Human Factors in Design, Engineering, and Computing, Vol. 159, 2024, 372–381
https://doi.org/10.54941/ahfe1005602

Explainable AI Solutions for the U.S. Coast Guard Command Center: A Human-Centered Collaboration

Audrey Haque[1], Anthony Lapadula[1], Jessamyn Liu[1], Sara Falkson[1], Karli Blanchard[2], Richard Coleman[2], and Amna Greaves[1]

[1]MIT Lincoln Laboratory, Lexington, MA 02421, USA
[2]United States Coast Guard, Boston, MA 02109, USA

ABSTRACT

With advances in artificial intelligence (AI) comes the responsibility to ensure that deployed AI solutions are ethical, useful, and safe. Explainable AI (XAI) has drawn increasing interest from the AI research community and seeks to provide understandable descriptions of how machine learning (ML) models generate their outputs. In short, XAI allows users to peek into the incredibly complex black boxes that most ML models have become. As successful adoption of new XAI tools necessitates designing "with," and not just "for," this paper explores the use of human-centered, participatory design in partnership with United States Coast Guard (USCG) command center watchstanders. Our process included traditional research methods such as interviews, observation, and contextual inquiry, as well as user experience (UX) workshop research methods such as experience mapping, post-ups, affinity diagramming, forced ranking prioritization exercises, and storyboarding. Our goals were to understand the unique problems and opportunities of the USCG's Search and Rescue (SAR) mission, collaboratively generate desirable XAI solution ideas with command center watchstanders, elicit watchstander ideas and requirements for explainability features, and prototype our ideas to better meet real-world operational needs.

Keywords: Explainable AI, XAI, Artificial intelligence, AI, Human-centered design, Participatory design, Human-centered AI, Human-machine teaming, USCG

INTRODUCTION

Despite advances and growing interest in artificial intelligence (AI), over 80% of projects featuring AI fail – double the rate for projects not featuring AI (Ryseff et al., 2024). Most AI projects fail for several reasons: misunderstanding the problem to be solved with AI, project teams prioritizing a technology solution over solving real problems for end users, and a lack of high-quality data to effectively train models (Ryseff et al., 2024; Wilson et al., 2024). Another major reason for failure is lack of user trust due to the opacity of AI systems (Haque et al., 2023) and general awareness of AI projects that reflect and perpetuate societal failures (Dwork and Minow, 2022). Notable AI projects with ethical issues include a Twitter bot being verbally abusive, Google's sentiment analysis creating homophobic content, and Amazon's AI recruitment tool facilitating discrimination against women (Dastin, 2018; Yampolskiy, 2019). These projects highlight the risks of black box models (Ehsan et al., 2022) and how bias in the real world leads to bias in available training data, which then allows AI trained on that data to both perpetuate existing biases and cause harm in the real world (Meske et al., 2022).

As the field of artificial intelligence advances, there is increased responsibility to ensure that deployed AI solutions are ethical, useful, and safe. Guided by President Biden's Executive Order governing the use and development of artificial intelligence (White House, 2023), the United States Department of Homeland Security (DHS) highlighted the critical need for explainability in all of its Science and Technology Directorate initiatives (Department of Homeland Security, 2024).

In the context of AI, explainability refers to the ability to explain the decisions made by a model in ways that humans can understand (Haque et al., 2023; Kore, 2022). Explainability can help identify and mitigate risks by providing users with useful information about how or why a model produced a particular output. If, for example, a computer vision model is trained to predict the species of a fish, an explainability feature highlighting what areas of the image were most useful in predicting the species might be able to identify potential problems, such as the model focusing on boat identification numbers instead of the fish (Shperber, 2017). In short, how well a model explains how it works can determine its fate, and what constitutes a "useful" explanation depends on several factors, including the viewer's domain, role, technical knowledge, and goals (Kore, 2022; Suresh et al., 2021).

This study investigates unique problems and opportunities within the USCG's Search and Rescue (SAR) mission and describes XAI solution prototypes collaboratively generated with USCG command center watchstanders.

PROCESS AND METHODOLOGY

In alignment with human-centered design principles, the MIT Lincoln Laboratory (MIT LL) team intentionally focused on people and solving the right problems instead of following one particular methodology (Norman, 2020). Successful creation and adoption of new systems necessitates designing "with" – and not just "for" – people, actively involving them in decision making (Bravo, 1993). Merely involving people in design or conducting user research does not make a project human centered; rather, human-centered design requires viewing humans as "central in every aspect of the design" – and not just as part of the system (Auernhammer, 2020). As such, we sought to engage with USCG watchstanders not only as subject matter experts, but as our collaborators and teammates in decision making.

