Ethical AI and the Future of Healthcare: Combining Academic Theory and Industry Practice to Ensure Patient-Centered Care

Dillon Plummer
Ethical AI and the Future of Healthcare: Combining Academic Theory and Industry Practice to Ensure Patient-Centered Care

Dillon Plummer*
PhD Candidate, Capitol Technology University

Article history
Submitted 23.11.2023 Revised Version Received 22.12.2023 Accepted 04.01.2024

Abstract

Purpose: This paper seeks to define a risk taxonomy, establish meaningful controls, and create a prospective harms model for AI risks in healthcare. Currently, there is no known comprehensive definition of AI risks, as applied to industry and society.

Materials and Methods: The temptation for current research, both in academia and industry, is to apply exclusively-tech-based solutions to these complex problems; however, this view is myopic, and can be remedied by establishing effective controls informed by a holistic approach to managing AI risk. Sociotechnical Systems Theory (STS) is an attractive theoretical lens for this issue, because it prevents collapsing a multifaceted problem into a one-dimensional solution. Specifically, the multidisciplinary approach—one that includes both the sciences and the humanities—reveals a multidimensional view of technology-society interaction, exemplified by the advent of AI.

Findings: After advancing this risk taxonomy, this paper utilizes the risk management framework of Lean Six Sigma (LSS) to propose effective mitigating controls for the identified risks. LSS determines controls through data collection and analysis, and supports data-driven decision making for industry professionals.

Implications to Theory, Practice and Policy: Instantiating the theory of STS into industry practices could be critical, then, for determining and mitigating real-world risks from AI. In summary, this paper combines the academic theory of sociotechnical systems with the industry practice of Lean Six Sigma to develop a hybrid model to fill a gap in the literature. Drawing upon both theory and practice ensures a robust, informed risk model of AI use in healthcare.

Keywords: Artificial Intelligence, Ethics, Sociotechnical Systems Theory, Lean Six Sigma, Academic-Industry Partnership
1.0 INTRODUCTION

Although intended to be the catalyst of new discoveries, the technology industry’s rapid developments in artificial intelligence (AI) have a problem: they aren’t addressing the associated risks outside of their own industry. This manifests itself in the hesitance and slow adoption of AI outside of the tech industry in 2020-2021 (Taulli, 2021). Even companies solidly rooted in the industry, such as giants like Google and Amazon, are concerned mainly with identifying and mitigating the risks that their AIs pose to other tech companies. This does not mean, however, that technological risks are limited to technologies as such. The risks are nested within social contexts, and pose risks outside of the latest technological trends, like fintech, martech, and adtech. Realizing this, business leaders in 2023 are moving to adopt and adapt to AI within their organizations: if Silicon Valley isn’t providing them scalable frameworks, and government isn’t putting forth solid regulations, these businesses must adapt, albeit at different paces, and with differing levels of success (Taulli, 2021; Korn, 2023; Claburn, 2023; Rajan & Rag, 2023; Pegoraro, 2023). These risks are most prevalent—and most damaging—in industries that serve humans in their most vulnerable states, such as healthcare.

Because the health and lives of patients are at stake in the healthcare industry, any exploration into healthcare tech must not be taken lightly. On the contrary, it must stand up to the intense scrutiny inherent in the industry: doctors must undergo specialized, rigorous training to be licensed (Health & Human Services, 2023); medical devices must meet strict regulatory requirements and submit to quality audits (U.S. Food & Drug Administration, 2022), and medical facilities must be accredited to receive certain privileges and funding (The Joint Commission, 2023). Applying AI to healthcare must be done with equal, if not more, caution and respect. First, however, it’s necessary to understand the culture that’s producing AI. It could be argued that technology is nothing more than the hardware and software themselves, and it would be reasonable for humanity to develop ethical frameworks around this idea. However, this paper argues that technology exists within a social context, and because of this, any ethical regulation, applied technology, or monitoring program must include the socio- part of the sociotechnical system.

Taking as an example the archetypical nexus of tech innovation, Silicon Valley, the tech industry is first and foremost a culture in itself. This can be understood through one of the fundamental tenets of Sociotechnical Systems Theory, which states that “technology is recognized as social practice within a specific institutional context” (Satori & Theodorou, 2022). One of these institutional contexts is startup culture, an informal subculture in Silicon Valley, where tech leaders and hopefuls create an unspoken social context among themselves.

