5-2022

Human-Centered Explainable Artificial Intelligence for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI

Srikanth Vemula
University of the Incarnate Word, vemula@uiwtx.edu

HUMAN-CENTERED EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR ANOMALY DETECTION IN QUALITY INSPECTION: A COLLABORATIVE APPROACH TO BRIDGE THE GAP BETWEEN HUMANS AND AI

by

SRIKANTH VEMULA

A DISSERTATION

Presented to the Faculty of the University of the Incarnate Word
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF THE INCARNATE WORD

May 2022

ACKNOWLEDGMENTS

I cannot express enough gratitude to my committee for their continued support and encouragement: Dr. Michael Frye, and Dr. Lucretia Fraga, my committee chair, Dr. Alison Buck in completing this research. I offer my sincere appreciation for the learning opportunities provided through out my PhD journey by my committee and will always be grateful for their constant support.

My completion of this project could not have been accomplished without the support of AVS Labs, CPS Energy, and DOD Army Research Grant. My deepest gratitude to my parents Mr. and Mrs. Murali Krishna Vemula, and my brother Avinash Vemula for keeping in check and being my motivation throughout in conducting this research and finishing this degree.

Finally, to my dissertation group, my classmates, friends, and other professors who helped me to explore the paths that I never travelled before. My heartfelt thanks.

Srikanth Vemula

DEDICATION

I would like to dedicate this work to three important people in my life. The first one is my mother, A strong and gentle soul whose unconditional love, encouragement, and prayers, teaches me how to be good to others and humble. The second, my dad, who taught me how to stop worrying about the uncontrollable things and focus on how to use what you have and make the best out of it. Lastly my little brother, who is my strength, without him I wouldn’t able to finish this research and I will always grateful to have a brother like him.

HUMAN-CENTERED EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR ANOMALY DETECTION IN QUALITY INSPECTION: A COLLABORATIVE APPROACH TO BRIDGE THE GAP BETWEEN HUMANS AND AI

Srikanth Vemula

University of the Incarnate Word, 2022

Consumption of electricity is becoming more significant and is an important part of our present-day society, which raises major difficulties in terms of maintaining power supply stability, affordability, and sustainability. In the quality inspection industry's use of AI, applications continue to advance to produce safer and faster autonomous systems that can perceive, learn, decide, and act independently. However, these AI systems' performance is limited by the machine's current inability to explain its decisions and actions to human users. Especially in energy companies, Explainable-AI (XAI) is poised to achieve fast reliability, explainability, and trustworthiness, which is currently lacking. Placing humans alongside XAI will establish a sense of trust that augments the individual's capabilities at the workplace. To achieve such an XAI system centered around humans, it is necessary to design and develop more explainable models. Incorporating this XAI system centered around human workers in the inspection industry is significant for the emerging generation of AI intelligent inspection systems that make the decision-making process more sustainable and trustworthy. In identifying the significance of and need for having explainable AI models centered around humans for quality inspection, there is a lack of trust between the inspection workers and AI, which creates uncertainty in using existing AI models by the inspection workers that are being developed. To address this gap, this

qualitative research study aimed to explore and understand the need for these human-centered XAI systems in energy industries in detecting anomalies in quality inspection.

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

TABLE OF CONTENTS—Continued

TABLE OF CONTENTS—Continued

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF FIGURES—Continued

Chapter 1: Background

Electricity plays a significant role in contemporary society in almost every aspect of human daily life nowadays. This is one of the reasons why it has become a highly dominant and dependable resource on earth. This raises major concerns about how the power supply can be sustained, bringing stability, affordability, and sustainability to the resources of the energy companies. For example, in Europe and the United States, a lack of incentives to invest in aging national power grid infrastructure is triggering a rise in power outages (Nguyen et al., 2018). Both short and long-term, these power outages can be detrimental to unprepared utility companies and inflict major financial damage on energy suppliers, manufacturers, and customers.

