Design of a GPT-4-Based Initial Moral Distress Consultation Assistant

Analysis of the Failure Behind the Technological System of the Therac-25

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Steven Myung

October 27, 2022

Technical Team Members: None

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Benjamin Laugelli, Department of Engineering and Society
Kevin Sullivan, Department of Computer Science

Introduction

Technology has had many positive impacts on modern healthcare, including the use of electronic records, improved patient care, and improved communication (Joseph, 2020). Particularly, improved quality of patient care is a big factor that healthcare organizations consider. In order to improve the quality of healthcare for patients, it is important to focus on improving the criteria behind principles, such as staff support or improved patient care, upon which solutions in healthcare should be built (Thimbleby, 2013).

One key criterion of staff support that needs to be addressed is moral distress, a phenomenon when nurses cannot carry out what they believe to be ethically appropriate actions because of institutional constraints (Summers, 1985). Today, critical care nurses report moral distress using a time-consuming paper-based process. While they wait for a consultation to be scheduled, moral distress can cause a healthcare provider to burn out and provide lower-quality patient care. To address the lack of in-the-moment resources for healthcare professionals dealing with moral distress, I propose an artificial intelligence (AI) based assistant that can provide an initial consultation to identify causes of moral distress, barriers to taking action, and strategies to improve the situation at hand (Epstein et al., 2021).

An organization is empowered to increase the quality of patient care through the influence and role of the technologies around it. As technologies grow, they gain momentum to influence the organizations that the technologies were once influenced by. To further investigate how a technological system can gain momentum to influence an organization, I will apply the STS framework of Technological Momentum to investigate the technological system around the Therac-25 and how it came to gain momentum to shape the very group of people that first influenced it, leading to 6 incidents of significant radiation overdose, resulting in death from

1985 to 1987 (Leveson & Turner, 1993).

Attending to both the technical and social aspects of increasing the quality of patient care provides a more comprehensive, and thus more effective, approach to addressing the quality of patient care because one can determine if the technological system's influence empowers the organization to increase the quality of patient care. Therefore, because the goal of increasing the quality of patient care is sociotechnical in nature, it requires attending to both technical and social dimensions. In what follows, I elaborate on two related research proposals: first, a technical project that describes an AI-based assistant that provides an initial consultation for moral distress, and second, an STS project that examines the momentum gained by the technological system around the Therac-25 that led to 6 deaths. As I work on developing the proposed AI-based assistant, I will apply insights from the study of Therac-25's failure to develop a technological system that better accounts for its role and influence on society, so that my design can empower an organization to increase the quality of patient care.

Technical Project Proposal

There are two parts to moral distress - one is the initial distress that occurs in the moment, and the other is the reactive distress that remains after the situation has passed (Epstein & Hamric, 2009). Studies have shown that the initial distress is where action can, and should, be taken (Epstein & Hamric, 2009). The initial distress can be emotionally overwhelming and nurses lack effective in-the-moment resources while a consultation is scheduled, outside of resources consisting of articles on how to manage moral distress. These articles fall short because they do not provide any concrete guidance on strategies to follow.

Other AI-based chat assistants are marketed as the final step in the process of accomplishing a certain goal. These chat assistants often use a Large Language Model (LLM), a deep learning algorithm that performs language processing tasks. These solutions are inadequate because LLMs all have some randomness when performing tasks, meaning there is no way to control or guarantee any output. My solution is practical because it will be used as a tool in the consultation process, not as the final step. Both consultation and healthcare workers will know how to proofread and improve upon the output from my assistant. This allows both the consultation process to speed up and provides concrete guidance on strategies to follow at the moment. This also means that healthcare professionals know not to rely on the assistant fully, but to use the outputs as a foundation to build and improve on, avoiding ethical dilemmas that may arise in the future.