If technology were developed in a vacuum, there would be no subculture. If this were true, there would be no aspirational Stanford students creating the next startup with no exit strategy, no (sub)cultural acceptance of dropping out of an Ivy League school with an idea that may disrupt the world as we know it, and certainly no angel investors funding risky startups that will either become household names or fail completely, forever forgotten in the innumerable unmarked graves of the Valley’s tech cemeteries (Lately, 2015).

The culture is fueled by a type of anti-corporate Newspeak: digital nomads depend on remote first companies, LinkedIn job offerings extol their “Chief People Officer’s” new unlimited PTO initiative, founders compete for the unofficial bragging rights conferred by proving that they had the scrappiest startup. This is, of course, before they rightsized their production to pivot to

2
focusing on being antifragile while increasing their runway. They tracked their burn rate to increase the power law that they feature in their slide deck when pitching to angel investors for their Series A funding. If all goes well, they become the most desired, most profitable, and—in a quasi-religious sense—the most revered in the Valley: a unicorn. (Lately, 2015)

Born out of this culture, modern AI development is defying almost all cautions about economic instability, bias, inequality, social responsibility, politics, and regulation. Although these cautions are coming from industry giants like Elon Musk, nevertheless, tech companies are still engaged in an ever-accelerating AI arms race (Knight, 2023). This is not technology for technology’s sake. Because there is clearly a culture behind AI innovation—complete with its own dialect—the tech industry needs a multidisciplinary research community to analyze the interaction of culture and tech. It’s critical to consider not only technological progress, but also the humanities. All fields of research should be involved in discussions about AI.

2.0 MATERIAL AND METHODS

An area most often neglected is the healthcare industry, comprising disparate institutions such as hospitals, long-term nursing facilities, medical device manufacturers, and payors (Retzinger & Retzinger, 2023). Since this extends beyond the patient experience into areas like manufacturing and finance, it’s critical that any healthcare AI has humanity at the center. In the same way, the constellation of other areas, such as patient safety and treatment effectiveness, must be patient-focused.

This problem is pervasive, both in the academic landscape and in broad swaths of industry: the remedy, therefore, must also be a combination of academia and industry practice. For its part, academia can contribute a theoretical framework of “sociotechnical systems” (STS), which takes a multidisciplinary approach—including both the sciences and humanities—to study and understand technology-based issues in the context of humanity writ large. The practical problem can be addressed by using, for example, Lean Six Sigma (LSS), a problem-solving process that is well-known in many industries (Council for Six Sigma Certification, 2018). Taken together, this combination of theory and practice, of sociotechnical systems and Lean Six Sigma, of the ivory tower and the boardroom, reveals a multidimensional view of technology-society interaction, and can offer a path to creating responsible AI for use in various industries.

Solving these problems—and the myriad others like them—requires collaboration between ethics and industry to use and develop responsible AI. In the healthcare industry, there are six quality goals defined by the National Academy of Medicine: safety, effectiveness, patient-centered care, timeliness, efficiency, and equity (National Academy of Medicine, 2019). Patient-centered care can be seen as an umbrella for the other five goals, since it encompasses safe, effective, timely, efficient, and equitable care for every patient. Acknowledging these goals, healthcare services should be centered around the patient in every context, and healthcare AI development and utilization are no different. With AI implementation, however, it is important to classify risks and roadblocks to ensure the model is properly and responsibly trained. Thus, each of the quality goals is susceptible to a corresponding sociotechnical problem type; the below table is not exhaustive, but denotes several problem areas:

---

1 A unicorn is a startup with a value of over $1 billion, pre-IPO
Table 1: Goals and Roadblocks in Patient-Centered Care

<table>
<thead>
<tr>
<th>Patient-Centered Goal</th>
<th>Sociotechnical Roadblock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>Managing expectations</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Educated guessing</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Rushing</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Busywork</td>
</tr>
<tr>
<td>Equity</td>
<td>Over-standardization</td>
</tr>
</tbody>
</table>

Apart from opposing their associated roadblocks, the five-quality metrics are often also competing against each other—especially in clinical settings where priorities can shift at a moment’s notice due to the patient's present condition. When attempting to overlay artificially intelligent technology onto these situations, the context becomes even more complex.

