***AI In Literature:* Volume 7, Issue 1**

**Journal Article: Friend Or Foe? Exploring the representations of AI in Mary Shelley’s *Frankenstein*.**

**Arianna Parris**

**Nottingham Trent University**

**English (BA) Hons**

**N1079102@my.ntu.ac.uk**

Throughout literature, science and AI have been portrayed in various ways - both good and bad, however Mary Shelley gives authors the founding thoughts for these portrayals by exploring the nature and morality of both to a whole new level in her groundbreaking, gothic novel, *Frankenstein*. The 1818 novel presents us with one of the earliest forms of artificial intelligence seen in literature - Victor Frankenstein’s monster, a sentient, artificially intelligent being - and Shelley subtly introduces questions of morality as well as the possible consequences of unrestrained science through the creature’s vicissitudes. Since the novel’s publication, Shelley’s thoughts and concerns for the future of AI have both been mirrored and rejected, with many feeling that this evolution will impact the world positively. This article will be exploring how Mary Shelley’s *Frankenstein* constructs the conflicting ethical factors regarding the creation of AI and how these ideas have been responded to, as well as the relevance of the gothic novel in today’s society. In addition, it aims to examine the author’s debate on the efficiency and detriment of AI through its theme of a foreshadowed danger in light of the swift movement of science during the Enlightenment Era, which manages to both dramatically challenge and concur with today’s portrayals of AI by authors and scholars, specifically Nick Bostrom, Tom Chivers and Nir Eisikovits.

Shelley’s novel quickly introduces its protagonist, Victor Frankenstein, whose character successfully depicts the idea and consequences of unrestrained science through his ardent desire to break the bounds of nature. Frankenstein earns the title as the ‘Modern Prometheus,’ as he seeks to play the role of God by manipulating life and death in hopes of becoming revered and remembered through his accomplishments, resulting in his final creation - a lifeform made from body parts of the deceased, and a variety of chemicals. His character speaks of a dissatisfaction “with the results promised by the modern professors of natural science,”[[1]](#footnote-0) and this desire to transcend natural limits proves to be a perfect embodiment of both today’s current society, and the Age of Enlightenment (approximately the time in which the novel was written). During the Enlightenment Era, some of the oldest social, religious and political beliefs began to be challenged, with discoveries that are now deemed conventional in the western world being the most pioneering breakthroughs of their time. Throughout this progressive era, the development of science was also at its peak thanks to preeminent scientists such as Isaac Newton who laid the fundamental foundations of physics through his laws of gravitation and motion and his development of calculus. Others such as Andreas Vesalius, who corrected traditional teachings and dissected corpses for his studies, allude closely to Victor Frankenstein’s contentious curiosity which, like many philosophers and scientists of the Enlightenment Era, raised ethical concerns. Shelley uses the character’s desires of crossing forbidden boundaries to convey a warning about these rapid scientific advancements, implying that the participation of nuanced experiments could eventually backfire or unravel things that should’ve stayed hidden. The novel ingeniously presents the potential consequences through a dramatised narrative as Frankenstein contemplates “whether [he] should attempt the creation of a being like [himself] or one of similar organisation,”[[2]](#footnote-1) and his lack of consideration for any ethical factors (or any factors in general) results in intense repercussions as the novel unfolds.

Philosopher, Nick Bostrom, centres his book, *Superintelligence: Paths, Dangers, Strategies,* on the idea of ‘The Control Problem,’ a conceptual framework referring to the potential malfunctioning of AI. His eighth chapter, *‘Is The Default Outcome Doom?*’ illustrates AI almost as a deceitful, self-aware being with bad intentions as he speaks of an “unfriendly AI with sufficient intelligence [realising] that its unfriendly goals will be best realised if it behaves in a friendly manner initially…”[[3]](#footnote-2) Throughout the chapter, Bostrom presents the idea of the increasing capabilities of AI eventually backfiring, referencing a ‘sandbox concept’ in which an AI seemingly appears to perform correctly in controlled environments whilst having intentions that defy its original purpose. This concept harmonises well with his ideas of an artificial intelligence purposefully malfunctioning in order to be reprogrammed and refined, resulting in its intelligence substantially surpassing the intellect of humans. Bostrom’s anxieties regarding the increasing power of AI correlate closely with Shelley’s central theme in *Frankenstein* - unintentional consequences. After being relinquished by its creator and ostracised by all those around him, the monster demonstrates the way in which forms of AI could eventually take control, with the creature educating himself through books, “while I improved in speech, I also learned the science of letters…through this work I obtained a cursory knowledge of history,”[[4]](#footnote-3) adapting independently and creating his own agenda - aiming to pursue revenge. Victor’s initial abandonment of his invention inadvertently leads to the death of his loved ones, as the monster abruptly develops cognitive human emotions, and through this, Shelley reveals the negative sides of the development of AI, depicting an exaggerated example of a malfunction. Overall, Shelley’s narrative manages to display Bostrom’s predictions of a defective intelligence with a premeditated plan beginning with “occasional mishaps,”[[5]](#footnote-4) leading to the development of, “systems that are smarter and have more common sense,”[[6]](#footnote-5) and resulting in an intelligence far superior to the human brain.

