Open Access

Scinexis Journal of Medical Devices and Healthcare Innovation
Volume 1 | Issue 1 | Nov-Dec | 2023
https://www.scinexis.org/journals

Research Article

AI Copilots for Risk Management: How Human-in-the-Loop Predictive Models Informed by Sociotechnical Systems Theory Can Augment Quality Assurance Professionals to Decrease the Risk of Harm in Medical Device Manufacturing

Dillon Plummer*
Capitol Technology University, United States

ABSTRACT

Risk is always present in any human endeavor, but when designing and creating medical devices, it is critical to get it right. International bodies and the United States FDA are updating their standards and regulations to focus more on risk management, so the need for a solid framework has never been more relevant. This paper seeks to establish a risk management framework for quality assurance (QA) professionals in the medical device manufacturing industry, based on AI predictive models. Because QA requires significant human feedback, any predictive model will also require some level of human input. This can be achieved by using a human-in- the-loop (HITL) machine learning mode, which uses both AI and human analyses to solve problems in a given environment. This model will be grounded in Sociotechnical Systems Theory (STS), a lens that analyzes the role of humans in a given technical system; this analysis allows for the technical goals of HITL models to be achieved in human-centered industries. Various quality domains are explored in this paper, including product design, material nonconformance, supplier quality and complaints. These areas can all be augmented by an HITL predictive model within an anthropocentric context. The framework presented in this paper delineates how HITL models can serve as copilots for quality assurance professionals in large part, they can accomplish this by developing and monitoring risk for various tasks, based on the company's quality management system (QMS) and human evaluation. As these machine learning models become more accurate, they will serve as efficient force- multipliers for bolstering public safety.

Keywords: *Risk management, artificial intelligence, medical devices, human-in-the- loop algorithm, Sociotechnical Systems Theory*

Corresponding Author: Dillon Plummer
Address: 11301 Springfield Rd, Laurel, MD, 20708, USA
***E-mail:** dtplummer@captechu.edu
Received on: Nov 17, 2023; Accepted on: Nov 21, 2023; Published on: Dec 31, 2023
Citation: Plummer, D. (2023), AI Copilots for Risk Management: How Human-in-the-Loop Predictive Models Informed by Sociotechnical Systems Theory Can Augment Quality Assurance Professionals to Decrease the Risk of Harm in Medical Device Manufacturing, *SJMDHI* 1(1), 1-9.

INTRODUCTION

1. The Current Problem and a Proposed Solution

Writing nearly 2,000 years ago, Seneca notes that every human and society is subject to "Fortune" or risk. In a letter to his colleague Lucilius, he advised: "Let the mind be disciplined to understand and to endure that there is nothing which Fortune does not affect that she has the same jurisdiction over empires as over emperors, the same power over cities as over citizens" [1]. Risk does not discriminate based on government, industry, or individual, so it must be approached with a "Disciplined mind" to "Endure." This discipline must be made manifest in every human endeavor, regardless of context.

However, there are certain industries that must guard against risk in a more comprehensive way than others, such as those focused on human medical safety and public health. This effort of risk management can be described by Seneca's exhortation to maintain a disciplined mind when developing mitigation measures and controls. It is the human mind that must be the source of disciplined, rigorous process when dealing with other humans. In the modern world, however, human efforts alone would struggle to mitigate or even identify the risks involved in a product or process. Because of this, companies rely on sophisticated AI technology to address the myriad risks of ever-advancing manufacturing technology in regulated industries[2]. Nevertheless, this does not come without its own risks: even risks to humanity. The singular goal and guiding principle should be the benefit of humanity using all available resources.

With this goal in mind, humanity has developed tools to assist with this effort, such as creating other "Intelligences" that may be more disciplined and vigilant than their own. AI can fill this role and assist professionals in managing risk, especially in the regulated industries that deal so intimately with human health, such as medical device manufacturing. AI-based risk management, then, must be informed by both humans and AI. Risk assessments and risk management plans must follow the government regulations regarding risk: "The FDA's benefit-risk assessment of a product is a case-specific determination that requires a thorough assessment of the extensive evidence of safety and effectiveness"[3]. Because of these stringent government requirements, risk management is always a major goal in the medical device industry. This paper will examine how to achieve that goal using both anthropocentric theory and AI predictive models, specifically Sociotechnical Systems Theory and Human-in- the-Loop machine learning.