An example of this complexity is the fact that the healthcare industry has benefited greatly from advances in technology, and not just with hardware like MRI machines: electronic medical records and telehealth appointments have made categorical shifts in accessibility for patients. On the other hand, there are instances where developing unregulated healthcare AI results in biased, unethical, or invasive practices, impacting people across business and society. The AI, Algorithmic, and Automation Incidents and Controversies (AI, Algorithmic, and Automation Incidents and Controversies, 2023a) database is an independent, non-partisan foundation that tracks and documents these negative ethical and social incidents across industries.

Some healthcare AI incidents include NarxCare (AI, Algorithmic, and Automation Incidents and Controversies, 2023b), a controversial black box algorithm sold to the US government for monitoring citizens’ prescription drug use and assigning them an addiction risk score; an AI predictive model for sepsis detection by Epic Care that claimed 76% accuracy and was put into use in real clinics, only to miss two-thirds of cases of sepsis and raise many false alarms (AI, Algorithmic, and Automation Incidents and Controversies, 2023c); and Idaho Medicaid cutting benefits to disabled people because the software vendor updated the black box algorithm (AI, Algorithmic, and Automation Incidents and Controversies, 2023d). These incidents demonstrate the need for a clearly-defined ethical framework to govern AI use by the healthcare industry.

To determine which ethical inputs matter most, it is important to narrow the scope of inquiry. Bernd Carsten Stahl, a leading researcher in the field of ethical technology, noted that we must always consider that “[i]t is rarely possible to draw a clear line between one particular component of [a technology, because] … the overall system is greater than the sum of its parts. And to complicate matters even further, any individual intelligent system (e.g. a fraud detection system in an insurance [sic] or an autonomous vehicle) is embedded in … a broader set of technical and social systems (Stahl, 2023). One answer to Stahl’s quandary lies in Sociotechnical Systems Theory, which treats every technology as an interconnected system between humans and technology.
Why Sociotechnical Systems Theory?

Scholars from Leeds University offer a high-level description of the core of the sociotechnical lens: “[Sociotechnical Systems Theory] has at its core the idea that the design and performance of any organizational system can only be understood and improved if both ‘social’ and ‘technical’ aspects are brought together and treated as interdependent parts of a complex system” (Leeds University, 2022). The choice to use STS flows naturally from the problem: it looks at technology as a social practice, and the problem is how to apply technology to healthcare, an industry highly-integrated into the fabric of society. STS encourages a melding of technological excellence with social responsibility.

One of the most detailed explanations of sociotechnical systems comes from the influential scholar Ibo van de Poel, known for his contributions to the ethics of applied technology: there are institutions, technical artefacts, human and artificial agents, and technical norms (van de Poel, 2020). His definitions and roles are summarized in the figure below. In this case, human agents are interacting with an AI, informed by a sociotechnical view of the clinical context (the institution). The AI is then providing outputs back through artificial agents, which are the smallest units of autonomous decision-making. The technical norms that mediate between the human and artificial worlds “[do] not ultimately rest on human intentions, as is the case with institutions, but on the (causal) laws of nature” (van de Poel, 2020). In a sociotechnical system with AI, the technical norms may be conceptualized as the code that programs and interacts with the AI system:

![Diagram of Sociotechnical Systems Theory]

