2009). If this is possible in an experimental condition, with exposure to a single narrative stimulus, it is certain that exposure to literature, and especially narrative literature, over the course of the lifetime may result in larger changes in personality. Indeed, it has been seen that individuals who engage in more reading of fiction, compared to those who engage in more non-fiction reading, demonstrate better performance on measures of social ability, as well as empathy measures, and also the tendency to become engaged in narratives (Mar et al., 2006). It is therefore possible that lifetime exposure to fictional narratives fosters alteration in very important elements of personality, an effect which emphasizes the importance not only of literature in human life but also that of examining the possible effects that AI literature will have on the human psyche and the human functioning in the long term.

This study questions if chatbots are capable of crafting a narrative which creates or integrates all of the elements of narrative engagement and is therefore able to result in a level of engagement within the story from the reader. If this is the case, these technologies could be used to engage students in educational capacities, assist writers or screenwriters in developing and editing their work if desired, engage in meaningful conversations with users (assuming the ability to write engaging fiction may transfer into the ability to communicate in an engaging way), write stories that may complement those written by humans, or take over some areas of jobs which may be more tedious or time consuming. More simply, however, it would also indicate that the ability to understand what constitutes an engaging narrative, and then to write it, is something that can be learned through analyzing the patterns of large amounts of text data, rather than experiencing the feelings and events oneself. Any of these conclusions may have important implications for the use of AIs and chatbots specifically, as well as how humans will interact with them.

### **The Current Study**

The present study attempts to contribute to the ongoing research into human and AI interaction, adding to the understanding of human attitudes towards AI programs and the ways in which AIs can be effectively and ethically utilized. This research investigates an unstudied form of AI content in the form of flash fiction, which should shed light on the capabilities of AI to create original narratives that produce engagement within consumers. Flash fiction is a category of short stories, with a maximum of 1,000 words, in which some artistic licenses are taken but an essentially complete story is still presented. This format of narrative was chosen over others in order to include the characteristic elements of character, dialogue, and setting, and carry out a relatively full plot, but also to be short enough to ensure continuous engagement in each story



from all participants. Short stories longer than 1,000 words may likely result in a lack of attention to the story and its elements, while selecting passages from existing, longer works may not result in a full use of plot and other elements which are involved in producing narrative engagement (Busselle & Bilandzic, 2009; Busselle & Bilandzic, 2008).

The first question under investigation in the present study is whether AI generated texts are able to produce engagement and emotions to the same extent that very similar human generated text can do. The second question attempts to determine if people are able to accurately detect between two passages and identify one passage as being written by a computer program (i.e. a version of the Turing Test). A tentative hypothesis is made that there will not be higher engagement scores in response to human generated texts than to AI generated texts, based on previous research that suggests that people do not significantly prefer human generated outputs over AI generated outputs, both when aware and when unaware of the existence of an AI generated product in the experiment (Kobis & Mossink, 2021; Demmer et al., 2023; Casal & Kessler, 2023; Hulman et al., 2023). While most studies suggest that people are largely unable to distinguish between AI generated and human generated content, this was not hypothesized here due to the unique nature of the presented stimuli, for which no true precedent in this paradigm exists. Most studies which attempt to have participants distinguish between these two types of texts have not tested creative literature (rather have utilized informational or academic type texts), and those that have, have not tested longer narrative creative texts. This makes it difficult to make a true prediction as to the relationship between stimuli and response; therefore, this study remains exploratory in nature as to whether participants will be largely successful in detecting the true origin of texts. A secondary hypothesis predicted that participants with stronger

previous voluntary interaction with various types of media will be more successful in accurately detecting passage origin. It was additionally hypothesized that knowledge and use of

ChatGPT would be positively associated with ability to detect passage origin.

### **Method Participants**

The participants for this study were drawn largely from an undergraduate college community recruited through advertisement around the college campus, word of mouth, and through advertisements to professors to spread the study. Participants were also gathered through use of SurveySwap.com. Efforts were made to include participants from the population of faculty and staff within the main college campus. There were no specific inclusion or exclusion criteria for participation in the study.

### **Stimuli**

To ensure a level of quality for the stimuli in the experiment, all flash fiction passages were taken from *Flash Fiction America: 73 Very Short Stories* (2023) a compendium of flash fiction from the past 20 years. The stories chosen had to be (a) between 500 and 1,000 words, (b) include more than 5 lines of dialogue, (c) include more than one named character, (d) have some mention of setting (i.e. not abstract). These elements were essential to creating a narrative which can then be extended to the types of narratives found in short stories and novels. The chosen flash fiction passages can be found in full in *Appendix A*.

The chosen flash fiction passages were then used to generate prompts to create the AI passages. ChatGPT 3.5 (OpenAI, 2022) was fed the prompts derived from these flash fiction stories in order to create passages which are similar in length, readability, characters, setting, and plot premise. The prompt used to generate these passages were designed to contain these elements: first- or third-person narration, details of the characters (namely occupation,

relationships, history, or aspirations, as important to the plot of the story), the main event of the story, and how long the story should be. The purpose in including these elements was to ensure that ChatGPT created a passage that include the most important parts of a narrative story which would be essential in evoking narrative engagement. The similarities in passages were also designed to limit the number of confounding factors which may lead to choose one passage over another, and instead attempts to isolate the patterns and specific elements of ChatGPT's writing. The full prompts aligned with each flash fiction story can be found in *Appendix B*.

## **Measures**

### *Narrative Engagement Scale*

The Narrative Engagement Scale (*Appendix C*; Busselle & Bilandzic, 2009) is designed to assess engagement experienced during reading or watching a text which presents a story, either fictional or nonfictional. The scale follows the concept of a mental model explained in Busselle & Bilandzic (2008), fully fleshed in the depiction of Figure 1. The scale features four subscales intrinsic to the concept of narrative engagement: narrative understanding, attentional focus, emotional engagement, and narrative presence. Narrative understanding includes items (ex. My understanding of the characters is unclear, which is reverse coded) that measure the ease with which a narrative can be understood, and how easily a mental model of the narrative can be created. Attentional focus includes items (ex. I found my mind wandering while reading, reverse coded) which measure the level to which the participant is fully involved with the reading to the extent that they engage in smooth narrative processing and are unaware of any drifting attention. Emotional engagement (ex. I felt sorry for some of the characters in the story) measures the extent to which the participant experiences overall emotional arousal in response to literature but does not require for any specific emotions to be identified. The last subscale, narrative presence,

essentially measures transportation into the story, the feeling of having left the actual world and entered the story, and subsequently feel a loss of attention to the real world and the sensation of entering another space and time (ex. At times in the story, the story world was closer to me than the real world). This scale has demonstrated correlation with enjoyment, identification, and transportation, main elements of engagement with narrative literature. Psychometric properties for this scale (construct and convergent validity) have also been demonstrated (Mahoney & Tang, 2020).

### *Origin Detection*

One question was used to measure participants' ability to determine which of the passages they were presented with seemed to be AI generated ("If you had to choose, which of the passages that you read do you think is more likely to have been written by an artificial or chatbot program (i.e. ChatGPT, Google's Bard, TextFX)?). Participants were only able to identify one of the passages as being likely to be AI written. A confidence check ("How confident are you in your response to the previous question", 5 item Likert, 1- I guessed to 5 Almost positive) was included to allow for dimensions in what would otherwise be a single item dichotomous response measure. Additionally, an open-response question was included to ask participants why they identified the passage that they did as the one likely written by AI.

### *Demographics*

Demographic questions measured level of education, college major, age, confidence in reading fiction (5 item Likert, 1- *Not Confident* to 5- *Extremely Confident*), and previous involvement with literature. This last item presents a list of 6 types of literature (academic, long fiction, short fiction, nonfiction, poetry, and manga/comics/magazines) and asks participants to respond how often they voluntarily engaged with the different types of media (5 item Likert,

1*Never or almost never* - 5*Every day*). Additionally, an item measured familiarity with ChatGPT (5 item Likert, 1-*Never heard of it* - 5-*Have heard of it and use it often*). This question is important for analyses and hypotheses of the current paper, as it has been demonstrated to be associated with accuracy in origin detection (Hulman et al., 2023).

## **Procedure**

The experiment was located on Qualtrics, where participants were informed that the purpose of the study is to understand how people relate and respond to literature, with no references to AI generated content. Participants were informed that they could withdraw at any point, that their data is anonymous, and that there were no risks inherent in the study. After completing the informed consent form, participants were presented with one passage. The origin of passages presented were counterbalanced to protect against order effects. However, the passages were paired so that the GPT generated passage was given with the flash fiction story that created its prompt. This allows a level of control so that engagement can be attributed to origin rather than to an individual interest in one story's plot or characters over the others.' After presentation with the first passage, participants advanced to complete the Narrative Engagement Scale (Busselle & Bilandzic, 2009). The second passage is then presented, followed by another presentation of the Engagement Scale. After reading the second passage and completing the Engagement Scale for that passage, participants were presented with the Demographics questionnaire (Appendix D). This presentation of questions included the origin detection question, confidence in their answer, and a qualitative open response item to report their reasoning in their choice.

## Statistical Analysis

To compare the level of engagement between AI stories and human stories, a paired t-test was conducted to determine if mean engagement scores were significantly different. The Narrative Engagement Scales were coded between -2 and 2, and the overall score for each passage was calculated as an average of the responses to each item for that passage. A ChiSquare Goodness of Fit test was run to determine if the accuracy for distinguishing between AI and human texts was significantly different from chance (50%). Regarding the secondary analyses, associations between familiarity with ChatGPT, interaction with textual media, and accuracy in origin detection were conducted with Pearson's point-biserial correlation coefficient ( $r$ ) due to the dichotomous nature of the origin detection variable. Exploratory analyses were conducted between variables utilizing logistic regressions and the appropriate statistical correlation as demonstrated in Khamis (2008). For qualitative analysis, attempts were made mainly to describe the type of allusions made to the different passages and the indicators within the texts that led individuals to select one passage as AI written.