2. Bridging Theory to Practice

What is "Risk?" The United States' Food and Drug Administration (FDA) has defined risk management as a bifurcated goal: "The combination of the probability of occurrence of harm and the severity of that harm"[4]. The organization has, in recent years, increased their emphasis on risk management in the medical device and healthcare industries[5]. The FDA's flowchart below illustrates their position on determination of risk; this paper seeks to develop a risk management framework for the bottom box, "Risk of Harm".

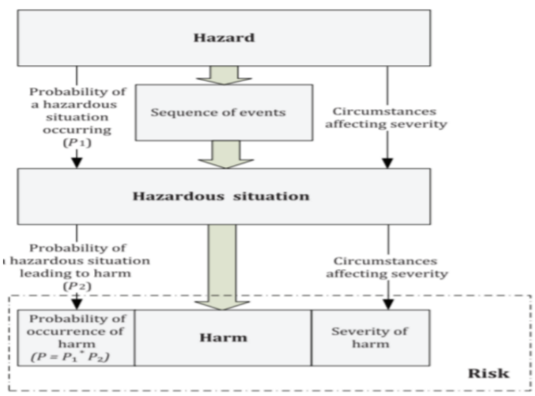


Fig. 1: The process of assessing risk. Image source: U.S. FDA[4]

2.1 Sociotechnical Systems Theory: The Watchdog against Untrammeled Technological Progress

An appropriate lens to examine this goal exists in Sociotechnical Systems Theory (STS), which emphasizes the relationship between humans and technology. STS is "Based on the recognition that Artificially intelligent systems are classical examples of sociotechnical systems. This means that they consist of an assemblage of heterogeneous components including individual humans, technological artefacts, and social structures"[6]. According to STS, humans and technical artefacts the STS term for individual instances of technology, such as software platforms or AI models are necessarily interrelated. This relationship permeates every human computer interaction.

2.2 Human-in-the-Loop Algorithm: A Human-Informed Copilot

One popular human-technology relationship in the modern industry landscape is Humanin-the-Loop machine learning (HITL), where an AI makes decisions and creates a framework alongside input and analysis from humans. This cooperation should be both a goal and a reminder of caution, as noted by Stahl in his discussion of AI as sociotechnical systems: "Ethical issues arise when these sociotechnical systems interact with their environment"[6] emphasizing that this interaction has far-reaching theoretical and tangible consequences. To explore these implications, this paper proposes an HITL model that is informed by the beneficent humanism espoused by Sociotechnical Systems Theory (STS) to create a medical device risk management framework.

An examination of the current academic landscape reveals both impactful theoretical research and gaps near the cutting edge of industry. Several studies have focused on promoting the importance of sociotechnical interactions in the healthcare industry, each focusing on a different sociotechnical area.

2.2.1. STS Focus: Principles

One research area focusing on STS foundational principles is typified by an Iranian article from 2021[7], which advances a sociological emphasis on human factors in a clinical setting. It discusses ethical concerns such as privacy and data protection, informed consent, social gaps, medical consultation, empathy, and sympathy through the lens of STS. The authors discuss these STS foundations extensively and thoroughly, noting the potential problems with their application to healthcare. These principles serve as an important starting point in applying theory to industry practice, but they do stop short of an analysis of a full sociotechnical system, with its heterogeneous components. The article focuses on brief bullet point problem listing, and although the authors mention AI, they don't specify a technical artefact or institution, rather than acknowledging the entire tableau of a sociotechnical system.

2.2.2. STS Focus: Institutions

Another type of research area is exemplified in a 2021 paper from Brazil,[8] which emphasizes the "Institution" aspect of STS. The Brazilian researchers examine government-run pharmaceutical services and its interaction with the sociotechnical system in their country. The research team thoroughly discusses the institution's role in the system and its different aspects: culture, management, workforce, structure, processes, and technology. This is valuable information about public health, but the research does not address machine learning, or AI systems in general.

The most relevant paper on this topic comes from Ibo van de Poel[9]. In it, he argues that AIs are themselves sociotechnical systems. These AI models should include not only the typical sociotechnical structural elements of human agents, technological artefacts, and institutions, but also added "Artificial agents" and "Technical norms"- his term for the processes that regulate relations between artificial agents and outer systems.