**Figure 1: The Core Relationships of a Sociotechnical System. Figure By Author.**

Using STS as a lens to identify and evaluate the associated risks results in a more comprehensive view of potential problems: it offers a multidisciplinary approach to not only question-solving, but question-asking. Ethical issues arise from either area of the diagram, such as a patient’s language barrier, or bias in the AI’s training dataset. After considering ethical issues are defined through the lens of sociotechnical systems, a framework must be chosen to instantiate particular, concrete solutions. Stahl clarifies in a 2023 paper: “Many of these [proposals to address AI ethics] aim to provide guidance to AI experts on how to ensure that ethical issues do not arise or can be mitigated. This includes work on opening up AI to critical scrutiny, for example [with] explainability or forensic examination of AI. … One approach to ensuring responsibility is to integrate AI design
and development in existing mechanisms aimed to ensure responsibility, such as risk management frameworks” (Stahl, 2023x). This need for ethical guidance is not abstract, as Stahl correctly argues, but requires concrete, comprehensive frameworks. He continues, stating that it’s possible to use existing mechanisms, which will help with industry adoption; AI is already a new concept, so adding on more novelty may deter business leaders, especially those outside the tech industry. This is echoed in a 2020 paper from Oxford: “mistakes or misunderstandings may lead to social rejection and/or distorted legislation and policies, which in turn cripple the acceptance and advancement [of healthcare AI]” (Morley et al., 2020). Other research teams have put forth similar ideas (Eitel-Porter, 2020; Farhud & Zohaei, 2021; Shah & Martin, 2023; van de Poel, 2020; Xing et al., 2023). All of the researchers are advancing the same idea: that AI implementation is difficult, and the most important factor is ease-of-use for non-tech-focused business leaders.

Outside of the tech industry, AI adoption depends heavily upon the level of stakeholder understanding and transparent explainability. Bringing this theoretical approach into real industry problems can be accomplished through instantiating it in an existing framework. LSS is an attractive fit, since its statistical approach shines when addressing problems with unknown causes. Additionally, it is robust enough to handle the complex and novel challenges brought by AI, and sufficiently flexible to remain relevant and useful to both the healthcare and AI fields. Because of this, Lean Six Sigma can bridge the gap between theory and practice.

**Why Lean Six Sigma?**

Initially developed in Toyota and Motorola factories, Lean Six Sigma is a collection of problem-solving methods used throughout various industries to deliver a product or service with consistent quality (Council for Six Sigma Certification, 2018). LSS takes consistency seriously: the goal is only 3.4 defects in 1,000,000 opportunities, or a 99.99966% success rate. This is what the name Six Sigma represents: the goal of improving a process to such a degree that there are no defects within ±6 standard deviations. The “Lean” qualifier for LSS is indicating that the goal of any given LSS system is to reduce various kinds of waste: excess inventory and excess worker movements.

Methodologies used in LSS projects, such as fishbone diagrams and control charts, emphasize clear data visualizations and cause/effect diagrams. In LSS, the methodology flows from the chart (Council for Six Sigma Certification, 2018), so the process and results are transparent from the beginning of the investigation to the end. The central problem-solving framework here is DMAIC: Define, Measure, Analyze, Improve, and Control. Professionals can apply the DMAIC principle to various goals by asking questions at each stage in the process. The general format is shown in Figure 2:
The ethical questions in each quality goal can be first identified as a sociotechnical system, then improved by an AI model, and finally monitored by Lean Six Sigma frameworks to verify responsible use. The analysis below details the problem-solving for each of the five quality goals set forth by the National Academy of Medicine.

3.0 FINDINGS

Problem Solving for Patient Safety

The first quality goal is safety. For example, a hospital unit can be a dangerous place for patient safety, especially regarding fall risk. There could be an AI model that uses data from cameras and movement sensors to predict a patient’s risk of falling; this can be evaluated as a sociotechnical problem, noting the main ethical issue of surveillance vs. safety. The figure below is one way to conceptualize this:

The sociotechnical risks involve using cameras and movement-monitoring technology to capture data from and about a patient, and how this affects their privacy, dignity, and safety. Essentially, the ethical risk is the tradeoff between dignity and safety. One data-driven way to monitor the AI performance is through a violin chart, a common Six Sigma tool. The figure below illustrates hypothetical data on actual falls (orange) vs. the times when the AI raises an alert that a patient fall has occurred (blue):
Figure 3: Violin Plot Depicting Actual Vs Predicted Falls. Figure By Author.

The chart clearly shows that the AI model is not performing well: it’s not catching a majority of the actual patient fall. The patient, care team, and family would need to discuss if the ethical risks inherent in the camera and movement monitoring system is worth the marginal safety improvement. This is the heart of marrying STS with LSS—the result is a robust framework that spurs meaningful conversation and real improvement, through displaying reliable, explainable data from a transparent algorithm.