From among the USCG missions that Sector Boston commonly conducts, we decided, in collaboration with our DHS partners, to select the Search and Rescue (SAR) mission. We worked with Sector Boston's command center watchstanders to conduct user research (Fessenden, 2021; Kore, 2022) and collaboratively select and prototype a SAR-related use case that would benefit significantly from explainability (Falkson, in preparation).

User Research Planning

Prior to each field visit, the MIT LL team wrote research plans that included research objectives, research questions, research methods, and flexible agendas, adjusting as needed to accommodate for unpredictable command center activity. After each field visit, we synthesized our findings, incorporated them into project decisions, and generated new sets of research objectives and research questions to be answered during our next visit.

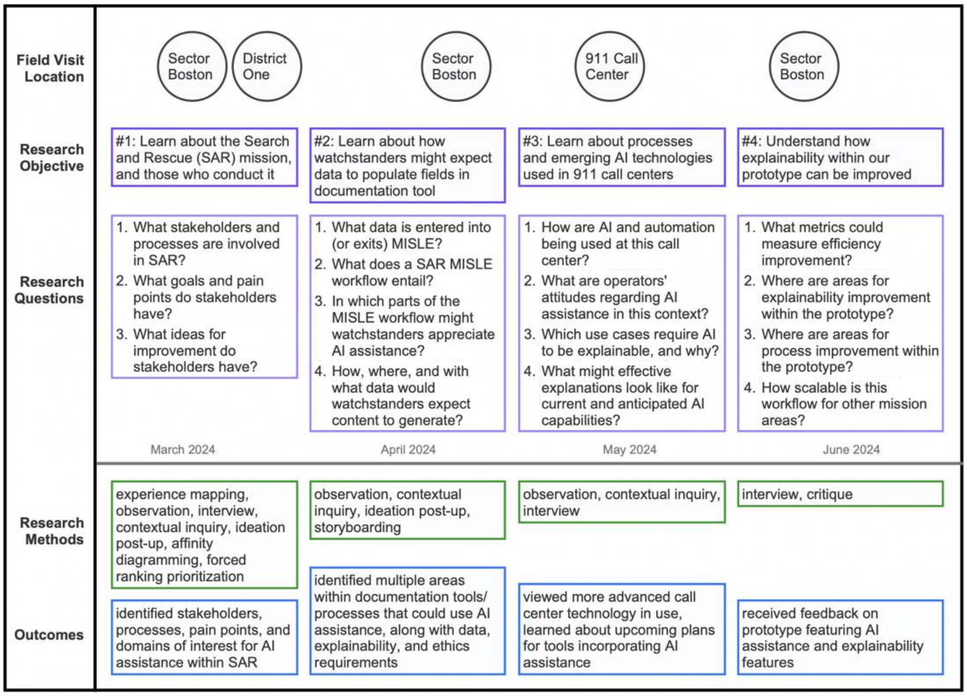

Figure 1: Timeline overview of our research objectives, research questions, research methods, and outcomes for each field visit.

User Research Methods

During our first visit to Sector Boston, we received an introduction to Search and Rescue Optimal Planning System (SAROPS), a Monte Carlo-based software tool used in USCG maritime search planning. By mapping the experiences (Gibbons, 2017) and processes of the people conducting SAR, we learned how SAR is performed, from the command center receiving the initial report of a missing person, to the coordinated response of assets, to the long case documentation process and eventual closure or suspension.

To gain a comprehensive understanding of the command centers, we conducted interviews (Neilsen, 2010), observed command center operations (Neilsen, 2024), and conducted contextual inquiries (Salazar, 2020), focusing not just on what people say, but on what they do (Hall, 2013). We also facilitated a "SAR of the Future" ideation post-up session (Gibbons, 2020), where each MIT LL and USCG participant generated at least 10 distinct ideas on how AI or automation could enhance future SAR operations. We then used affinity diagramming (Gibbons, 2020) to group ideas by topic, helping us identify themes of interest. Finally, we applied forced ranking prioritization (Gibbons, 2020), where each USCG participant voted on the topic clusters, selecting one topic they were most interested in and one topic they thought would be most impactful to the USCG.