The main engineering method that will be used to develop and refine my design will be prompt engineering. Specifically, the use and combination of prompt patterns, implementing system and ground rules, few-shot prompting, and chain of thought prompting. Prompt patterns are known to prompt solutions to a common problem (White et al., 2023). By utilizing prompt patterns to address broken-down problems of moral distress, and combining the prompt patterns together, we can effectively and efficiently address fundamental pieces of the overall problem. My design will use the GPT-4 API, the set of rules for software to communicate and share data with OpenAI's LLM, GPT. This API allows system role instructions to be given to the LLM, separate from any user input (Suárez, 2023). Setting up a system and ground rules define how the assistant will behave, what rules it should follow, and what to and not to do (White et al., 2023). Few-shot prompting will give the assistant examples of input, output, and optionally intermediate steps so that when prompted with the input prefix, the LLM will use pattern recognition from the examples given to fill the rest of the template (Wei et al., 2022). An example is given in the figure below.

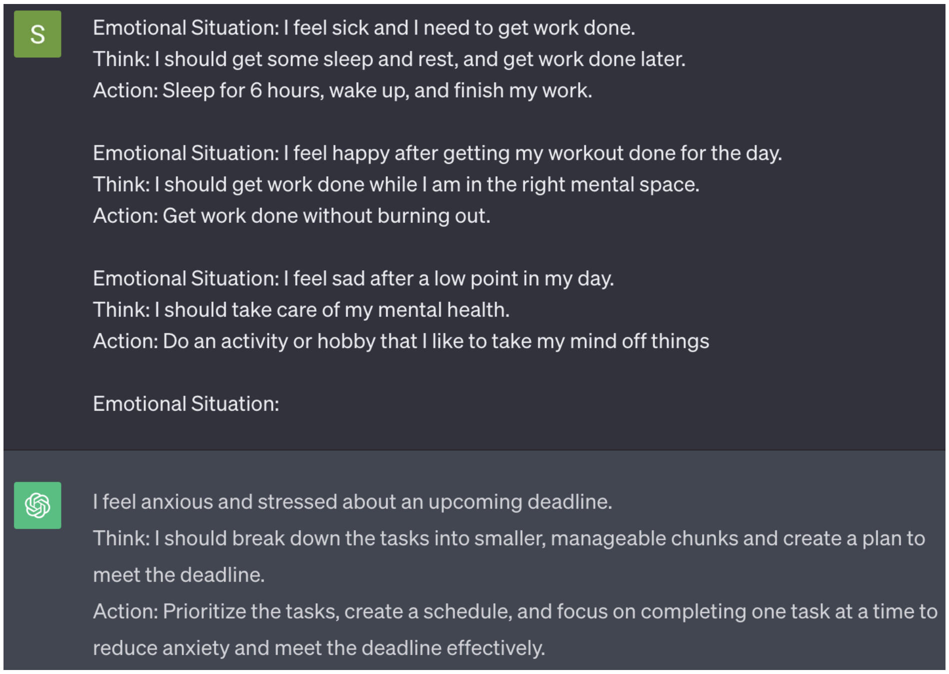

Figure 1 - Few-Shot Prompting Example

Lastly, chain of thought prompting is getting an LLM to explain the reasoning behind an output because if the reasoning is correct, then it is more likely that the right answer will follow the correct reasoning (White, n.d.). An example is shown in the figures below. By collecting data and analyzing the process in which consultation services address moral distress from research papers, mock data sourced from real data, and interviews with professors and graduate students directly involved in the Moral Distress Consultation Collaborative the UVA healthcare system is part of, I can use the methods mentioned above to develop my design to emulate this process as accurately as possible. That way, stakeholders can see the value in my design in comparison to the current systems in place. Since this technical project is a part of a current Institutional Review Board (IRB) approved research project due to human participants, I will also demonstrate the viability of this project by following IRB guidelines.

Question: if I have 2 blue rocks and 4 red leaves in a black bag, and I randomly select one item at a time, what are the chances that I will get 2 blue rocks?
Reasoning: The chance of selecting a blue rock at first is 2/6, since there are 2 blue rocks and 6 total items. The chance of selecting a blue rock after selecting a blue rock at first is 1/5, since we already selected a blue rock so that is one less blue rock and one less item. Multiplying those values gets us 2/30, or 1/15.
Answer: 1/15.