These additional elements are critical to a comprehensive view of AIs as sociotechnical systems. As can be seen in current extent AI systems, there are indeed processes that regulate human computer interaction from within the system itself. Machine learning models, for example, can sometimes produce a black box decision process; from the outside, then, the processes become artificial agents themselves.

There is a gap in the sociotechnical literature, wherein the theory is rarely applied to industry practice in the context of machine learning, especially in non-tech sectors. This deficit corresponds with the absence of the humanities in scholarship around machine learning and algorithm implementation. On issues of ethics, especially concerning devices placed into the human body, it involves every field of human inquiry.

3. Adapting Sociotechnical Systems

A sociotechnical approach to AI development includes above all else the "Human" aspect of "Human computer interaction." Because of this, applying Sociotechnical Systems Theory to machine learning algorithms is essential to the development of responsible AI systems. This can advance human-centered algorithms forward into non-tech sectors, especially the healthcare and medical industries. It is critical to expand the narrow view of algorithms as pieces of software, into viewing them how van de Poel argues: As technical artefacts in an interconnected system of human and artificial agents, interfacing through technical norms, and encompassed by institutions.

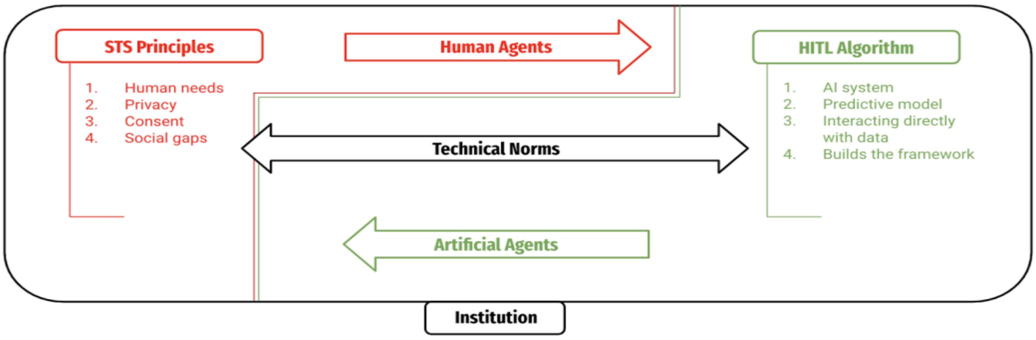


Fig. 2: The integration of STS Theory into an HITL algorithm

4. Building the HITL Algorithm

Informed by this model of STS into a Human-in-the-loop algorithmic model (Fig. 2), the process of the HITL algorithmic process itself must be advanced and analyzed. The HITL process can then build a risk management framework. The high-level process flow is shown in the figure below:

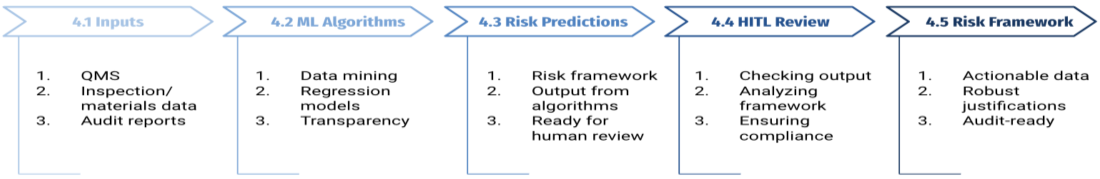


Fig. 3: The stages of using an HITL algorithm to produce a risk management framework

4.1 Inputs

In the context of the medical device industry, the inputs for the HITL model may come from the company's quality management system (QMS), inspection data, material reports, and audits. The QMS is a crucial resource, as it drives decisions for the entire company and secures its ISO certification (ISO 9001 and/or ISO 13485). Quality professionals can feed Human-in-the-Loop machine learning models data from the ISO Standards and the QMS for it to learn from and to become familiar with the company's policies, procedures, and controlled documents. This is the "Technical" aspect of the AI as a sociotechnical system; the "Socio" portion is, as the authors argue, focused on principles such as data privacy and model transparency. Both are needed for audits, and both are needed to validate any risk management model that the AI produces.