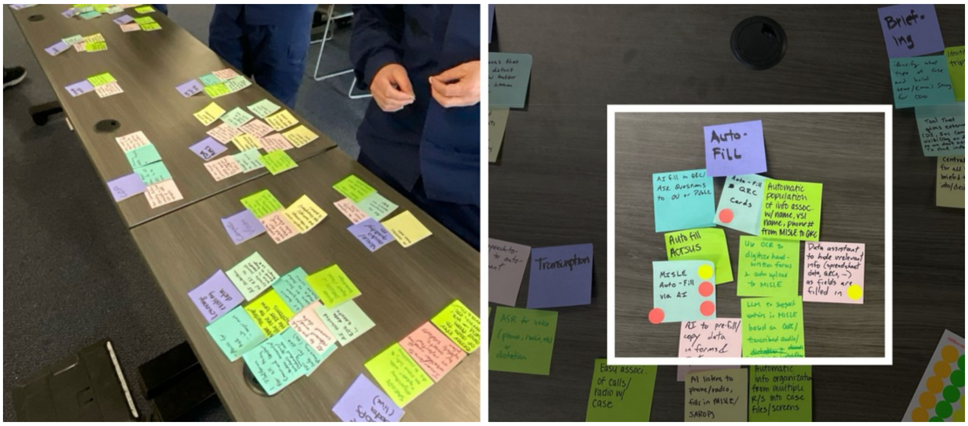

Figure 2: Documentation of "SAR of the Future" ideation post-up, affinity diagramming, and voting, showing the top-voted category of data entry automation.

As the command center did not have any incoming SAR cases during our visit, we observed watchstanders collaborate as they responded to a simulated SAR case. We interviewed key members of the command center, including the on-duty Command Duty Officer (CDO) – the most senior person on watch, responsible for overseeing mission execution; the Situation Unit (SU) – the person responsible for maintaining maritime domain awareness; the Operations Unit (OU) – the person responsible for coordinating assets and communicating with callers in distress; and the Communications watchstander – the person responsible for listening and responding to multiple channels of radio communication within the command center.

Guided by our research to that point, during subsequent field visits we utilized another collaborative post-up activity to elicit information about specific and desirable use cases within the realm of data entry automation. Each individual was asked to generate specific ways in which AI, automation, or technology might be used in a SAR mission within the data entry automation domain. In addition, individuals were asked to specify what data sources might be needed to accomplish each task, how one would verify the accuracy of each completed task, and accuracy, ethical, or legal concerns with automating the task. This information would give us insight into implementation, ethics, or legal requirements associated with the proposed task being automated.

In our brainstorming activities, to avoid stifling creativity, individual expression, and ideas that could be built upon, we explicitly welcomed ideas that did not require AI. Following this, we shortlisted the use cases that required AI, were feasible for us to implement and test, and that we believed would have a high impact. To further elicit implementation, ethics, and explainability requirements, we created three scenarios based on shortlisted use cases and asked command center watchstanders to storyboard them. In three groups, watchstanders storyboarded three different scenarios: 1) AI creates narratives and timelines, 2) AI analyzes an incoming call and prompts watchstanders with questions to ask, and 3) AI populates vessel information and property outcomes. Watchstanders storyboarded each scenario in two formats: one in which AI performs the task well, and one in which AI performs the task poorly.

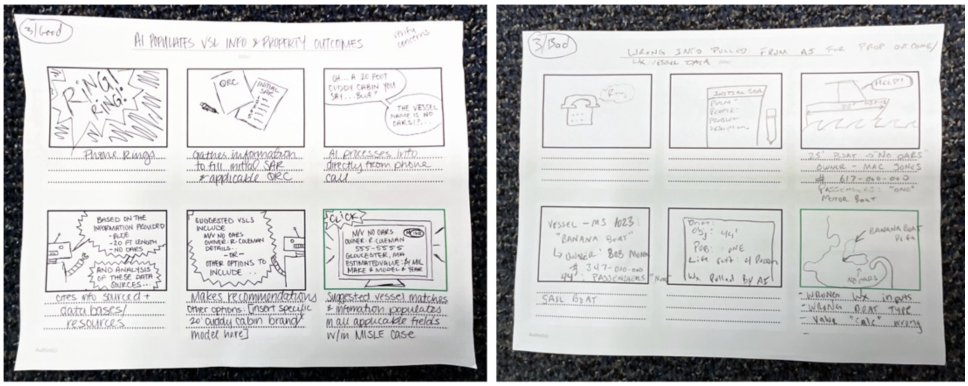

Figure 3: Documentation of storyboarding use case "AI populates vessel information and property outcomes" – with a successful AI story on the left, and an unsuccessful AI implementation on the right.