To avoid such kinds of catastrophic outcomes, electric utilities are usually expected to conduct regular visual inspections (Katrasnik et al., 2009) of their electrical grids to avoid power outages and ensure a safe and stable energy supply. These inspections are tedious, time-consuming, and expensive, and yet are vital steps to be performed by electric companies. It is impossible to safely run a transmission and distribution network if damage checks or risk evaluations are overlooked. Typically, these inspections are carried out using a combination of airborne surveys via low-fly helicopters, and field surveys via foot patrol and tower climb. Infield surveys are carried out by a team of two inspectors, who walk from pylon to pylon and visually inspect the powerlines using binoculars, sometimes with infrared and corona detection cameras, and cover a short range of inspection. In airborne surveys, the inspection is typically conducted by a pilot team and a camera operator. The pilot flies the helicopter over the power lines while the camera operator takes pictures (Katrasnik et al., 2009). During this inspection process, many utility companies and contractors take pictures of potential defects and anomalies, while others take pictures of the whole power grid, which includes pictures of conductors and

other powerline components (e.g., insulators, poles, cross arms, and transformers) and surrounding objects (e.g., vegetation encroachment). These images are manually inspected one by one to identify potential damages, and to determine if there is any action that needs to be taken. As this whole process is carried out for large areas, it can take an enormous amount of time to find defects in the power grid or poles within a county or city, and is expensive, tedious, and risky, which affects the safety of the inspection workers as well.

Context of the Study

Lately, in order to address the issues around these traditional methods of inspections, Unmanned Aerial Vehicles (UAV) have become a viable option in terms of cost and data processing. In combination with Machine Learning (ML), Deep Learning (DL), and Computer Vision (CV) technologies, alongside UAVs, utility companies are focusing on looking into this approach. This research approach of converging Artificial Intelligence (AI) equipped UAVs in a civilian application brings much flexibility to enhancing human operators' ability to decide quickly and fix problems in time, which avoids delays and shortages. One other advantage of using UAVs is that the capability of equipment with advanced camera payloads provides an ability to conduct aerial inspection with greater accuracy, effectively making the process itself less tedious, less expensive, and much faster.

Due to the above advantages when using UAVs, the inspection industry overall found it attractive to try this approach, as it overcomes the difficulties that are involved with conducting inspections, but the new approach also enhances the range and productivity of inspections by boosting the coverage, volume, and quality of the data capture process. In doing so, there are several instances in which different techniques were developed to address these issues by using DL and CV in combination with UAVs. One such instance would be using the instance

segmentation method. In a recent paper by Vemula and Frye (2020a), a method was proposed that identifies the powerline components, and segments out each component in real-time while flying, which shows the use of AI and UAVs. Vemula et al. (2021) suggest a novel approach. A heterogeneous system, consisting of two autonomous systems, one UAV and one Unmanned Ground Vehicles (UGV), works collaboratively alongside humans to conduct powerline inspection. The same researchers published another DL method for conducting powerline inspection to make decisions based on the trained model to mask out the individual components of the powerline with a complete workflow in conducting inspections (Vemula & Frye, 2020b). Several other research studies mainly focus on developing new approaches or methods, and all these developed AI models are centered around the Blackbox approach. This means that the AI model does not explain why the model made such a prediction that creates an amount of uncertainty. The developer who is involved in the development of the model does not know why that decision was made and cannot assure when the user can trust the AI model. This huge gap in terms of the model explainability, due to which trust issues arise between the human and the AI, is the core of this research study. In the inspection industry, it is imperative to establish a sense of trust between the AI models that are developed with the inspection workers, and how it can bridge the trust within by establishing a collaborative approach. For instance, one scenario can be when a model detects the anomalies in the crossarm component of a powerline, the explainable model should not only detect the anomaly but also provide an explanation that a human inspection worker would understand of why it made such decision. This type of approach plays a significant factor in developing AI models for powerline inspections by designing the systems around humans, which possess explainability and trust within those models. When humans can trust and rely on conducting the quality inspection, identifying anomalies if they exist and

keeping the power grid in good maintenance to avoid long and short-term shortages, their capabilities are augmented using explainable AI. With that said the following sections discuss the problem statement and the research questions that this research is investigating and focusing on, the literature review, the purpose of this research, and a proposed methodology to address these questions.

Problem Statement

As the quality inspection industry's use of AI, applications continues to advance, companies produce safer and faster autonomous systems that can perceive, learn, decide, and act independently. However, these AI systems' performance is limited by the machine's current inability to explain their decisions and actions to human users. Especially in energy companies, Explainable-AI (XAI) is imperative to achieve speed, reliability, explainability, and trustworthiness, which are currently lacking in the present models developed.

Purpose

Placing humans alongside XAI will establish a sense of trust that augments the individual's capabilities at the workplace. In order to achieve such an XAI system centered around humans, designing and developing more explainable models is necessary.