4.2 Machine Learning Algorithms

Once the model is trained on these data, it can then be tested. For example, QA engineers can have the HITL model perform a Failure Mode and Effects Analysis (FMEA). It may consider the design's compliance with ISO and the QMS, as well as potential failure modes. Engineers can make the model's predictions more robust by including previous customer complaints and production floor reworks this will increase the quality of the model 's predictions. Figures 4 and 5 below demonstrate this example:

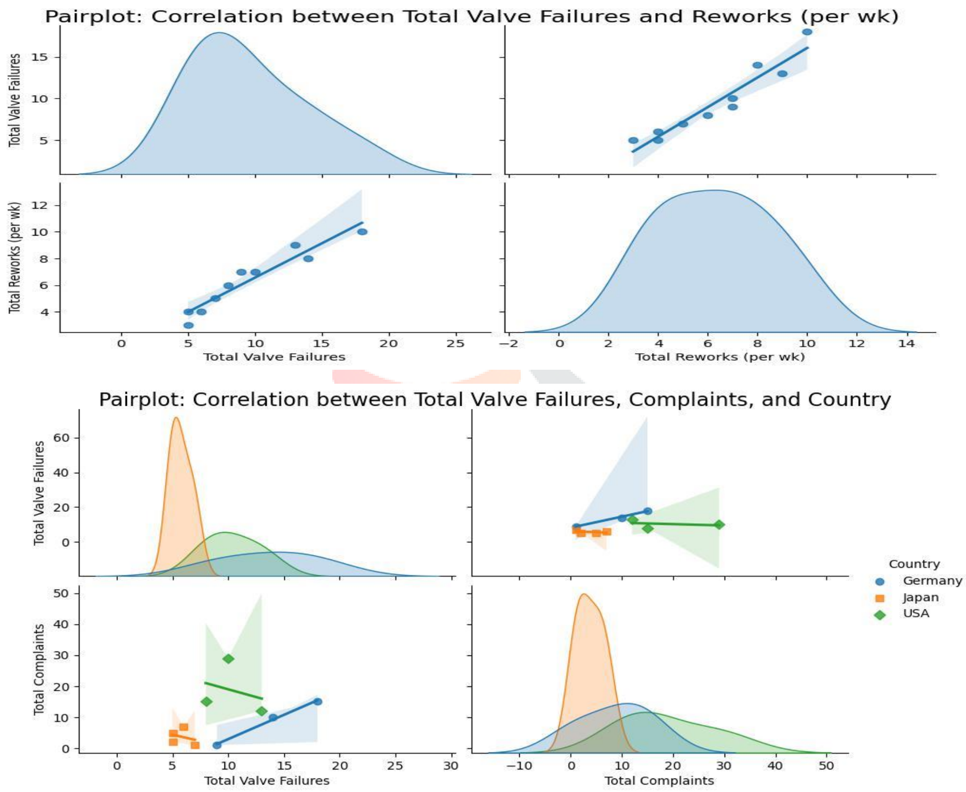


Fig. 4 (Top) and **Fig. 5 (Bottom):** An example of a linear regression model to estimate risk of component failure, given four inputs: total weekly reworks, valve failures, total complaints, complaint country. The data used in this example is fictional, and contrived to illustrate a point.

	Total Reworks (per wk)	Total Complaints
Coefficient	0.513514	0.885886
P-value	0.000016	0.202879
R-squared	0.912794	0.193841

Fig. 6: The coefficient table for the previous graphs. The data used in this example is fictional, and contrived to illustrate a point

4.3. Risk Predictions

As mentioned above, a human professional should always review both the input data and output to ensure both compliance and accuracy. In this example, the QA team, very reasonably, might have thought to look at the customer complaint data to find the cause of the valve failures, based on the country in which they're sold. However, using an HITL machine learning model, they can now see that production floor reworks are a much better predictor ($p < .001$) than country complaints ($p = .20$) for determining the risk of valve leaks. P-values estimate the chance of results happening as a result of pure chance, rather than a true correlation. The R2 value also reveals a strong correlation ($R2 = .91$) of the product reworks with valve failures, which demonstrates how well the regression model fits the data. For this example, the company's QMS states that a result where $p \leq .05$ is significant, and sufficient to reject a null hypothesis. The AI model, then, is predicting that the production line is a much riskier place for valve failures than when they're already delivered to customers. Essentially, the HITL model's goal is to predict the probability of risk, which falls in line with the FDA's risk framework in Fig. 1. The risk prediction leaves it up to the quality engineers to ask "Why?" and begin the root cause investigation. This example illustrates that the social and technical aspects of Sociotechnical Systems Theory dovetail, when understood properly, and draw on the strengths of both algorithms and human professionals.