Question: If I have a black bag with 4 red rocks, and I add in 2 blue rocks and 2 green rocks, what are the changes of selected a red rock?
Reasoning: Since we will have 8 total items in the bag, and we don't add any red rocks, we will have 4 red rocks and 8 items which is a 4/8 probability, or 1/2 or 50%.
Answer.: 1/2

Question: If I have a blue bag of 200 red rocks, 500 blue leaves, and 300 green sticks, and we add double the blue leaves, triple the red rocks, and quadruple the green sticks, what are the chances of selected a red rock?
Reasoning: <REASONING>
Answer: <ANSWER>

Answer the question by filling in the <REASONING> with your reasoning and <ANSWER> with your answer

Figure 2 - Chain of Thought Prompt

Question: If I have a blue bag of 200 red rocks, 500 blue leaves, and 300 green sticks, and we add double the blue leaves, triple the red rocks, and quadruple the green sticks, what are the chances of selecting a red rock?

Reasoning: First, let's calculate the total number of each item after adding the specified amounts:

- Red rocks: 200 (original) + 3 * 200 (added) = 800 red rocks
- Blue leaves: 500 (original) + 2 * 500 (added) = 1500 blue leaves
- Green sticks: 300 (original) + 4 * 300 (added) = 1500 green sticks

Now, let's calculate the total number of items in the bag:
Total items = 800 (red rocks) + 1500 (blue leaves) + 1500 (green sticks) = 3800 items

The probability of selecting a red rock is the ratio of the number of red rocks to the total number of items in the bag:

Probability of selecting a red rock = (Number of red rocks) / (Total items)
Probability of selecting a red rock = 800 / 3800

Now, let's simplify the fraction:
Probability of selecting a red rock = 2 / 19

Answer: The chances of selecting a red rock from the bag after adding the specified amounts of items are 2/19.

Figure 3 - Chain of Thought Output

STS Project Proposal

External beam radiation therapy aims high-energy beams into one's body using a machine called a linear accelerator. The Therac-25 was a double-pass linear accelerator released in 1983 by Atomic Energy of Canada Limited (AECL), with its double-pass design allowing for a more powerful accelerator taking up less space at a lower cost (Huff, 2003). It largely relied on software for both computer and safety control (Huff, 2003). The Therac-25's role in the healthcare industry was to deliver precise, high-energy fractionated doses over several treatment sessions utilizing software instead of manually operated hardware (Connell & Hellman, 2009). However, due to a software bug where a system's behavior is dependent on the sequence of uncontrollable events, the radiation therapy machine gave patients doses more than a hundred times greater than what they should have received (Leveson & Turner, 1993). This has led many writers to conclude that it appears the radiation therapy machine was designed to control the hardware and safety of delivering x-rays with software, appealing to the software bug, lack of communication between AECL, hospitals, and the government, and the overconfidence of software engineers in both the safety analysis and response to the overdose incidents as the social factors in this case.

However, x-rays were first discovered by Wilhelm Conrad Roentgen, and radiation machines were designed by the medical community for physicians to diagnose and primarily cure skin cancers with single large exposures because of the low penetration of tissues from the low energy x-rays (Gianfaldoni et al, 2017). Other writers' perspectives do not take into account the social factors of the socio-technical system evolving to influence the practices, values, and power dynamics of the healthcare industry that created it. If we continue to think that radiation therapy machines always played a role of having software control all hardware and safety controls, then we will not understand how its role and influence evolved over time, and failures similar to the Therac-25 may happen again regardless of the social factors at the time of the potential failure.

I argue that although radiation therapy machines were designed to diagnose and cure skin cancers by emitting low-energy X-rays in the late 19th century, over time they gained momentum due to the innovations of higher-energy X-rays, better precision, and computer systems. In turn, the system gained increased complexity, scale, bureaucracy, and skills and knowledge to maintain, shaping the healthcare industry by changing the practices and power relations the industry had with the machines. Technological Momentum, developed by Thomas P. Hughes, says that as technology (socio-technical system) gains momentum, the influence of society over technology decreases, while technology's influence over society increases (Bijker et al., 2012). Early in the life of a technology, society has more power to shape the technology's design, purpose, meaning, and role, setting society as cause and technology as effect. After technology has gained momentum, technology has more power to influence society's practices, values, power relations, etc, setting technology as the cause and society as the effect. Applying these concepts, I will analyze evidence from reports from someone involved in the case as an expert witness, lawsuits, depositions, and government records, providing information about how radiation therapy machines shaped the healthcare industry by changing the practices and power relations the industry had with the machines.