## Results

75 participants fully completed the survey and had their data analyzed for this study. Participants were aged between 18 and 62, 71.6% of participants had English as their first language, and 28.4% spoke a language other than English as their first, largely European. The majority of participants had completed a bachelor's degree or higher ( $n = 45$ , 60.81%). Specific demographics and their breakdown by story theme are depicted in Table 1.

No significant difference in engagement with AI passages and human passages was demonstrated by a two-tailed t-test ( $t(72) = -1.432$ ,  $p = .157$ ,  $d = -.135$ ). Nominally, the engagement with human texts was lower ( $m = 3.31$  vs  $m = 3.52$ ), but this was not significant

statistically. A one-way ANOVA for engagement between AI and human texts by story theme did not find any significant effects (Table 2). Exploratory analyses related to engagement were also conducted (Table 3). A medium, significant negative interaction was found between engagement with AI generated texts and familiarity with ChatGPT ( $\rho = -.302, p > .001$ ). AI engagement score was also significantly related to order presentation, although with a small effect ( $\rho = .295, p = .010$ ). An independent samples t-test demonstrated that those presented with the human passage followed by the AI passage experienced significantly lower AI engagement effects ( $t(72) = -2.369, p = .021, d = -.561$ ). Otherwise, engagement scores were not related to any other variables collected.

A Chi-Square Goodness of Fit test determined that the overall accuracy of participants' AI origin detection was not significantly different from chance ( $\chi^2(1) = 1.946, p = .163$ ). Story theme did not significantly interact with accuracy ( $\rho = -.065, p = .583$ ). Nominally, more people were able to accurately detect the AI passage ( $n = 43$ ) than who incorrectly identified the human passage as AI generated ( $n = 32$ ). There were no significant correlations between accuracy with any of the variables collected (Table 3). A logistic regression for accuracy with all other variables included as predictors returned no significant predictors of origin detection accuracy ( $R^2 = .057$ ; Table 4). Of the NES subscales, Narrative Understanding completed on the AI text was associated with the accuracy of origin detection ( $r = -.229, p = .048$ ), with higher scores on this subscale associated with incorrect origin determinations. Interestingly, levels of accuracy within education levels were scattered (Table 5). Within those who completed only high school, 75% were right ( $n = 3$ ), within those who did not graduate college 52% were right ( $n = 13$ ), in those who possessed a bachelor's 63% were right ( $n = 17$ ), in those with a master's 41.7% were right ( $n = 5$ ) and in those with a doctorate, 83.3% were right ( $n = 5$ ).

Other exploratory analyses between variables were conducted (Table 3). Familiarity with ChatGPT was significantly associated with having a language other than English as a first language ( $r = -.391, p < .001$ ), and engagement with textual media was significantly related to confidence in ability to read English literature texts ( $\rho = .330, p = .004$ ). Education level was significantly and strongly related to age ( $\rho = .678, p < .001$ ), and to familiarity with ChatGPT ( $\rho = .231, p = .046$ ). Otherwise, there were no other significant relationships between the variables collected.

Qualitative answers were categorized as mentioning one or more of six themes identified within the overall responses. When in doubt regarding placement in a specific category, categorization would not be made to that category. The six categories were Clarity/Complexity, Scripted/Simplicity, Grammar/Organization, Character Depth, Emotionality, and Plot/Content (Table 6). Six answers were omitted due to not enough clear reasoning given. Table 7 depicts the number of answers counted in each theme. Reference themes did not vary significantly within accuracy for any of these themes except for Scripted/Simplicity, within which significantly more references were made by individuals who made correct origin detections ( $\phi = .276, p = .017$ ). References to themes did not vary by story theme (Table 8). References to any of the qualitative themes was not associated with engagement with either of the texts (Table 9).

Full study data is available in the study repository at <https://github.com/camillearc/Thesis-Materials>.

## **Discussion**

### **Text Engagement**

With the expansion and development of artificial language models implemented in numerous areas, it is essential to understand the possibility that these language models produce



works that humans interpret as human generated. In the current field investigating these questions, none have utilized narrative literature in any form as a stimulus, missing a large corpus of data that these models are trained on and may generate. This study is therefore novel in its use of narrative literature as a stimulus and provides important evidence regarding the intersections between engagement, detection of AI texts, and other variables affecting these relationships.

In line with the tentative first hypothesis, there were no significant differences of engagement between texts, although engagement with AI stories was nominally larger than engagement with the human stories. It may very well be the case that the form of literary fiction seems to many people to be hard to understand, has less coherent structures, includes unnecessary details, or that it is “trying too hard” to be creative, and that these elements of this specific form of literature interfere with origin detection and with engagement. This may not preclude the existence of a relationship between engagement and AI/human authorship but rather reflects the specific characteristics of the specific type of narrative texts chosen for this project. Studies which utilize narrative texts which are seen as less “literary” may reflect different levels of engagement between the stories. It may also be possible that AI stories replicate human writing and the feelings garnered in response to them such that engagement with the different text origins is not different.

The Narrative Understanding subscale of the NES completed in response to the AI texts was the only subscale of the NES which was associated with origin detection, such that those with a higher score on the subscale were linked with inaccurate origin detections. It seems possible that individuals who understood the AI generated story took this easy understanding to indicate that the text was written by a human when in fact, the simpler story was the AI

generated one. In some cases, however, individuals may have experienced similar or slightly lower scores on this subscale but utilized other concepts to make their origin determinations. It may be the case that understanding of a text is the most important element in determining whether a text may be artificially generated. However, engagement with the AI text overall was not associated with accuracy, and it seems to be the case that experience with the other subscales detract from the ability to make correct judgements.

Engagement with AI texts was found to be negatively related to familiarity with ChatGPT, such that individuals who use ChatGPT more frequently experience less cognitive engagement with the text generated by AI. This finding suggests the possibility that knowledge of the patterns and language used in these artificial programs decreases the cognitive engagement felt with the text. It may be that they associate the AI text passage with the type of language that AI produces, and that there is an internal lessening of engagement due to this internal, although perhaps unconscious, association of the AI text with AI outputs. However, this lessening of engagement did not lead to a direct identification of the less engaging text as AI generated, as discussed later.

Additionally, engagement with AI texts was related to order presentation, such that individuals who read the AI story after the human story demonstrated smaller levels of engagement with the AI text than individuals who read the AI text before the human text. This indicates that when reading an AI text in the light of a human text which had just been read, the AI text is significantly less engaging. It may be possible that in this order of presentation, individuals are unable to escape from comparing the passage and their experience of it back to the passage they had just read, while those who view the AI text have no immediate experience to refer to. This might then suggest that the AI texts are in themselves less engaging, despite the

lack of an overall effect of origin on engagement scores. It is very possible that the differences in engagement felt by the individuals themselves was not large enough to be subjectively noticeable, and that the difference in engagement has to be larger in order to have an effect on accuracy.

### **Accuracy of Origin Detection**

Regarding the question of distinction between texts, participants were not able to distinguish between AI generated and human written short literary fiction stories at a level greater than chance. This accuracy in origin detection was unrelated to any other measured variable, including education, familiarity with ChatGPT, level of engagement with different textual medias, and confidence in one's choice. This finding aligns with those of past research which have found that individuals are unable to distinguish AI reliably or significantly from human authorship. Nominally, more participants correctly identified the AI origin passage (43 participants) than mislabeled the AI text as human (32 participants) but this did not reach significance. This finding indicates that AI texts are similar to human texts in many ways and are able to "pass" as human writings in many cases, even in a literary form which may seem to require more creativity and technical writing ability than others.

It may very well be the case that engagement is not a method through which individuals may make their decisions regarding text origin, or that larger engagement is not an indicator to individuals of human writing. This may be suggested given that the directionality of engagement and accuracy were oppositional, where engagement was slightly higher with AI texts and accuracy was slightly more effective than chance, although neither of these were significant. However, qualitative answers frequently made references to elements of the Narrative

Engagement Scales' factors (i.e. comprehension of plot) as reasons for origin decisions, so this finding would be surprising. It is perhaps more likely that engagement is a method of distinguishing between texts, but that engagement between the texts in this study was not sufficient for participants to use it as a method of decision-making. The specificity and more impenetrable format of literary short fiction may interfere with engagement with human texts, hindering an engagement distinction between the AI and human texts. Other forms of literature which may be less dense or "artistic" may reveal different patterns between engagement and origin detection.

The secondary hypotheses in this study were that there would be an association between the accuracy of origin detection and both the knowledge of ChatGPT and textual media interaction. However, there was no statistical significance in either of these relationships. It was believed that individuals with more exposure to linguistic media would better recognize the patterns and systems of natural human language in opposition to artificial language. This may not have been found in this case because literary fiction is a specific subset of fiction that not all people, even those with high exposure to other forms of literature, may interact with frequently. It will be important in the future to test different forms of fictional literature within the paradigm of human and AI writing comparisons to determine which forms of literature AI can reproduce similarly. However, this relationship between exposure to textual media may also not exist if AI can mimic the patterns and elements of human well enough that any extra levels of exposure to textual media does not impact abilities to detect origins.

The finding that experience and knowledge of ChatGPT is an important predictor of distinguishing authorship has been made before (Hulman et al., 2023), and it may be the case that the single item measure was not enough to fully measure knowledge of and experience with

ChatGPT. It is certainly possible that there is not a wide enough distribution of ChatGPT knowledge and experience in this sample to perform adequate analyses, and that utilizing larger samples may bring in wider experiences to detect possible effects between these variables.

A possible paradigm (Figure 1) suggests that the engagement with the human text is key to the relationship between ChatGPT familiarity and origin detection. Difficulty engaging with the human text may interfere with a large, conscious distinction between engagement with the AI and human texts, which may have indicated to the individual which passage is the AI generated text. Other significant effects are incorporated into this theoretical model. Familiarity with AIs may result in decreased AI text engagement, as may viewing the AI text after the human one, while difficulties engaging with the human text may interfere with a noticeable difference in engagement between the texts, which may in turn interfere with accurate origin detection.