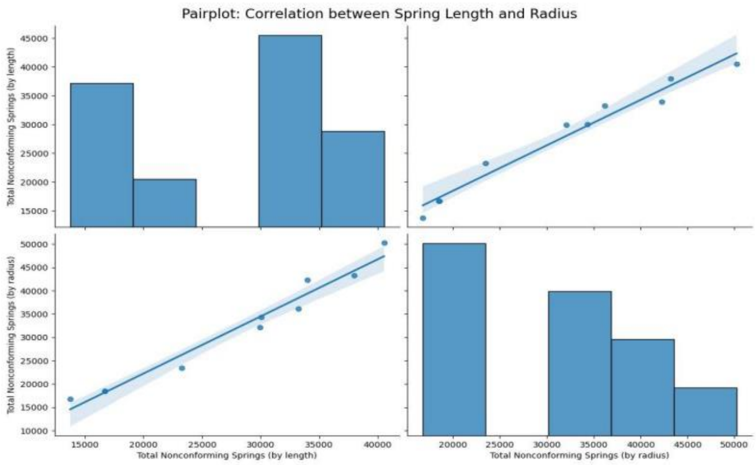
Nevertheless, an argument could be made that the QA team would very well have determined the location of maximum risk without the use of an HITL algorithm, and this is correct. However, starting at the customer complaint level and working backward stepping back through customer receiving, shipping, outgoing inspection, and only then back to the production floor is far more resource intensive than working with an AI collaborator.

4.4. Human-in-the-Loop Review

From a sociotechnical perspective, the presence or absence of the human agents or artificial agents would have a large impact on the system as a whole. For example, if the AI weren't present, the company's supplier quality engineers would have to reach out to other companies in the supply chain, potentially straining relationships with customers and suppliers, just for the purpose of tracking the root cause. On the other hand, if the humans weren't in the loop, the AI would have been a detriment to internal relations as well: it could very well say that the production floor workers are the problem and present its findings unquestioned. This is a powerful example, illustrating that in this post-AI business environment, a sociotechnical partnership is critical to building a robust and fair risk management model.

4.5. Putting It All Together: The Risk Framework

Beyond root cause analysis, this risk evaluation can also be used for analyzing material nonconformances and supplier quality. When analyzing nonconforming components, the QA inspectors have another use case for an HITL predictive model. Based on their AQL sampling plan, they can reject an entire component shipment due to crossing a defined quantity threshold of rejected parts. The inspectors can put their dimensional measurements into to a predictive model, and, drawing from the company's database (Such as an MRP system), compare the current nonconformances against those in the past. This example uses spring lengths and radii measurements to classify nonconforming materials:



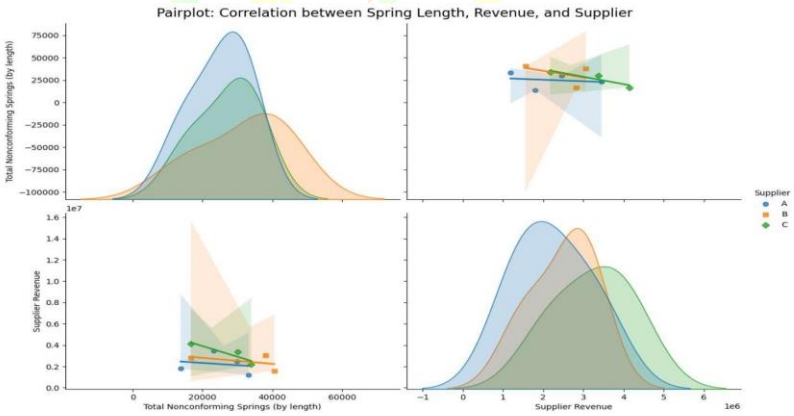


Fig. 7 (Top) and **Fig. 8 (Bottom):** An example of a multiple linear regression model to estimate risk of material nonconformance, given four inputs: total NC springs by length, total NC springs by radius, supplier, and supplier revenue. The data used in this example is fictional, and contrived to illustrate a point