Problem Solving for Treatment Effectiveness

The next quality goal is to ensure a given treatment is effective. Taking physical therapy as an example, consistency and followup are key, as is continual monitoring. An AI model could help with this, drawing on data from wearable tech and periodic patient surveys. From a sociotechnical perspective, the ethical risks would surround data privacy and algorithmic intrusiveness. Here are some sociotechnical considerations:

Figure 4: The Sociotechnical System of an AI Predictive Model. Figure By Author.
The main ethical concern, from a sociotechnical perspective, is the intrusiveness of the AI-based intervention into the patient’s daily life. A Lean Six Sigma solution would be creating a fishbone (Ishikawa) diagram to investigate the root cause of a patient’s perception that the tech is intrusive. In this hypothetical example, the patient is a retired elderly person who isn’t very tech-savvy. The LSS-based diagram could look like the following figure:

**Figure 5: A Fishbone Diagram Depicting Patient Issues with a Wearable Monitor.**

*Figure By Author.*

The fishbone diagram here highlights the patient’s problems with the data collection instruments. The LSS framework begins at the “head” of the fish, stating the problem, then working back through the various “ribs,” each representing a different reason behind the problem, as in the figure above. This is usually an initial step for brainstorming and identifying which areas to investigate to determine the root cause and contributing causes.

**Problem Solving for Timeliness, Efficiency, & Equity**

The interrelated goals of timeliness and efficiency can be supremely important when evaluating if an AI is promoting health equity. Time is valuable, both for patients and their medical care team. Related to the goal of timeliness, efficiency is a hallmark of a well-constructed professional structure, but busywork impedes this goal. Busywork takes away time and energy from humanity, in a way that machines generally aren’t subject to. This meta-work often is only tangentially related to the actual task being completed—the one that requires human skill. Technology, such as an AI model, could take this aspect over, not only freeing human professionals to perform their skilled work more often, but also completing the busywork part of their profession with more speed and accuracy than a human could. The ethical issue here, however, surrounds data privacy and accuracy; AI models are not, at the time of writing, always accurate in their transcription or classifying medical insurance coding, especially given the specialized terms, conditions, and drug names in the industry. Sociotechnical and ethical concerns could look like the following figure:
In this example, there are a number of ethical implications; in an effort to improve timeliness of care and medical team efficiency, the healthcare industry must be extremely careful when applying language models into clinical settings. There are risks of data misuse, misconstruing meaning, language barriers, mixing names and properties of drugs, and biases against, for example, cultures who employ storytelling-based histories as opposed to factual-timeline-based histories. Using these as inputs, then, to train a personalized generative AI for an individual patient’s care seems like a high-risk/high-reward gamble; at the time of writing, neither academia nor industry has any concrete controls in place to handle the advent of “gen AI” models, and remain opaque even to their developers. This is an area for further research in the coming months and years. Academicians can inform industry leaders by researching, synthesizing and offering solutions from an ethical perspective, informed by the humanities-focused approach of Sociotechnical Systems Theory.

4.0 CONCLUSION AND RECOMMENDATIONS

When considering the unique ethical challenges in the healthcare field, it is critical to acknowledge that it is an industry that is hyper-focused on humanity and human needs. Bringing technology in should always be in service to that goal, and must be consistently monitored to ensure that it stays on track.

Practitioners can determine which ethical principles are at the most risk by using STS to understand the social and ethical underpinnings of any human-computer interaction. Industry leaders can then focus their initiatives on maintaining responsible, explainable AI models, validated through LSS, that demonstrably serve human health and minimize risk. In this way, academia and industry can work together to build frameworks that place meaningful controls on new technology, even in the absence of official regulation. Acknowledging and supporting a synthesis—of academia and industry, of tech leaders and healthcare practitioners, of humanity and technology—will integrate them all for mutual benefit, leading to a brighter future for public health across the globe.

Figure 6: A Sociotechnical System Diagram Illustrating the Human-AI Relationships in a Primary Care Setting

Figure By Author.
REFERENCES


Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. Scientific Reports (13)1. https://doi.org/10.1038/s41598-023-34622-w


**License**

*Copyright (c) 2024 Dillon Plummer*

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a [Creative Commons Attribution (CC-BY) 4.0 License](https://creativecommons.org/licenses/by/4.0/) that allows others to share the work with an acknowledgment of the work’s authorship and initial publication in this journal.