Exploratory analyses between origin detection and all other variables collected were also not significant. The logistic regression conducted testing the contribution of different variables with the outcome of origin detection did not determine any variable to be a significant predictor of accuracy. It is possible that predicting which individuals will be accurate in origin detection may be based off obscure or more subjective variables that were not considered here.

### **Qualitative Results**

The single qualitative item in this study allows for insights into the reasonings for passages identified as AI and any common features that people use to distinguish between texts, and what they associate with human writing and artificial writing. Of the six themes, only references to the theme of Scripted/Simplicity showed differences within accuracy, with significantly more people who referenced this theme correctly identifying the AI passage. References were made to a lack of depth in the story's meaning, repetitiveness, familiarity to previously known plotlines, simple character relationships, and guessable elements. Stories that

seemed human were referenced to create unexpected tension and suspense, have deeper meaning, have a ‘dream-like’ writing style, and be more imaginative, regardless of accurate origin detection. This finding suggests that the presence or lack of scriptedness and simple meanings is a telling feature of AI writing, either because it is very prevalent and noticeable within these writings or because people are especially sensitive to these elements of stories. It is possible that scriptedness is an element of writing that individuals associate specifically with AI generation, and that if one notices a story to be overly scripted or simple, they may attribute that to an AI authorship. Other themes may not be associated with AI authorship as strongly in individual’s minds, and noticing these themes may not then lead directly to attribution of AI origin.

No statistical significances were seen between engagement and references to qualitative themes; however, the qualitative results seem to suggest the possibility of engagement as a method for making decisions on text origin. Many qualitative answers reference constructs of Narrative Engagement. The concept of narrative understanding, of a comprehension of plotlines and character information, is referenced in many qualitative answers and encompassed within the theme of Clarity & Comprehension, wherein individuals referenced being confused by plotlines or one story making more sense than another. This confusion was referenced in some cases to detract from the ability to follow the storyline and therefore to take the reader out of engagement with the passage. Regarding the engagement construct of attentional focus which considers one’s mind wandering or being drawn into the story, less references were made, perhaps due to the short length of the passages. The same could be said for narrative presence, which considers the extent to which a person experiences transportation out of the real world and into the story world, and to which no references were made. References were made, however, to emotional engagement, with individuals referencing the level of emotional connection they felt that the

characters or the story evoked. On the whole, it seems as though elements of engagement were indeed considered when making decisions regarding passage origin, and that the engagement felt between the texts may have interfered more with origin detection given that references to the presence of engagement existed in both the AI and human texts.

Perhaps the hallmark of the qualitative analysis section of this project are the contradictions made between participants. One participant indicated that insight into a main character's life outside of their job was indicative of human origin, while another participant within the same story theme indicated that they believed that this same element was one that AI would add in. In this case, the latter participant was misled, misattributing this same element to an AI author rather than a human one. Two more participants, also in the same theme, each indicated the presence of "random details" as either an indicator of AI writing or of human writing. Repetition of words, phrases, or specific descriptions seemed to be a consistent indicator of AI – the 5 participants who specifically referenced repetition were all correct in their origin detection. References to the plot and content of stories were the most frequent, including presence of details, cohesive or disconnected story lines, and differences in plots between passages that indicated which passage was AI. These contradictions make clear the need for more in-depth, rigorous, and larger analysis of qualitative responses to AI texts, in addition to direct linguistic analysis of the texts.

### **Limitations**

While this study investigates some very important questions and provides evidence regarding human engagement with AI texts and abilities to distinguish between them, there are a number of limitations in this study, some of which are inherent in such an exploratory and new field of questioning. Firstly, the choice of stimuli and development of prompts creates multiple

difficulties. The stimuli used here are literary flash fiction, and direct generalization of any findings here to other forms of text or literature should not necessarily be considered. It is possible that findings will extend to these other forms but should not be assumed. Additionally, selection of human passages could be done more rigorously, or passages could be selected or generated in many other ways. The generation of prompts to be fed into ChatGPT also presents a difficulty in determining the level of detail and direction to be included, and the prompts for this paper may have been better at matching human literature passages if more detail was given. Additionally, no comprehension or attention checks were administered, and therefore it could be argued that participants did not attend to the passages enough to accurately answer questions or notice differences. However, qualitative answers do demonstrate a certain level of attention paid to passages.

Another limitation of this study is the number of participants and the demographic makeup of the sample. The majority of the participants had a bachelor's degree or had not graduated college, limiting any possible interaction analysis between education and results. It is very possible that there may be effects of education on engagement or detection accuracy that were not detected in this study due to the limitations in demographics that a study with a large participant sample and larger distribution of education would be able to detect. The same may be said for any other associations that may be detected with a larger sample size.

An additional limitation revolves around the measurement of certain variables, especially whether participants were correct or incorrect and the level of textual engagement. The labels for different points on the scale were not entirely continuously additive, and it is difficult to say how participants understood the scale. It is likely that they followed the general conception of gradual increases over a Likert scale, and paid attention mostly to the anchoring labels, but it may be the



case that the results on this measure are inaccurate due to this measurement error. Additionally, by making the accuracy variable a forced choice, where the participant marks one passage as AI, thereby indicating that the other must be human, it is often difficult to say whether choices were made arbitrarily. It is true that the inclusion of a confidence measure for this choice mitigates this concern, but an accuracy item which allowed ‘Both’ or ‘Neither’ options may provide a more nuanced and accurate look on individuals’ interactions with these texts. This organization was not used in this study largely due to concerns about statistical power when analyzing between four choices, but future studies with more power should consider providing more answer options and not indicating with certainty that any stimuli was artificially generated, thereby allowing for analysis of multiple dimensions.

### **Future Research**

There are numerous different paths for this research to take in the future. Replications of this study with larger sample sizes is essential for gaining statistical power and to capture various levels of education and expertise with forms of literature. Utilizing multi-dimensional measures of accuracy will be essential in understanding directionality – whether human writing seems to resemble AI, or whether AI writing resembles human. The qualitative answers here suggest that both may be possible. Studies of objective differences between human and AI writing (e.g. Sardinha, 2024) may be essential in determining whether these are the same elements which individuals report as the reasons for their decisions, whether these are elements of writing that are objective but not noticed by readers, or whether there are influences on origin detection which are undetectable using objective comparisons between texts.

The stimuli utilized here is a very specific type of literature, of short literary fiction, and the results will likely not stretch to other forms of literature. Future research should certainly

consider the linguistic variances that are the most susceptible to interference from artificial writings and test these accordingly. Testing the ability of these programs to generate scripts, marketing materials, speeches, webpages, educational materials, and even academic research literature (as has been done before) will be essential to understand the abilities of these programs to ‘pass’ as human. As the abilities of these programs advance, they may become more effective at mimicking human speech, and it would be important to replicate results utilizing more advanced programs. Additionally, changing the specificity of the prompt to the AI program and editing the type of information that is given to it may also be very important in determining the abilities of these programs to produce different forms of literature and to respond to different prompts, and how humans respond to literature generated by different types of prompts.

A very intriguing area of future research should be the effects of origin knowledge and attitudes towards AI in affecting engagement with AI literature or other materials and accuracy of origin detection. These factors would likely affect the ability to distinguish passage origin or reveal interesting information about whether positive or negative attitudes towards AI may affect engagement with passages suspected or known to be AI. Additionally, measuring in what ways people interact with AI programs and their attitudes about the outputs may be important to discover how these factors might influence ability to distinguish between AI products.

Perhaps most importantly, although more broadly, is the ethical research on the use and dissemination of these programs and their materials. This has already become an issue surrounding the use of author’s material in training these models and in the television media industry, as discussed previously. Understanding the abilities and perception in the public of these programs is essential to considering how they will be used and what ethical considerations

they may break, and what limitations should or can be placed on their use. It is certainly a critical time for dictating what will happen in the coming years regarding these technologies.

It is certainly up for more deliberation and research to determine the extent to which the impacts of AI in the literature and cognitions spheres will be on the whole beneficial or debilitating, but what is certainly not up for debate is the fact that the decisions, creations, and precedence set down in the current times will stretch into the future, and careful considerations should be attributed to many different paradigms and questions within this overarching field of research.

### References

- 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. <https://doi.org/10.3389/fnhum.2012.00192>.
- Bedingfield, W. (2023b, May 8). *Hollywood's Screenwriters Are Right to Fear AI*. Wired; Condé Nast. <https://www.wired.com/story/hollywood-screenwriters-artificial-intelligenceguardrails/>.
- Butlin, Patrick, Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S.M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M.A.K, Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv Preprint*. [Arxiv.org/abs/2308.08708](https://arxiv.org/abs/2308.08708).
- Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, 18(2), 255–280. <https://doi.org/10.1111/j.1468-2885.2008.00322.x>
- Busselle, R. & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, 12(4), 321-347. <http://dx.doi.org/10.1080/15213260903287259>
- Busselle, R., & Vierrether, T. (2022). Linking epistemic monitoring to perceived realism: The impact of story-world inconsistency on realism and engagement. *Media Psychology*, 25(5), 1–17. <https://doi.org/10.1080/15213269.2022.2032180>
- Child, B. (2023, May 12). AI is coming for Hollywood scriptwriters – this is how they are going to do it. *The Guardian*. <https://www.theguardian.com/film/2023/may/12/ai-artificialintelligence-generating-screenplays>

Coyle, J. (2023, May 5). *A.I. is one of the main reasons that Hollywood writers are on strike:*

“Too many people are using it against us and using it to create mediocrity.” Fortune;

Fortune Media. <https://fortune.com/2023/05/05/writers-strike-hollywood-ai-scripts/>

Csikszentmihalyi, M. (1997). *Finding flow: The Psychology of Engagement with Everyday Life*.

Basic Books.

Cupchik, G.C., Oatley, K., & Vorderer, P. (1998). Emotional effects of reading excerpts

from short stories by James Joyce. *Poetics*, 25, 363-377.

del Barco, M. (2023, May 18). *Striking Hollywood scribes ponder AI in the writer's room*. NPR.