Using the storyboarded scenarios and the information elicited about implementation, ethics, and explainability, MIT LL created prototypes that combined multiple storyboarded features to illustrate how AI and XAI features might be utilized in the command center to facilitate the required time-intensive documentation processes. Prototypes built using Axure RP, a UX prototyping tool, were informed by existing explainability methods for text summarization (Norkute et al., 2021) and utilized best practices for designing human-centered AI experiences, such as maximizing user control, allowing users to provide feedback, and utilizing data sources within explanations (Google PAIR, 2021; Kore, 2022).

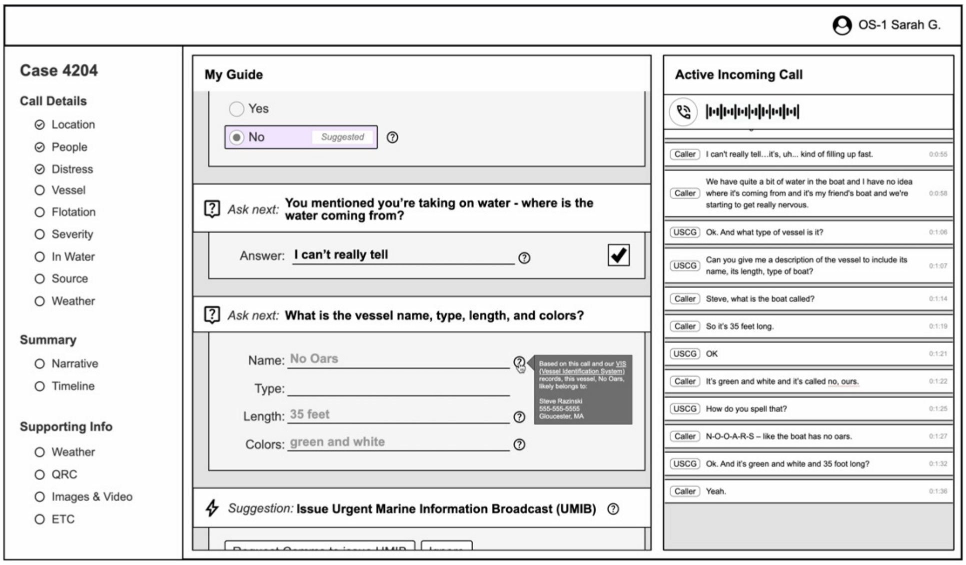

Figure 4: Screenshot of a digital prototype shows how a command center watchstander answering a SAR radio call can receive tailored follow-up questions, suggested answers (which can be accepted or corrected), and information about how suggested answers were generated from the audio call source, as well as data from other connected systems, such as the USCG's vessel information system (VIS) or weather sources; note the underlining in the rightmost pane indicating relatively uncertain speech-to-text translations ("no, ours" instead of the correct "No Oars").