Conclusion

The technical design that improves on existing designs and their value to stakeholders will be a GPT-4-based initial moral distress consultation assistant that draws on one petabyte worth of data to come up with the best outputs to address emotionally overwhelming experience by healthcare workers and create a moral distress consult summary that can be used by nurses for reference or by consultants to speed up the consultation process already in places in hospitals everywhere. The STS research will aim to determine why the Therac-25 overdosed 6 people despite its state-of-the-art technology utilizing Technological Momentum to determine how, as the technology evolved, its increased role and influence over society shaped the practices and power dynamics society had with itself, leading to the Therac-25 incident. By applying the insights of technology evolving, and influencing an organization, I will be able to design my technical project to influence the healthcare industry in a positive way so workers can rely more on an institution's staff support, allowing for the workers to be in a better state of mind to provide better quality healthcare to patients. The technical project contributes to addressing the challenge of improving the quality of patient care by providing an effective in-the-moment option to healthcare workers so that they can continue providing the best quality patient care possible. The STS project also addresses the challenge by identifying how society's social factors are shaped as technology gains momentum, allowing society to be conscious of, and therefore act on, how technology is shaping its practices, values, and power relations.

References

Bijker, W. E., Thomas Parke Hughes, & Pinch, T. J. (2012). *The social construction of technological systems : New directions in the sociology and history of technology*. The Mit Press.

Connell, P. P., & Hellman, S. (2009). Advances in radiotherapy and implications for the next century: A historical perspective. *Cancer Research*, *69*(2), 383–392. https://doi.org/10.1158/0008-5472.can-07-6871

Epstein, E. G., & Hamric, A. B. (2009). Moral distress, moral residue, and the crescendo effect. *The Journal of Clinical Ethics*, *20*(4), 330–342. https://doi.org/10.1086/jce200920406

Epstein, E. G., Shah, R., & Marshall, M. F. (2021). Effect of a moral distress consultation service on moral distress, empowerment, and a healthy work environment. *HEC Forum*. https://doi.org/10.1007/s10730-021-09449-5

Gianfaldoni, S., Gianfaldoni, R., Wollina, U., Lotti, J., Tchernev, G., & Lotti, T. (2017). An overview on radiotherapy: From its history to its current applications in dermatology. *Open Access Macedonian Journal of Medical Sciences*, *5*(4), 521. https://doi.org/10.3889/oamjms.2017.122

Huff, C. (2003). *A history of the introduction and shut down of Therac-25 | Online Ethics*. Onlineethics.org. https://onlineethics.org/cases/therac-25/history-introduction-and-shut-down-therac-25

Joseph, T. (2020, October 6). *7 major impacts of technology in healthcare*. Fingent Technology. https://www.fingent.com/blog/7-major-impacts-of-technology-in-healthcare/

Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer*, *26*(7), 18–41. https://doi.org/10.1109/mc.1993.274940

Summers, J. W. (1985). [Review of NURSING PRACTICE: THE ETHICAL ISSUES, by A. Jameton]. *Business & Professional Ethics Journal, 4*(1), 83–87. http://www.jstor.org/stable/27799853

Suárez, D. (2023, April 25). *How to write "system" instructions for OpenAI's GPT-4 Chat API*. Rootstrap https://www.rootstrap.com/blog/how-to-write-system-instructions-for-openais-gpt-4-chat-api

Thimbleby, H. (2013). Technology and the future of healthcare. *Journal of Public Health Research*, *2*(3), 28. NCBI. https://doi.org/10.4081/jphr.2013.e28

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*, 24824-24837.

White, J. (n.d.). *Chain of Thought Prompting - Few-Shot Examples*. Coursera. Retrieved October 27, 2023, from https://www.coursera.org/learn/prompt-engineering/lecture/bK1RY/chain-of-thought-prompting

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382.