<https://www.npr.org/2023/05/18/1176876301/striking-hollywood-writers-contemplate-ai>

Demmer, T.R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI-

versus human-created art and impact on aesthetic experience. *Computers in Human*

*Behavior*, 148. <https://doi.org/10.1016/j.chb.2023.107875>.

Dijkstra, N., & Fleming, S.M. (2023). Subjective signal strength distinguishes reality from

imagination. *Nature Communication* 14, 1627 <https://doi.org/10.1038/s41467-023-37322>.

Eekhof, L. S., Kobie van Krieken, Sanders, J., & Willems, R. M. (2023). Engagement with

narrative characters: The role of social-cognitive abilities and linguistic viewpoint.

*Discourse Processes*, 60(6), 411–439. <https://doi.org/10.1080/0163853x.2023.2206773>

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the

labor market impact potential of large language models. *ArXiv Preprint*.

<https://arxiv.org/abs/2303.10130>

Eysenbach G. (2023). The role of ChatGPT, Generative Language Models, and artificial intelligence in medical education: A Conversation with ChatGPT and a call for

papers. *JMIR medical education*, 9, e46885. <https://doi.org/10.2196/46885>

- Floridi, L., Taddeo, M., and Turilli, M. (2008). Turing's imitation game: Still an impossible challenge for all machines and some judges. *Minds and Machines*, 19.145–50.
- French, R. (1990). Subcognition and the Limits of the Turing Test. *Mind*, 99. 53–65.
- French, R. M. (2000). *The Turing Test: the first 50 years*. *Trends in Cognitive Sciences*, 4(3), 115–122. doi:10.1016/s1364-6613(00)01453-4
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, 6(1).  
<https://doi.org/10.1038/s41746-023-00819-6>
- Gunderson, K. (1964). Descartes, La Mettrie, Language and Machines. *Philosophy* 39. 193–222.
- Haenlein, M. & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4).  
<https://doi.org/10.1177/000812561986492>
- Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series*, 1387. 10.1088/1742-6596/1387/1/012020
- Hayes, P. and Ford, K. (1995) Turing Test considered harmful. In Proc. Fourteenth IJCAI-95, Montreal, Canada (Vol. 1), pp. 972–977.
- Hogan, P. C. (2018). *Literature and Emotion*. New York: Routledge.
- Adam Hulman, Ole Lindgård Dollerup, Jesper Friis Mortensen, Fenech, M., Norman, K., Henrik Støvring, & Troels Krarup Hansen. (2023). ChatGPT- versus human-generated answers to frequently asked questions about diabetes: A Turing test-inspired survey among employees of a Danish diabetes center. *PLOS ONE*, 18(8), e0290773–e0290773.  
<https://doi.org/10.1371/journal.pone.0290773>

Johnson-Laird, P.N & Oatley, K. (2022). How poetry evokes emotions. *Acta Psychologica*, 224. <https://doi.org/10.1016/j.actpsy.2022.103506>.

Kang, J., Hong, S., & Hubbard, G. T. (2020). The role of storytelling in advertising: Consumer emotion, narrative engagement level, and word-of-mouth intention. *Journal of Consumer Behaviour*, 19(1), 47–56.

Kasneci, E., Sesler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

Khamis H. (2008) Measures of Association: How to Choose? *Journal of Diagnostic Medical Sonography*. 24(3):155-162. doi:10.1177/8756479308317006

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 1–13. <https://doi.org/10.1177/00336882231162868>

LaCurts, K. (2011). Criticisms of the Turing Test and why you should ignore (most of) them. Official blog of MIT's course: Philosophy and Theoretical Computer Science.

Landgrebe, J. & Smith, B. (2019). There is no general AI: Why Turing machines cannot pass the Turing test. *arXiv Preprint*. <https://arxiv.org/abs/1906.05833>.

Liang, J., Hwang, G., Chen, M., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*. doi: 10.1080/10494820.2021.1958348

- Maddaus, G. (2023, March 21). *WGA would allow Artificial Intelligence in scriptwriting, as long as writers maintain credit*. Variety. <https://variety.com/2023/biz/news/writers-guildartificial-intelligence-proposal-1235560927/>
- Mar, R. A., Oatley, K., Djikic, M., & Mullin, J. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition & emotion*, 25(5), 818–833. <https://doi.org/10.1080/02699931.2010.515151>
- Mar, R. A., Oatley, K., dela Paz, J., Hirsh, J., & Peterson, J. B. (2006). Bookworms vs. nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and simulation of fictional social worlds. *Journal of Research in Personality*, 40, 694–712.
- Mauldin, M.M. ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize competition. Proceedings of the AAAI Conference on Artificial Intelligence. <https://aaai.org/papers/00016-aaai94-003-chatterbots-tinymuds-and-the-turing-testentering-the-loebner-prize-competition/>.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>
- McClelland, T. (n.d.). *Will AI ever be conscious?* Clare College. <https://stories.clare.cam.ac.uk/will-ai-ever-be-conscious/index.html>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023, August 18). *A comprehensive overview of large language models*. ArXiv.org. <https://doi.org/10.48550/arXiv.2307.06435>
- Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 5374.



Oatley, K. (2002). Emotions and the story worlds of fiction. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 39-69).

Lawrence Erlbaum Associates, Inc.

OpenAI. (2023). GPT-4. *OpenAI*. <https://openai.com/research/gpt-4>.

OpenAI. (n.d.) ChatGPT. *OpenAI*. <https://openai.com/chatgpt>

Pikhart, M. (2020). Intelligent information processing for language education: The use of artificial intelligence in language learning apps. *Procedia computer science*, 176, 1412–1419. <https://doi.org/10.1016/j.procs.2020.09.151>

Rainie, L., Funk, C., Anderson, M., & Tyson, A. (2017). AI and human enhancement: Americans' openness is tempered by a range of concerns. *Pew Research Center*

Russell, S. J., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd. ed.). Pearson.

Sardinha, T.B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics* 4.

Schmälzle, R., Wilcox, S., & Jahn, N. T. (2022). Identifying moments of peak audience engagement from brain responses during story listening. *Communication Monographs*, 89(4), 1–24. <https://doi.org/10.1080/03637751.2022.2032229>

Tang, T. & Mahoney, L.M. (2020). Narrative Engagement Scale. In *Communication Research Methods III: A Sourcebook* (eds. Elizabeth E. Graham and Joseph P. Mazer). Routledge.

The Author's Guild. (2023). *Author's Guild Open Letter to Generative AI Leaders*.

Actionnetwork.org. <https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders>

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236).

van Laer, T., Feiereisen, S. and Visconti, L. M. (2019). Storytelling in the Digital Era: Relevant

- Moderators of the Narrative Transportation Effect. *Journal of Business Research*, 96, pp. 135-146. doi: 10.1016/j.jbusres.2018.10.053
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., & Isaac, W. (2021). Ethical and social risks of harm from Language Models. *ArXiv Preprint*. <https://arxiv.org/abs/2112.04359>
- Winston, P.H. (2011). The strong story hypothesis and the directed perception hypothesis. *Proceedings of the AAAI Fall Symposium Series*, 345–352.
- Worth, S. (2004). Fictional spaces. *Philosophical Forum*, 35, 439–455.

Table 1. Sociodemographic Characteristics of Participants

Baseline characteristic	Call Theme		Dinosaur Theme		Kiss Theme		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Language								
First Language	16	59.3	21	77.8	16	80.0	53	71.6
English								
First Language other than English	11	40.7	6	22.2	4	20.0	21	28.4
Highest educational level								
High School	2	7.4	0	0.0	2	10.0	4	5.4
College (not graduated)	9	33.3	10	37.0	6	30.0	25	33.8
Bachelor's Degree	10	37.0	9	33.3	8	40.0	27	36.5
Master's Degree	4	14.8	7	25.9	2	5.0	12	16.2
Doctorate	2	7.4	1	3.7	3	15.0	6	8.1

*Note.*  $N = 75$ . Participants were on average 28.47 years old, with a standard deviation of 12.29 years.

Table 2. One-Way ANOVA for the Means of NES Engagement Scores by Story Theme.

		ANOVA				
		Sum of Squares	df	Mean Square	F	Sig.
NES Human Score	Between Groups	6.809	2	3.404	2.970	.058
	Within Groups	81.394	71	1.146		
	Total	88.202	73			
NES AI Score	Between Groups	2.729	2	1.364	1.317	.274
	Within Groups	73.535	71	1.036		
	Total	76.264	73			

Table 3. Correlations Between all Variables

Correlations												
		1. NES Human Score	2. NES AI Score	3. Age	4. Education	5. First Language English	6. Textual Media Engagement	7. Confidence in Reading Ability	8. Familiarity with ChatGPT	9. Confidence in Choice	10. Passage Presentation Order	11. Choice Accuracy
2	Correlation Coefficient	.077										
	Sig. (2-tailed)	.512										
	N	75										
3	Correlation Coefficient	-.022	.157									
	Sig. (2-tailed)	.852	.179									
	N	75	75									
4	Correlation Coefficient	-.140	-.034	.678**								
	Sig. (2-tailed)	.231	.771	<.001								
	N	75	75	75								
5	Correlation Coefficient	-.028	.075	.072	-.095							
	Sig. (2-tailed)	.814	.520	.541	.418							
	N	75	75	75	75							
6	Correlation Coefficient	.179	.058	.000	.158	-.004						
	Sig. (2-tailed)	.125	.623	.997	.175	.972						
	N	75	75	75	75	75						
7	Correlation Coefficient	.161	.140	-.049	.096	.225	.330**					
	Sig. (2-tailed)	.167	.231	.675	.410	.052	.004					
	N	75	75	75	75	75	75					
8	Correlation Coefficient	-.034	-.302**	-.132	.231*	-.391**	.046	.007				
	Sig. (2-tailed)	.774	.008	.258	.046	<.001	.697	.950				
	N	75	75	75	75	75	75	75				
9	Correlation Coefficient	.187	-.091	-.108	-.027	-.056	.068	.044	.049			
	Sig. (2-tailed)	.108	.438	.356	.815	.633	.560	.709	.678			
	N	75	75	75	75	75	75	75	75			
10	Correlation Coefficient	-.074	.295*	.110	.056	-.024	-.040	-.013	-.146	-.052		
	Sig. (2-tailed)	.530	.010	.348	.633	.836	.735	.915	.211	.656		
	N	75	75	75	75	75	75	75	75	75		
11	Correlation Coefficient	.011	-.099	-.031	.016	.062	-.015	.005	.156	-.007	-.209	
	Sig. (2-tailed)	.922	.400	.791	.894	.595	.901	.968	.180	.952	.072	
	N	75	75	75	75	75	75	75	75	75	75	
Story Theme	Correlation Coefficient	.046	.058	.033	-.025	-.178	.133	.120	.080	.024	-.005	-.026
	Sig. (2-tailed)	.615	.526	.721	.802	.104	.150	.247	.428	.811	.963	.815
	N	75	75	75	75	75	75	75	75	75	75	75

Note: The correlations involving Accuracy and First Language are calculated with Pearson's pointbiserial correlation coefficient, and all other correlations with Story Theme are conducted with Kendall's  $\tau_b$ ; all other correlations are Spearman's rank correlation coefficient, following guidelines regarding data type and appropriate correlations (Khamis, 2008).