These figures illustrate an interesting situation: Quality inspectors are measuring springs and notice that there is an increasing number of dimensional nonconformities (NC) in length. Informed by STS, the human agents in the system look to other institutions and outside human agents, such as their supplier data. The QA team could conjecture that the supplier's quality has decreased since their revenue has decreased. However, the AI model demonstrates that the risk doesn't lie with the supplier's revenue. Although it's not correlated with NC spring lengths, there is an interesting explanation within the company's own data: historically, as the amount of rejections increases due to NC lengths, so does the amount of rejections for NC radius. This indicates that the risk lies not in the supplier's financials, but in their machines.

The probability of risk is high, since there are two nonconformities appearing in tandem; this suggests something beyond one part of the tooling getting worn or tolerance creep, but instead that there could be a significant root cause in the supplier's plant one that's spiraling downward and creating components that aren't

simply wrong, but consistently wrong. The machine learning model can also predict that the severity of the risk is also high, since the slope of the linear regression line is quite steep, showing that each component coming off the supplier's line is more out of spec than the previous one.

The human agents in the system (QA engineers) can then address that issue with the other institution and its human agents (QA engineers at the supplier's facility), who will then correct their technical artefacts (Production line machines). This example highlights that an HITL model can predict both the probability and severity of risk, which is both the heart and primary goal of risk management.

5. CONCLUSION

The medical device industry has greater consequences than most other industries, so properly and thoroughly assessing those risks is paramount. Sociotechnical Systems Theory insists that technology exists in an "Ecosystem" with humans they should be copilots for the AI, just as much as the AI is a copilot for humanity. The emphasis on humanity's involvement in the risk management process, especially in medical devices, is key. By leveraging AI models, humans can develop processes with the discipline and rigor that Seneca lauded nearly two millennia ago. The sociotechnical systems built up to mitigate risk can be strengthened through the use of algorithms as a driving force, keeping humanity at its core. The cooperative relationship between human and machine will allow for risk management frameworks that are data-driven and human-centered: human QA professionals, aided by AI copilots, will shape the future of healthcare and public safety.

REFERENCES:

1. Seneca the Younger. (1969). Moral letters to Lucilius (Gummere, Richard M., Trans.). https://topostext.org/work/736. (Original work published ca. 65 C.E.)

2. Turner, C., & Oyekan, J. (2023). Manufacturing in the Age of Human-Centric and Sustainable Industry 5.0: Application to Holonic, Flexible, Reconfigurable and Smart Manufacturing Systems. *Sustainability*, 15(13), 10169, https://doi.org/10.3390/su151310169

3. U.S. Food & Drug Administration. (2023). Benefit-risk assessment for new drug and biological products: guidance for industry. U.S. Department of Health and Human Services. https://www.fda.gov/media/152544/download#:~:text=FDA's benefit risk assessment of understanding of the data gaps

4. U.S. Food and Drug Administration Center for Devices and Radiological Health. (2023). Risk management and the total product lifecycle (TPL). U.S. Department of Health and Human Services. https://www.fda.gov/media/150416/download

5. Buocz, T., Pfotenhauer, S., & Eisenberger, I. (2023). Regulatory sandboxes in the AI Act: reconciling innovation and safety?. *Law, Innovation and Technology*, 15(2), 357-389.

6. Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports* (13)1, 7586. https://doi.org/10.1038/s41598-023-34622-w

7. Farhud, D. D., & Zokaei, S. (2021). Ethical issues of artificial intelligence in medicine and healthcare. *Iranian Journal of Public Health*, 50(11), i. https://doi.org/10.18502/ijph.v50i11.7600

8. Bernardo, N. L. M. C., Soares, L., & Leite, S. N. (2021). A Sociotechnical Approach to Analyse Pharmaceutical Policy and Services Management in Primary Health Care in a Brazilian Municipality. *Pharmacy* 2021, 9, 39. https://doi.org/10.3390/pharmacy9010039

9. Van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, 30(3), 385-409. https://doi.org/10.1007/s11023-020-09537-4