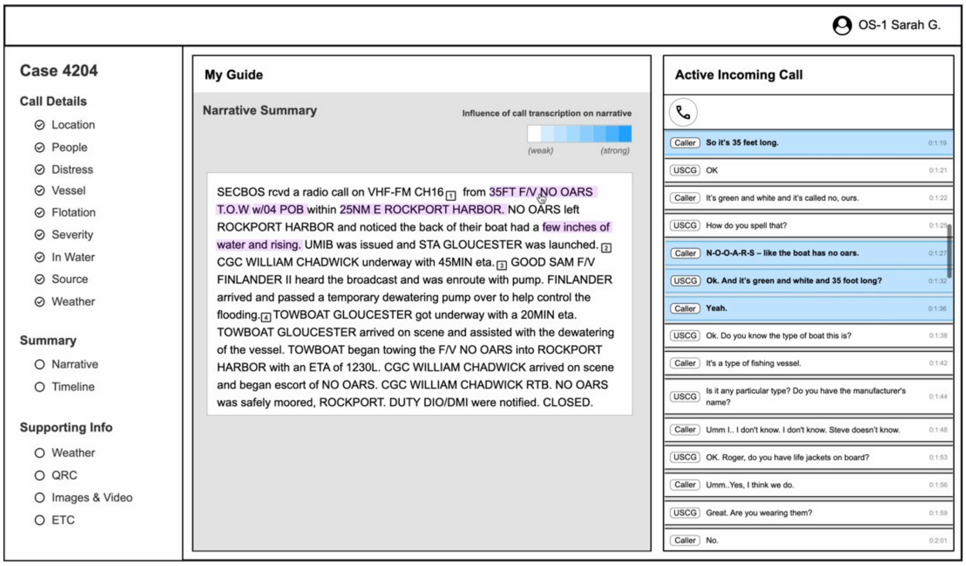

Figure 5: Screenshot of a digital prototype displays how a call narrative (one of the many time-consuming inputs into a USCG documentation tool) can be generated from transcribed command center calls, highlighting in blue the key excerpts that influenced the generated text "35FT F/V NO OARS".

RESULTS

This section details findings from our user research, focusing on watchstander ideas for technology improvement, concerns about unsuccessful implementation, and needs for explainability.

User Interests

Collaborative post-ups contained ideas to improve USCG decision making by leveraging existing data, generating courses of action, recommending assets to launch, updating existing tools with live data, improving radio call quality, automating data entry into multiple systems, asking a chatbot questions about doctrine and assets, and assisting with the briefing process. The top-voted category – automating data entry – contained ideas to automatically populate QRCs (Quick Response Checksheets), MISLE (Marine Information for Safety and Law Enforcement), and other systems; automatically populate case information, such as name, vessel name, and phone number from MISLE to QRC; suggest data entry in MISLE based on QRCs and transcribed audio; digitize handwritten forms and auto-upload to MISLE; and populate MISLE and SAROPS with data synthesized from phone and radio calls. Because MIT LL prioritized participatory design and human-centered principles, we focused on a category of use cases that was highly desirable to our end users – namely, automating data entry. Through the ideas generated, as well as interviews, observation, and contextual inquiry, we discovered that data entry processes were time consuming and tedious, and that AI would be welcomed in this domain. In effect, the watchstanders had identified a subset of their duties for which AI-enabled, XAI-explained human-machine teaming would be highly desirable.

User Concerns

Interviews, post-ups, and storyboards revealed user concerns about caller privacy, overall safety of AI when used in "life or death" situations such as SAR, having to spend extra time correcting AI-generated content, and potential liability if the AI failed to perform as expected. These concerns highlight the need for explainability in AI systems, as well as the need for continuous collaboration with end users to ensure that solutions are mindful of their concerns.

Key Takeaways

Overall, the prototypes based on successful AI implementation storyboards received very positive feedback. Watchstanders expressed high interest in using the prototyped features, if developed, and expressed positive feedback regarding some of the explainability features: "I really like how you can double check if the information generated is accurate. You can go directly to the source in the transcript. … I like that the radio buttons say 'suggested.' It gives me confidence that I can verify the options."

Suggestions for improvement included removing the call transcription during an incoming call to minimize distraction and bias, reformatting question hierarchy, pulling more information from other systems, showing additional formats for latitude/longitude conversion, and incorporating i911, a platform that allows mariners to share their location with USCG.

Further immersive scenario testing with real data would be needed to identify necessary and desirable functionality changes.

CONCLUSION

As successful adoption of new XAI tools necessitates designing "with" and not just "for," this paper summarizes the application of human-centered, participatory design in partnership with USCG watchstanders. Our work included collaborative generation of desirable use cases, elicitation of explainability requirements, and creation of wireframe prototypes to illustrate the utility of applying AI and XAI techniques to a common, time-consuming, high-consequence event: receiving and responding to a SAR call at a USCG command center. Iterations of prototypes yielded a proposed system that the participants judged to be of potentially high operational value. Importantly, the prototypes contained several explainability features to help watchstanders determine whether or not AI models are working as expected. Future work would include creating functional software prototypes to test the effectiveness of the proposed explainability features.

ACKNOWLEDGMENT

The research was conducted with the U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T) under contract 70RSAT23KPM000040. The DHS S&T sponsor team, led by Dr. Farrokh Vatan and including Dr. Sean Warnick, Dr. Krishna Vemuru, and Amidu Kamara, provided valuable guidance throughout this effort. This work would not have been possible without the ongoing collaboration with our partners at USCG Sector Boston.