Table 4. Logistic Regression for Accuracy

Logistic Regression Statistics: Model 2 for Accuracy (10 variables, n=75)								
	R-squared (McFadden)	Adj.R-Sqr.	RMSE	Mean	# Fitted	ROC area	Critical z	Conf. level
	0.057	0.000	0.475	0.573	75	0.66	1.960	95.0%
Logistic Regression Coefficient Estimates: Model 2 for Choice.variable (10 variables, n=75)								
Variable	Coefficient	Std.Err.	z-statistic	P-value	Lower95%	Upper95%	VIF	Std. coeff.
Constant	0.669	1.381	0.485	0.628	-2.037	3.375		
Age	-0.000487	0.027	-0.018	0.985	-0.052	0.051	1.899	-0.003
Confidence in Choice	-0.041	0.280	-0.146	0.884	-0.589	0.508	1.079	-0.021
Education Level	0.044	0.331	0.132	0.895	-0.605	0.693	1.891	0.025
English 1st Language	0.710	0.634	1.120	0.263	-0.533	1.953	1.366	0.177
ChatGPT Familiarity	0.375	0.293	1.279	0.201	-0.200	0.950	1.815	0.232
NES AI Score	0.042	0.263	0.159	0.874	-0.473	0.557	1.337	0.025
NES Human Score	0.072	0.241	0.299	0.765	-0.400	0.544	1.170	0.044
Passage Presentation Order	-0.761	0.545	-1.396	0.163	-1.829	0.308	1.181	-0.207
Confidence in Reading Ability	-0.161	0.323	-0.498	0.619	-0.793	0.472	1.276	-0.076
Textual Media Engagement	-0.102	0.403	-0.253	0.800	-0.892	0.688	1.228	-0.038

Table 5. Accuracy of Origin Detection by Education Level

		Accuracy				Total	
		Incorrect		Correct			
		N	%	N	%	N	%
Education Level	High School	1	3.1%	3	7.0%	4	5.3%
	College, not Graduated	12	37.5%	13	30.2%	25	33.3%
	Bachelors	10	31.3%	17	39.5%	27	36.0%
	Masters	8	25.0%	5	11.6%	13	17.3%
	Doctorate	1	3.1%	5	11.6%	6	8.0%
Total		32	100.0%	43	100.0%	75	100.0%

Table 6. Qualitative Analysis Dimensions and Examples.

Dimension and Explanation	Example Quote
<p>Scripted and Simplicity</p> <p>Mentions of story or writing sounding scripted or predictable or of presence or lack of imagination/depth in the writing.</p>	<p>“Was less engaging seemed like followed more basic plot putting together storylines that happened before.”</p>
<p>Clarity vs. Complexity</p> <p>Noting that passages were either too complex to understand or very clear and understandable.</p>	<p>“The passage was just very hard to understand and follow. I couldn't stay engaged with it.”</p>
<p>Plot and Content</p> <p>Differences in storyline between passages, inclusion or lack of elements of real life or activity in story, inclusion or lack of details, and cohesive or disconnected storyline.</p>	<p>“Passage 1 was not as cohesive as passage 2. Details were added in passage 1 that were not needed.”</p>
<p>Character Depth</p> <p>Mentions a lack of depth in characters as a notifier of AI, or deep and engaging characters as a notifier of human authors.</p>	<p>“Additionally, I thought that the characters were more developed and individualized in the second passage, with the main character</p>

being given some insight into her life outside of her night shift job.”

Emotionality	“I Think the first one had much more depth, words were carefully chosen. the second one transmitted less emotions”
Mentions of presence or lack of emotions in the story or evoked in the reader.	
Grammar and Organization	“Too many grammatical errors. I would guess that a human author would be a better editor.”
Mentions of grammar (quotation marks, language choice, use of certain parts of speech) as well as passage organization (eg. spaces between paragraphs) as notifier of either AI or human.	

Table 7. Accuracy within Qualitative Themes

Theme	Incorrect	Correct	Total	Phi Value	Phi Sig.
SS	4	16	20	.276	.017*
CC	9	11	20	-.028	.805
CD	3	4	7	-.001	.991
E	5	6	11	-.023	.840
PC	15	17	32	-.073	.525
GO	6	10	16	.054	.638

Note: ‘SS’ is scripted/simplicity; ‘CC’ is clarity/complexity; ‘CD’ is character depth; ‘E’ is emotionality; ‘PC’ is plot/content; ‘GO’ is grammar/organization. Phi values and significance are calculated for accuracy in origin detection by each theme.

\*. Correlation is significant at the 0.05 level (2-tailed).

Table 8. Qualitative Themes by Story Theme

Theme	References	Phi Value	Phi Sig.
SS	20	.176	.312
CC	20	.252	.092
CD	7	.106	.655
E	11	.089	.743

PC	34	.044	.929
GO	16	.114	.615

Table 9. Correlations Between Qualitative Themes and Engagement

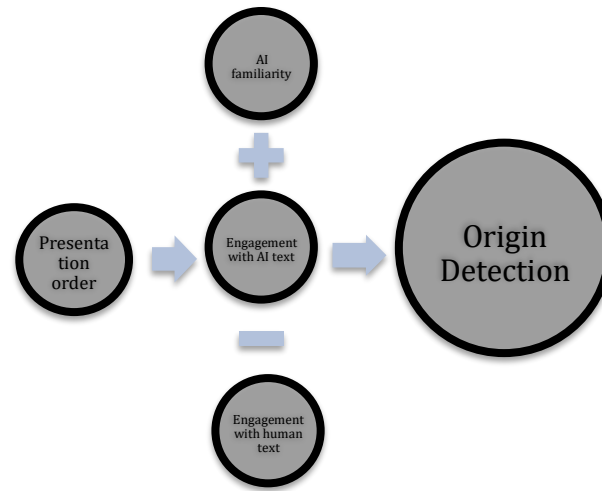


		1	2	3	4	5	6	7	8	9
1. NES Human Score	Pearson Correlation	--								
	N	75								
2. NES AI Score	Pearson Correlation	.069	--							
	Sig. (2-tailed)	.556								
	N	75	75							
3. SS Theme	Pearson Correlation	-.077	.064	--						
	Sig. (2-tailed)	.509	.587							
	N	75	75	75						
4. CC Theme	Pearson Correlation	.110	-.152	-.227*	--					
	Sig. (2-tailed)	.349	.193	.050						
	N	75	75	75	75					
5. CD Theme	Pearson Correlation	-.074	.034	.014	-.090	--				
	Sig. (2-tailed)	.528	.771	.906	.443					
	N	75	75	75	75	75				
6. GO Theme	Pearson Correlation	-.132	-.172	-.093	-.093	-.167	--			
	Sig. (2-tailed)	.261	.140	.426	.426	.152				
	N	75	75	75	75	75	75			
7. E Theme	Pearson Correlation	.000	-.007	.006	.091	-.003	-.032	--		
	Sig. (2-tailed)	.998	.949	.961	.438	.977	.786			
	N	75	75	75	75	75	75	75		
8. PC Theme	Pearson Correlation	-.107	.027	-.065	-.246*	-.016	-.147	.001	--	
	Sig. (2-tailed)	.361	.818	.582	.033	.892	.207	.993		
	N	75	75	75	75	75	75	75	75	
9. E and CD Themes	Pearson Correlation	.009	.032	.034	.034	.593**	-.126	.766**	.019	--
	Sig. (2-tailed)	.938	.787	.775	.775	<.001	.280	<.001	.873	
	N	75	75	75	75	75	75	75	75	75

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Figure 1. Theoretical Model of Interactions of Engagement, Origin Detection, and Variables



Appendices.

## Appendix A

### ChatGPT Flash Fiction Passages<sup>1</sup>

#### PROMPT 1:

Could you please write for me a short story between 700 and 800 words. This story should follow a female first-person narrator who works as a calling agent for a diet product company and talks to a customer on the phone to inform them that their package is on the way. Your of the story should include a good amount of dialogue and some personal details about the other woman on the phone, and should in some way be 'distilled, cogent, concentrated, layered, allusive, textured, insightful, and unpredictable'.

---

The incessant hum of fluorescent lights filled the cramped office space as I settled into my worn-out chair. My fingers tapped rhythmically on the keyboard, entering data into the computer with a mechanical precision that had become second

---

<sup>1</sup> All ChatGPT passages are generated within a brand-new conversation with the intent to avoid interactions from previous exchanges in the generation of the passage.

nature. As a calling agent for a diet product company, my days were consumed by the repetitive drone of scripted conversations, scripted smiles, scripted life.

The call queue light flickered, signaling an incoming call. I took a deep breath, ready to immerse myself in another predictable conversation about weight loss miracles. The familiar beep sounded, and I answered with the rehearsed enthusiasm that came with the job.