Tom Chivers, an award-winning science writer, appears to downplay the potential hazards linked to artificial intelligence, which is made very apparent through the title of his book: *The AI Does Not Hate You*. Chivers highlights the anthropomorphism that many attach to the idea of AI which he describes causes unnecessary alarm as he believes the two do not correlate. Throughout the book, Chivers points out the main difference between humans and AI - the fact that humans are living, functioning and conscious organisms with complex emotions and thoughts, whereas AI merely mimics human behaviour and is neither conscious nor able to feel emotions. Through this, Shelley’s portrayal of Frankenstein’s monster as a form of AI becomes increasingly unrealistic, as her portrayal of a sentient character appears less relevant when applied to an insentient entity. Similarly, Nir Eisikovits discusses the unlikely future of a man-made intelligence that is capable of feeling the way humans do in his article, *AI Isn’t Close to Becoming Sentient,* stating that “the real danger lies in how easily we’re prone to anthropomorphize it.”[[7]](#footnote-6) The philosopher acknowledges important figures such as Elon Musk and Stephen Hawking who publicly exaggerated the potential future ahead, describing it as one of the greatest threats to humanity, as well as the myriad of science fiction shows and books that frequently instil fear, and when both opinions are applied to *Frankenstein*, the text becomes increasingly less important in regards to AI as a whole. As the monster narrates his experiences after his isolation, Shelley makes the character’s natural sentience exceedingly apparent as he mentions his “feelings were those of rage and revenge,”[[8]](#footnote-7) “a kind of insanity in [his] spirits that burst all bounds of reason and reflection.”[[9]](#footnote-8) Despite not entirely eradicating the true meaning and warnings introduced by Shelley through the story, the clear difference between today’s insentient AI and an anthropomorphic creature, created using human body parts, makes her novel feel almost implausible as it clearly portrays the habit Eisikovits points out, the human projection of our own traits that do not apply to other organisms or objects.

Overall, Mary Shelley’s *Frankenstein* certainly proves itself to be ahead of its time through its references to the potential harm that could lie ahead if restrictions fail to be implemented to the progression of science. Despite being an intensely dramatised depiction of what the future of AI may hold, the story most definitely captures the ethical values and dangers in a metaphorical sense. The debate of the novel’s importance remains controversial due to its exaggerations and sentient representations, however through its themes of lack of control, unintentional consequences and creation as a whole, the ‘Modern Prometheus’ most definitely challenges modern writers like Tom Chivers who seem to downplay just how complicated the invention of an intelligence like ours could become. Like Shelley, Nick Bostrom leaves room for his readers to ponder the never-ending ethical conflict around the topic through his reflections of ‘The Control Problem,’ revealing the way in which modern writers not only dismiss the warnings, but also build on them. As a result of this, the timeless classic demonstrates a form of relevance through not just its controversial storyline, but the way in which it can be interpreted and responded to.

Bibliography:

Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies,* 1st edn(Oxford: University Press, 2014)

Chivers, Tom, *The AI Does Not Hate You: Superintelligence, Rationality And The Race To Save The World*, (Bloomsbury: 2019)

Eisikovits, Nir, *AI Isn’t Close to Becoming Sentient - The Real Danger Lies in How Easily We’re Prone to Anthropomorphize It, The Conversation, (2023) <*[*https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525*](https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525)*> [accessed 8 January 2024]*

Needle, Flori, *Is AI Sentient? Could It Ever Be? Experts Weigh In*, (2023) <<https://blog.hubspot.com/marketing/sentient-ai#:~:text=Sentient%20AI%20Is%20Part%20of%20the%20Not%2DSo%2DNear%20Future.&text=The%20next%20type%20is%20theory,interactions%20based%20on%20this%20understanding>.>[accessed 28 December 2023]

Shelley, Mary, *Frankenstein*, ed. By Kathleen Scherf, 3rd edn (Hertfordshire: Wordsworth Editions Ltd, 1993)

*The History Of Revolutionary Thought*, <<https://evolution.berkeley.edu/the-history-of-evolutionary-thought/pre-1800/comparative-anatomy-andreas-vesalius/>>[accessed: 2 December 2023]

1. Mary Shelley, *Frankenstein*, ed. By Kathleen Scherf, 3rd edn (Hertfordshire: Wordsworth Editions Ltd, 1993), p.37. [↑](#footnote-ref-0)
2. Shelley, p.42. [↑](#footnote-ref-1)
3. Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies,* 1st edn(Oxford: University Press, 2014)p.117 [↑](#footnote-ref-2)
4. Shelley, p.92. [↑](#footnote-ref-3)
5. Bostrom, p.117. [↑](#footnote-ref-4)
6. Bostrom, p.117. [↑](#footnote-ref-5)
7. Nir Eisikovits, ‘AI Isn’t Close to Becoming Sentient - The Real Danger Lies in How Easily We’re Prone to Anthropomorphize It*’, The Conversation, (2023) <*[*https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525*](https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525)*> [accessed 8 January 2024]* [↑](#footnote-ref-6)
8. Shelley, p.104. [↑](#footnote-ref-7)
9. Shelley, p.106. [↑](#footnote-ref-8)