DISTRIBUTION STATEMENT

REFERENCES

Auernhammer, J. (2020). Human-centered AI: The role of human-centered design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), Synergy - DRS International Conference 2020. [Conference held online, August 11–14]. https://doi.org/10.21606/drs.2020.282

Bravo, E. (1993). The hazards of leaving out the users. In S. Schuler & A. Namioka (Eds.), Participatory design: Principles and practices (pp. 3–12). Lawrence Erlbaum Associates.

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

Department of Homeland Security. (2024). Artificial intelligence roadmap. Department of Homeland Security. https://www.dhs.gov/sites/default/files/2024-03/24_0315_ocio_roadmap_artificialintelligence-ciov3-signed-508.pdf

Dwork, C., & Minow, M. (2022). Distrust of artificial intelligence: Sources and responses from computer science and law. Daedalus, 151(2), 309–321. https://doi.org/10.1162/daed_a_01918

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., Riener, A., & Riedl, M. O. (2021). Operationalizing human-centered perspectives in explainable AI. In Extended abstracts of the 2021 CHI conference on human factors in computing systems (CHI EA '21) (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3411763.3441342

Falkson, S., Greaves, A., Haque, A., Lapadula, A., Liu, J., Nam, K. Explainable AI (XAI) for DHS: Year 1 Report. MIT Lincoln Laboratory project report. In preparation.

Fessenden, T. (2021). User research for nonexistent projects [Video]. Nielsen Norman Group. https://www.nngroup.com/videos/user-research-non-existent-products/

Gibbons, S. (2017, November 5). UX mapping cheat sheet. Nielsen Norman Group. https://www.nngroup.com/articles/ux-mapping-cheat-sheet/

Gibbons, S. (2020, March 1). Foundational UX workshop activities. Nielsen Norman Group. https://www.nngroup.com/articles/workshop-activities/

Google PAIR. (2021). People + AI guidebook. Google. https://pair.withgoogle.com/guidebook/

Hall, E. (2013). Just enough research. A Book Apart.

Haque, A. K. M. B., Islam, A. K. M. N., Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Technological Forecasting and Social Change, 186 (Part A), 122120. https://doi.org/10.1016/j.techfore.2022.122120

Kore, A. (2022). Designing human-centric AI experiences: Applied UX Design for artificial intelligence. Apress.

Meske, C., Bunde, E., Schneider, J., Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. Information Systems Management, 39(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Neilsen, J. (2010, July 25). Interviewing users. Nielsen Norman Group. https://www.nngroup.com/articles/interviewing-users/

Neilsen, J. (2024, February 7). Field studies done right: Fast and observational. Nielsen Norman Group. https://www.nngroup.com/articles/field-studies-done-right-fast-and-observational/

Norkute, M., Herger, N., Michalak, L., Mulder, A., Gao, S. (2021). Towards Explainable AI: Assessing the usefulness and impact of added explainability features in legal document summarization. In Extended Abstracts of the 2021 CHI conference on Human Factors in Computing Systems (CHI EA '21) (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3411763.3443441

Norman, D. (2020). Principles of human-centered design [Video]. Nielsen Norman Group. https://www.nngroup.com/videos/principles-human-centered-design-don-norman/

Ryseff, J., De Bruhl, B., Newberry, S. J. (2024). The root causes of failure for artificial intelligence projects and how they can succeed (RR-A2680-1). RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2680-1.html

Salazar, K. (2020, December 6). Contextual inquiry: Inspire design by observing and interviewing users in their context. Nielsen Norman Group. https://www.nngroup.com/articles/contextual-inquiry/

Shperber, G. (2017). What I've learned from Kaggle's fisheries competition. Medium. https://gidishperber.medium.com/what-ive-learned-from-kaggle-s-fisheries-competition-92342f9ca779

Suresh, H., Gomez, S. R., Nam, K. K., Satyanarayan, A. (2021). Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. In Proceedings of the 2021 CHI conference on Human Factors in Computing Systems (CHI '21) (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445088

White House. (2023, October 30). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. The White House. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Wilson, K. A., Wong, J. C. S., Landree, E. (2024, May 2). How the DHS AI roadmap could reach its intended destination. RAND Corporation. https://www.rand.org/pubs/commentary/2024/05/how-the-dhs-ai-roadmap-could-reach-its-intended-destination.html

Yampolskiy, R. V. (2019). Predicting future AI failures from historic examples. Foresight, 21(1), 138–152. https://doi.org/10.1108/FS-04-2018-0034