"Hello, this is Lisa from TrimTech, and I'm thrilled to inform you that your package is on its way!" I chirped into the headset, my voice layered with practiced cheerfulness.

A hesitant pause lingered on the other end before a woman's voice, laden with both weariness and curiosity, responded. "Oh, finally. I've been waiting for that. It's been a long time since I ordered it."

"Thank you for your patience," I replied, scanning through the customer's profile on my screen. "I see your name is Sarah. How has your day been so far?"

There was a hint of a sigh before Sarah answered, "Well, you know, the usual. Juggling work, family, and trying to stay healthy. That's why I ordered this package, hoping it'll make things a bit easier."

I sensed a weariness in her voice, as if her life was a puzzle missing a few crucial pieces. "I completely understand, Sarah. We all strive for that elusive balance. So, tell me, what made you choose TrimTech?"

A subtle laugh escaped her lips. "I guess it was the promise of a shortcut, a way to reclaim control without sacrificing too much. But life is never that simple, is it?"

I was taken aback by her honesty, the unscripted vulnerability in her words. In that moment, the conversation shifted from a routine call to a genuine exchange between two women navigating the complexities of modern life.

"It rarely is," I agreed, my tone softened. "But sometimes, these small steps can be a catalyst for bigger changes. Have you tried any other programs before?"

There was a moment of silence before she confessed, "A few, but nothing stuck. Life has a way of derailing the best-laid plans."

As Sarah spoke, I imagined the layers of her life, the hidden struggles and triumphs beneath the surface. I was drawn to her openness, an unexpected connection forming through the sterile medium of a phone line.

"You're not alone in that, Sarah. Life has its own script, and we're all just trying to find our place in the narrative," I mused, stepping outside the confines of the company's script for the first time.

Her laughter, genuine and warm, echoed through the line. "You're right, Lisa. I never expected a call like this to turn into a philosophical conversation."

"Life is full of surprises," I replied, a smile playing on my lips. "And speaking of surprises, how about we add a personal touch to your package? We have a new recipe book that complements our products perfectly. I think you'll love it."

Sarah's curiosity piqued. "A recipe book? That sounds interesting. What kind of recipes are we talking about?"

"It's a mix of healthy and delicious meals, tailored to support your weight loss journey. I'll make sure it's included in your package," I assured her, enjoying the deviation from the usual sales pitch.

As we continued to chat, the conversation became a dance of shared experiences and genuine connection. I learned about Sarah's passion for painting, her dreams, and the subtle joys that anchored her in the chaos of everyday life. The scripted barriers between caller and customer dissolved, leaving behind a conversation that was distilled, cogent, and textured with the nuances of shared humanity.

Before ending the call, I couldn't resist sharing a piece of advice that transcended the confines of diet products and sales targets. "Sarah, life may not always go according to plan, but sometimes the detours lead us to unexpected joys. Embrace the surprises and find beauty in the unplanned."

As I hung up, I couldn't shake the feeling that this call had been more than just another task on my to-do list. In the sterile office environment, I had discovered a genuine connection, a reminder that beneath the scripted routines of life, there was always room for authentic moments of connection and understanding.

#### PROMPT 2:

Could you please write a short story, between 700 and 800 words, in first-person narration, with the female narrator being a previous EMT now working at a dinosaur centered amusement park to teach people about dinosaurs, where there are also life-size dinosaur models. This narrator becomes involved in a discussion with a child about the park and the dinosaurs. Your story should include a good amount of dialogue and personal details, and should in some way be 'distilled, cogent, concentrated, layered, allusive, textured, insightful, and unpredictable'.

---

The sun hung low in the sky, casting long shadows over the dinosaur-centered amusement park. Nestled among the towering life-size dinosaur models, I marveled at the intricate details of their ancient forms. My role here wasn't just a job; it was a passion—a journey back in time to teach people about the magnificent creatures that once roamed the Earth.

On this particular day, a wave of excitement rippled through the air as families and children explored the park. Among the crowd, a young boy caught my eye. His eyes gleamed with curiosity, fixated on the colossal Tyrannosaurus rex that stood before him.

I approached the boy and his family, a smile playing on my lips. "Hey there! Are you enjoying the park?"

The boy nodded eagerly, his enthusiasm contagious. "Yeah! These dinosaurs are awesome. Are they real?"

I crouched down to his eye level, noting the dinosaur-themed T-shirt he wore. "No, they're not real. But they're lifelike replicas, designed to help people like you learn about dinosaurs and the world they lived in."

His eyes widened, absorbing the information. "Wow, that's cool! What's your favorite dinosaur?"

I chuckled, thinking about the countless times I'd been asked that question. "It's tough to pick just one, but I've always had a soft spot for the Velociraptors. They were like the ninjas of the dinosaur world, smart and fast."

The boy grinned, clearly intrigued. "Velociraptors sound awesome! Do you have any here?"

I pointed toward the Velociraptor exhibit, nestled in the shadows. "Right over there. You should check it out; they're quite fascinating."

As the boy darted toward the exhibit, his family following close behind, I couldn't help but feel a sense of accomplishment. The park wasn't just a collection of life-size models; it was a gateway to a world long gone—a world I hoped to make accessible and exciting for each visitor.

A little while later, I found the same boy near the Triceratops exhibit, engrossed in studying the massive herbivore. I approached, noting the glint of curiosity in his eyes. "Finding everything okay?"

He nodded, his attention still fixed on the Triceratops. "Yeah! These dinosaurs are so big. How do you know so much about them?"

I smiled, appreciating his curiosity. "Well, I've always been fascinated by dinosaurs. I studied paleontology in college, and now, I get to share what I know with people like you."

He looked up at me with wide eyes. "You get to talk about dinosaurs all day? That's the coolest job ever!"

I chuckled, "It's pretty amazing, I won't lie. But the best part is seeing the excitement and curiosity on faces like yours."

As we continued our conversation, I sensed a layer of genuine interest in the boy's questions. His curiosity wasn't just about the dinosaurs; it was an exploration of possibilities, a thirst for knowledge and wonder that made my job all the more fulfilling.

He pointed toward the Pteranodon soaring overhead. "What about those flying dinosaurs? How did they fly?"

I explained the concept of flight in Pteranodons, their wingspan, and adaptations. His eyes widened with each detail, a textured tapestry of awe and fascination weaving itself in his expression.

As we chatted, an insightful question escaped his lips. "Do you think there are still dinosaurs out there, somewhere?"

The unpredictability of his inquiry caught me off guard. I smiled, appreciating the depth of his imagination. "Well, not the dinosaurs you see here, but birds are considered descendants of dinosaurs. So, in a way, there are still some dinosaur relatives flying around today."

The boy's eyes lit up, the notion of modern-day dinosaurs adding a layer of wonder to his understanding. We explored the park together, discussing various species, their behaviors, and the mysteries of a world long before ours.

As we approached the exit, the boy turned to me with a grin. "Thanks for telling me all about the dinosaurs. I want to be a paleontologist when I grow up!"

His declaration filled me with a sense of pride and accomplishment. The park wasn't just a place for entertainment; it was a platform for inspiration and learning.

Before parting ways, the boy's family expressed their gratitude, and I handed the young paleontologist a small dinosaur figurine—a memento of his adventure. The unexpected joy in his eyes mirrored the magic I hoped to instill in every visitor.

As the sun dipped below the horizon, casting an amber glow over the park, I reflected on the day's interactions. The park wasn't just a collection of life-size models; it was a conduit for shared experiences, sparking curiosity and leaving an indelible mark on the minds of those who ventured into our prehistoric world.

### PROMPT 3:

Could you please write a story told in first-person narration, between 900 and 1,000 words of the gathering of a group of friends, a total of 8 people, where one woman has recently gotten a tongue piercing, and she is showing it off to the group. The narrator wonders aloud what it would be like to kiss someone with a tongue piercing, and one of the members of the party decides to kiss her to find out. There should be a level of personal detail and description surrounding the kiss, and should in some way be 'distilled, cogent, concentrated, layered, allusive, textured, insightful, and unpredictable'.

----

The night was a canvas of laughter and shared memories as our group of friends gathered in the cozy living room. A tapestry of personalities, eight individuals with histories woven together in the intricate fabric of friendship. As we settled into the warmth of camaraderie, I couldn't help but notice the mischievous glint in Emma's eyes.

Emma, the vivacious redhead of the group, had recently ventured into the world of body piercings with a new addition that demanded attention—a silver stud glinting in the dim light of the room, nestled between her lips. She reveled in the glances and curious inquiries, using her newfound accessory as a conversation starter.

"Guess what, guys?" Emma exclaimed, her voice carrying the excitement of a storyteller about to unveil a secret. "I got a tongue piercing!"

The room erupted in a chorus of gasps and exclamations, each friend taking their turn to inspect the subtle rebellion adorning Emma's tongue. I couldn't help but be intrigued by the glimmering accessory, a small rebellion against the mundanity of our everyday lives.

As the conversation flowed, I found myself musing aloud, "I wonder what it's like to kiss someone with a tongue piercing." The room fell momentarily silent, the weight of my words hanging in the air like an unspoken challenge. The unexpected question had sliced through the chatter, leaving us all suspended in a moment of contemplation.

Without missing a beat, Jake, our resident free spirit with an easygoing smile, grinned mischievously. "Well, there's only one way to find out."

Laughter erupted, a collective release of tension as Jake's boldness broke the spell of the unexpected. Amidst the banter and teasing, Emma shot Jake a playful yet curious glance. The atmosphere had shifted, a charged energy buzzing in the room as we all became spectators to the unfolding narrative.

In the dim glow of the room, Jake leaned in, his eyes never leaving Emma's. The air thickened with anticipation as the room hushed, a collective held breath punctuated only by the distant hum of the city outside. The kiss was a promise, a daring exploration that transcended the physical act, delving into uncharted territory of friendships and desires.



As their lips met, there was an unexpected grace to the moment, a subtle acknowledgment of the unspoken curiosity that had lingered in the air. Emma's hands instinctively found their place on Jake's shoulders, her fingers weaving through the strands of his hair. The kiss was a dance, a choreography of emotions that shifted between playfulness and something deeper, an exploration of sensations both novel and familiar.

The taste of metal lingered on Jake's lips, a textured nuance that added a layer of complexity to the kiss. It was an allusion to a realm of experiences outside the ordinary, an exploration of boundaries that extended beyond the mere act of affection. The piercing, once a conversation piece, became a tangible connection, a shared secret that bound them in that fleeting moment.

As they parted, the room erupted into applause and cheers, the tension dissipating into a chorus of laughter. The unexpected kiss had been a collective adventure, a shared experience that transformed the evening into a memorable chapter in the book of our friendships.

Emma's cheeks flushed with a mix of surprise and amusement as she shot a sidelong glance at me, the originator of the whimsical query. In that exchange, we shared a silent acknowledgment of the unpredictable nature of life and the beauty that could be found in seizing the unexpected.

The night continued with its tapestry of shared moments, laughter echoing against the walls as we reveled in the joy of each other's company. The air had changed, charged with the energy of the unexpected kiss, a reminder that life's beauty often lies in the unplanned, the unscripted, and the courage to embrace the unknown.

As we bid each other farewell in the early hours of the morning, the memory of that kiss lingered, a layer of complexity woven into the fabric of our friendships. The unexpected had unfolded in the most unpredictable of ways, leaving us with a shared story that would be recounted with laughter and smiles in the years to come.

## Appendix B

### Human-Generated Short Story Passages

#### **PASSAGE 1:** “Pounds Across America” Meg Pokrass

Pokrass, M. (2011). “Pounds Across America.” *Damn Sure Right*, Press 53.

On Tuesday afternoon I line up with other petite brunette actresses, silently, our eyes underlined with dark liner. When it's my turn to walk onstage, the assistant casting director asks me to smile, inspects my teeth for flaws. She has purple hair, a nose ring, and a T-shirt that says

2nd Butch Bitch. She looks me over—back to front to back. Says they'll call if I make the cut.

I work in the fringes of Midtown Manhattan on the night shift, which allows me days to audition. My coworkers are mainly out-of-work actors. Our job is calling people who've ordered our diet product from a TV infomercial.

The floor manager creates a sales contest to motivate us, calls it POUNDS ACROSS AMERICA! We're all nervous, fluttering and bullying each other. I pile Three Musketeers bars next to my coffee. A bite, then a sip, then a call. I wave at Jeremy, who's been on the night shift the last month.

The prize is Broadway show tickets for two. I dial, opening my Three Musketeers.

"Yep?" a tired female voice says.

"Hi. Is this Janet?"

"Depends," she says.

"This is Martha Tiffany with Dr. Feldman's weight loss system! Congratulations, Janet! We've shipped your trial order and you should be receiving it anytime!" "Jingle-jangle-jesus!" says Janet D. Higgins, 190 pounds, in Racine.

"Janet, Dr. Feldman is having us call every customer individually so we can design your unique program. How many pounds do you need to lose?"

I can't help reaching for my Three Musketeers bar. I hear the pop of a fart from the young recruit behind me.

"Fifty," she says, followed by a puff of air."

"Great. How fast would you like to do that, Janet?" I ask, tonguing the caramel nougat. "Three weeks? Heh!"

"Let's see, I'm just looking at the chart," I say.

I turn to see what's happening. Dawn (who started when I did) is doing her shtick for a group in the back, saying, "Pee—niss," in a Mickey Mouse voice. "Pee-niss, pee-niss, pee-niss!"

Janet screams, "Mommy needs a little time-out too, honey."

“Janet, we’re looking at . . . ( here the script suggests to improvise) . . . two to three to four months if you follow the easy step system!”

I look over at Jeremy, his new haircut. He just did a national soda commercial—knows he’s hot. He’s rolling a joint under his desk, not really caring if he gets caught.

“I got to try something,” Janet says. I hear a child yelling.

“Let me get to the other reason I called . . . and this has to do with what we just talked about. We care about your success as much as you do, Janet, and we don’t want you to have a gap in your continuation—an important concept in weight loss. We’re real backed up here, Janet!

People are waiting for months to receive orders because of the success they’re achieving.” The script says, WAIT NOW FOR REACTION.

“Oh,” she says. “I guess that’s good then. Was your name Martha Tif-ney?”

“Martha Tiffany Reynolds,” I say.

I wave at Jeremy near the window grid flipping me off like he always does. I stick out my tongue and he gives me his rat face. We spent last weekend in bed and he’s probably bored already.

Janet tells me in a hushed tone that I sound like a super, no B.S. gal.

“You do too, sweetheart—we love you here,” I say.

She says she’s a waitress. Her husband died on the way home from work one-and-a-half years ago, crushed by a semi. She has a toddler named Trevor. He’s a handful, and needs a good preschool. She hopes to be able to afford one soon.

Sweat is forming under my breasts and pits even though the airconditioning is blasting. I say the last line of the script a bit early, feeling my full bladder, pressing it with my hand to make it worse. “You. Deserve. Success.”

“She gives me her credit card number, saying, Shit yes! to the Supreme Success Package (the most expensive).

“I bet you’re pretty and thin, Martha Tifney!” she says before she hangs up.

----

After work I bring Janet’s order sheet home under my shirt. I read off each name as I tear the sheets into bits: Kelly, Nita, Jen, Marla, Iris, Nancy, and Janet. They will be mystified when there’s no charge on their statements and they receive nothing else.

I take off my clothes and stand naked in front of the bathroom mirror. Look at myself from different angles. The way a casting director would.

**PASSAGE 2:** “Bone Wars” Laura Citino.

Citino, L. (2017). “Bone Wars.” *Ghost Parachute*.

This kid is trying to tell me that Brontosaurus didn’t really exist. He wears a black T-shirt halfway to his knees, skinny like they all are. These kids get bused in from tri-county or the group homes because a day at Suffolk City Dino Park is a cheap way to occupy restless eyes and destructive little hands. Most of them get right to vandalizing. Some of the park people hate these kids, but I like them. They’re easy, really, so easy to help.

“I don’t know, buddy,” I say. “We got one right here.” I gesture like I’ve been taught to the animatronic sauropod creaking its head back and forth above us. From where I stand, I can see the bolts in its neck, green and yellow paint rusting. I’m supposed to be doing rounds to make sure the tri-county kids aren’t clogging toilets or scrawling dick on the flat boulders along the walking path. This job is low stakes. No flashing lights or emergency tracheotomies, no rollovers in cornfields. A positive step forward, was what I wrote on my application.

“Brontosaurus was a mistake,” the kid says. His nose runs and he keeps slinging the back of his hand across his face. Snot doesn’t freak me out. In EMT school I always had the strongest stomach. Blood, vomit, the strange clarity of viscera, no problem. “They thought they discovered another species but they didn’t,” he says. “It’s just an Apatosaurus with a funny head.”

My first day, José quizzed me on dinosaur knowledge in the break room while he sorted through OSHA videos. Velociraptor means Swift Seizer. Dimetrodon means Two-Sized Tooth.

José can’t wait for me to see the Dino Park in the fall, how the flaming leaves enliven the machinery. I like his faith in my future here.

I say, “Did you know that Brontosaurus means Thunder Lizard?”

He looks at me with suspicion, caught off guard by how much of a dweeb I am. These kids are hard to impress. Even the dumbest ones just know things, but I’m trying to get into the habit of automatic goodwill toward others, the belief that someone, somewhere, is having a worse day than I am. The potted ferns rustle and I pull José from his hiding place, pluck the joint from his lips. José loves kids. He places his long fingers on the back of my neck. “Oh, Carlotta,” he says. “These are the most majestic creatures to ever walk the planet. We have to help him.

This could save his life.”

I flinch. “He’s right here.”

The kid says hi. He says his name is Carter.

José pinches his eyebrows together, rearranges his face into one of deep concern. “Carter, are you telling my beautiful friend here that Mr. Brontosaurus isn’t actually real?” He winks at me. The kids always ask if the adults here are dating or married, as if that’d be the most natural thing.

Carter stomps his foot. “Brontosaurus just isn’t the right name of that thing. Somebody switched the skulls.” He has the face of someone who has never been told, I believe you. José says that the admittance of failure is not my strong suit.

Carter says, “Well, she stinks.”

Kid, I can admit failure anytime. The ease with which I admit failure would blow your fucking mind. The history of prehistoric discovery in America is fraught with lies and deception, one-of-a-kind fossils destroyed out of hand simply for the pleasure of undercutting another man’s work. How’s that? And I knew the guy was dead. The car was crushed in like a pop can, this guy hung by his seat belt, bleeding out at a speed that seemed impossible to me. His femur was white and clean. But I thought maybe he needed to breathe. That’s why I jammed my little knife into his throat. That’s why I did it, even though I didn’t need to, because I was surprised by all that white.

José grabs my hand and holds it tight.

Here’s another. When I was a kid, I visited a park just like this one. I screamed the whole drive home because I thought the dinosaurs were alive. Carter says that we create and destroy in equal measure and José says, Babe, he’s got a point, but I knew all their names. I knew they would follow me home. We found the bones. Even when we are wrong, the bones will still be real.

**PASSAGE 3:** “The Kiss” Pamela Painter.

Painter, P. (1999). *The Long and Short of It*. Carnegie-Mellon University Press.

No one can guess so she finally tells us.

Actually Mona doesn’t tell, she sticks out her tongue at us and there it is—a gold ball the size of a small pea, sitting in the creased rose lap of her glistening tongue.

We all lean forward from our pillows on the floor, seven of us, the wrung-out remnants of a grad party in the low-candle stage. Inge asks Mona where she got it done (Cambridge as an undergrad), Raphi our host asks her why (she likes something in her mouth), my boyfriend wants to know what it tastes like (no taste). We’re still peering into her mouth so she

lifts her tongue slowly, the rosy tip pointing up toward her nose. There on the silky downward slope is another gold ball. “A bah—bell,” she says, her tongue still showing off. And it is. A tiny gold barbell piercing her tongue.

We settle back into our pillows and she closes her mouth.

My boyfriend, a chef at Valentino’s, is probably wondering which taste buds sit in the middle of the tongue and if they are affected and how. Inge, the etymologist, is mouthing the word tongue no doubt marveling at how the tongue loves to say that word. I can tell we’re all wondering something. Our tongues feel heavy in our mouths, empty except for the privileged gold fillings and ivory bondings of the middle class.

It occurs to me that Mona’s not able to enjoy her barbell. Enjoy the way the tip of my tongue visits a rough molar, soothes a canker sore moistly healing on its own, or wetly licks the hairy friction of chapped lips.

I say I wonder what it’s like to kiss her?

Everyone shifts and nods as if they were wondering the exact same thing. We turn to the man Mona came with, who shrugs and says he doesn’t know. They just met three hours ago at Huddle’s Pub.

Well, who’s going to kiss her? Inge says.

We all look first to the man she came with and then at the other eligible male. No one counts mine, which disappoints him and he lets it show. The man she came with weighs thirteenth century Inca bones after reducing them to ashes in an autoclave the size of a toaster. Raphi, our host, is a religion major—the Hellenistic culture -- who thinks the world is fast approaching a non-religious end. “I’ll kiss her,” he offers, then defers to the man she came with.

“Wait a minute. Maybe not,” Jorie says, holding up her hand. She and I are in gender studies. “Don’t you think we should run it by Mona first?” We all turn belatedly to Mona.

“Oh,” she says, “It’s all right with me.” The gold ball doesn’t show when she talks. I wonder if it makes a dent in the roof of her mouth.

The man she came with says, “I’ll kiss her.” Neither man is looking at Mona.

“You choose,” Inge says to Mona.

Mona shrugs and points to Raphi. “You offered first.” He grins.

In perfect sync, they both stand up.

“No, do it here,” we all say, “here in front of us.” But we needn’t have worried; they had no intention of leaving.

Mona and Raphi face each other above us. They are the same height. Mona’s hands rest on the hips of her black jeans, her elbows jut out, claiming space to equal Raphi’s greater weight. He has his hands deep in his pockets. We are all aware of his hands in his pockets.

“They stand inches apart—two inches apart. She tilts toward him first, just her shoulders and head, and then he catches her tilt, catches her mouth with his mouth. They kiss. They kiss tenderly and well for two people who have just met. Their heads glide with their mouths and their shoulders move ever so slightly. I imagine his tongue filling her mouth, sliding toward the ball, searching, pressing, perhaps turning it, rolling it; her tongue letting him. I imagine their hands aching to touch the other person but refraining as if to abide by some set of rules. No one looks away.

Minutes, but probably seconds, later they stop. “It’s pretty far back,” Raphi says, and we all swallow with him.

Mona turns and sticks out her tongue to show us she thinks not, and we see it’s not so far, really. Perhaps an inch and a half.

She turns back to Raphi and they kiss again and we all watch them kiss, even better the second time: harder, deeper, her tongue and his tongue, her generously letting him, that slight tilt, their scrupulous hands.

They pull away. We have all been holding our breath.

Well?

They settle themselves cross-legged and facing each other. I imagine another night such as this for them, moving away from the kiss toward the questions and answers of getting to know someone, and that moment when they invite their hands to join their kiss.

We listen as Raphi describes to Mona the amazingly hard muscle of her tongue, the cool surprise of the tiny gold ball, the flick past the ball underneath. They tilt toward each other.

Raphi’s hands talk.

Mona is smiling that smile. She’s got what she wanted. The man she came with leaves first.

My boyfriend leaves with me, but we go home separately. We all go home with something missing on our tongue.

### Appendix C

Narrative Engagement Scale (Busselle & Bilandzic, 2009)<sup>2</sup>

#### Narrative Understanding Subscale

At points, I had a hard time making sense of what was going on in the story. (-)

My understanding of the characters is unclear. (-)

I had a hard time recognizing the thread of the story. (-)

---

<sup>2</sup> Items are ranked on a 7-point Likert scale: (7) strongly agree; (6) agree; (5) somewhat agree; (4) neutral; (3) somewhat disagree; (2) disagree; (1) strongly disagree. The scale has also been modified slightly to refer to a ‘story’ rather than a ‘program.’ (-) Indicates reverse scoring.

**Attentional Focus Subscale**

I found my mind wandering while reading the story. (-)

While reading the story I found myself thinking about other things. (-)

I had a hard time keeping my mind on the story. (-)

**Narrative Presence Subscale**

During the story, my body was in the room, but my mind was inside the world created by the story.

The program created a new world, and then that world suddenly disappeared when the story ended.

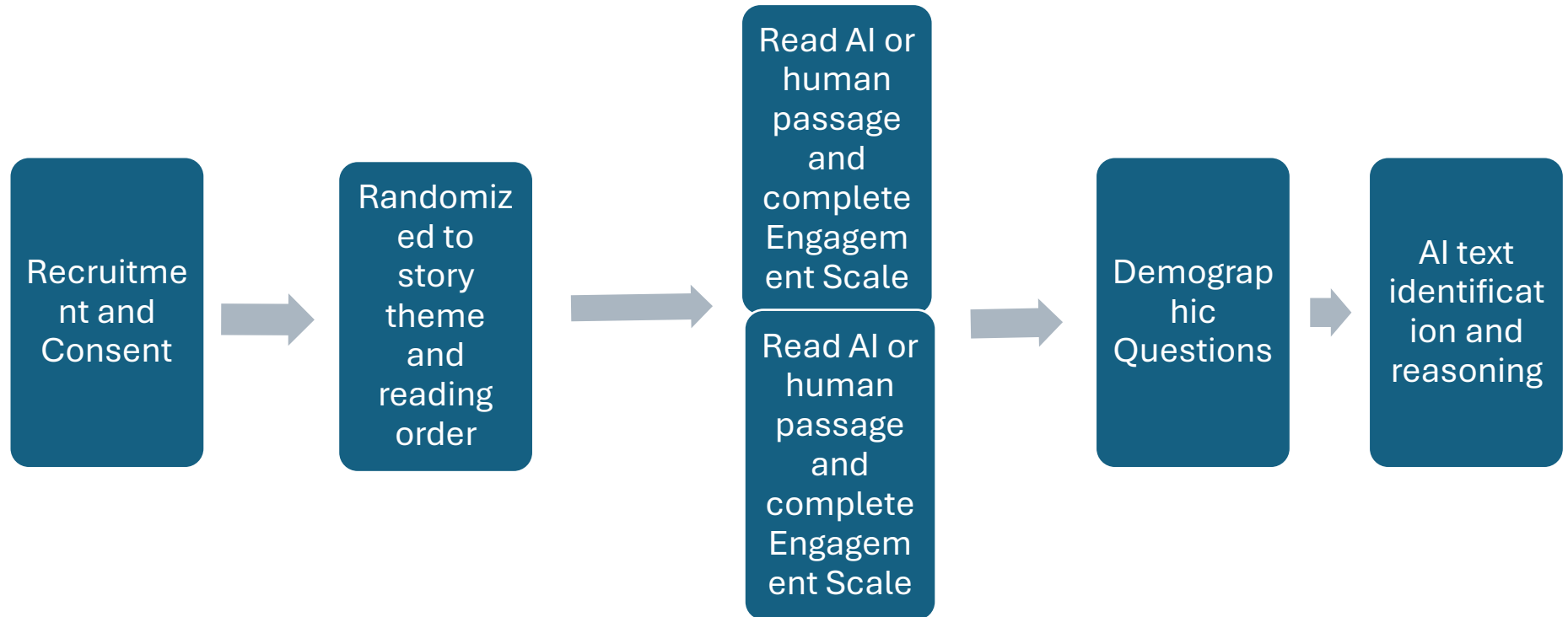
At times during the story, the story world was closer to me than the real world.

**Emotional Engagement Subscale**

The story affected me emotionally.

Through the story, when a main character succeeded, I felt happy, and when they suffered in some way, I felt sad.

I felt sorry for some of the characters in the story.





**Appendix D** Demographics Questionnaire

1. How old are you?
2. What level of education have you reached?
  - a. High school or GED
  - b. Some college
  - c. Bachelor's degree
    - i. Ask major
  - d. Master's degree
    - i. Ask title
  - e. Doctorate
    - i. Ask title
3. How often have you voluntarily engaged with each type of media in the past six months (ex. one sitting of reading)?
  - i. Academic (i.e. journal articles, textbooks, theoretical papers)
  - ii. Long Fiction (i.e. novels)
  - iii. Short fiction (i.e. short stories, flash fiction)
  - iv. Nonfiction (i.e. biographies, scientific, history, lifestyle)
  - v. Poetry
  - vi. Manga, comics, or magazines
    1. (1) Never or almost never (2) Less than 10 times (3) Every month or so (4) Weekly to every other week (5) Every day
4. Is English your first language?
  - a. Yes
  - b. No
    - i. If no: What is your first language? 1. [enter text]
5. How confident are you in your reading abilities for texts such as novels, short stories, and
  - a. (1) Not confident (2) Fairly Confident (3) Neutral (4) Very Confident (5) Extremely Confident
6. What is your familiarity with the online program ChatGPT?
  - a. (1) Never heard of it (2) Heard of it, do not know what it is (3) Have heard of it and know what it is (4) Have heard of it and used it a few times (5) Have heard of it and use it often

Origin Detection Questions.<sup>3</sup>

1. If you had to choose, which of the passages that you read do you think is more likely to have been written by an artificial or chatbot program (i.e. ChatGPT, Google's Bard, TextFX)?
  - a. Passage 1
  - b. Passage 2
2. How confident are you in your response to the previous question?
  - a. (1) I guessed (2) Not confident (3) Fairly Confident (4) Very Confident (5) Almost Positive
3. What were your reasons for identifying the passage that you did as the one written by an AI?

---

<sup>3</sup> Included within the Demographics questionnaire at the end of the testing.

a. Open text response

---