Significance

Incorporating an XAI system centered around human workers in the inspection industry is significant for the emerging generation of AI intelligent inspection systems that make decision-making processes more sustainable. When I worked closely with the electric company, I observed distrust and a lack of explainability in the AI models. I also observed that the developers who created those AI models knew how the model was trained and created, but could not explain why the model was behaving in a certain way after it was trained. This observed

phenomenon led me to investigate how these factors could be addressed by placing humans at the center when developing AI models for conducting inspections. This idea of inculcating a human-centered vision of innovation in the electric companies will open a new approach to augmenting human capabilities during inspections, and will promote co-development, co-existence, and a sense of explainability between AI and humans. One major challenge would be designing such human-centered XAI systems and diffusing humans and AI as a collaborative system that helps build trust and explainability during inspections.

A qualitative research study was used to explore and understand the need for these human-centered XAI systems in energy industries to detect anomalies in quality inspection by providing effective and trustworthy experiences between the AI and inspection workers. For this study, a modified framework for innovation was used to answer the research questions that are posed in three stages. Stage 1 and Stage 2 were focused on answering research question 1. Stage 3 was focused on answering research question 1a.

Research Questions

- Research Question 1 (RQ1): How might we design a Human-Centered XAI (HC-XAI) system that augments human capabilities in conducting visual inspection for identifying anomalies?
- Research Question 1a (RQ1a): How might this HC-XAI design foster social innovation and sustainability through this shared and collaborative approach?

Overview of Methodology

To answer the research questions, a modified framework for innovation was based on a double diamond methodology, and the research process was divided into three stages.

Stage 1: Discover Synthesis Phase

Stage 1 comprised the discover phase, where participant selection, data collection, data analysis, and validating the collected data was carried out. Three individuals from three companies were selected, two of whom were directly involved in advocating the incorporation of human-centered design into AI. The last company was City Public Service (CPS) Energy, where I wanted to know more about the difficulties that are involved in conducting the inspection process. Data was collected using interviews and observations. Computational grounded theory was used to analyze the data and obtain the categories. Affinity mapping was used to organize the categories and shuffle them to form themes, visually organize thoughts into groups and shuffle the ideas in order to analyze the data collected. Once the data was analyzed, member checking was used to validate the data by returning to the participants to check for accuracy and resonance with their experiences. By the end of this stage, a better understanding of the problem statement was achieved and the study was ready for the second stage.

Stage 2.1: Redefine Phase

In this stage, the problem statement was redefined based on the findings collected from the first stage. This redefined problem statement included the research insights, intended audience, and the pain points discovered during initial research, and define what factors need to be investigated and focus on in developing the HC-XAI system in the prototyping phase in Stage 2.2.

Stage 2.2: Prototyping Phase

In this stage, machine learning and computer vision technologies were used to construct and develop explainable AI models, focused on creating this prototype of an HC-XAI system for quality inspection. Before doing that, a feasibility mapping was generated to determine what could be achieved and what could not be achieved within the scope of the time frame. Once this

was established, a prototype was developed and made ready for the implementation and evaluation phase in Stage 3.

Stage 3: Implementation and Evaluation Phase

In this stage, usability testing was used to validate the prototype. Specifically, Concurrent Think Aloud (CTA) moderating techniques were used that helped to understand the thoughts of the participants. For this validation, research planning meetings were conducted with the participants where informed consent forms were given to participants and their observation and feedback were collected and recorded in validating the system. Where there was a need to incorporate some of the missing elements that these participants thought were necessary, those were incorporated by going back to the development phase. Once everyone was satisfied with the system, feedback was taken from participants who were involved in testing the system and assessing whether social innovation could be fostered through this system. This is how both research questions were answered in this research study.

Nomenclature

- AI - Artificial or Augmented Intelligence
- CV - Computer Vision
- DL - Deep Learning
- HCD - Human-Centered Design
- ML - Machine Learning
- XAI - eXplainable AI
- HC-XAI - Human-Centered eXplainable AI
- UAV - Unmanned Aerial Vehicles Participant Selection

In Stage 1, three participants were used whose experience was focused on the topics

related to the use of human-centered AI for solving socially challenging problems, advocating human-centered AI design and its importance. One participant possessed experience in handling and overseeing responsibilities in conducting powerline inspections. These three participants were recruited through email and in Stage 3, were used for testing and obtaining feedback on the HC-XAI system and how this system might augment human capabilities during inspection and identifying anomalies.

Precautionary Measures

There was minimal risk involved in this study, since the testing of the system took place by video that was recorded during inspection flights that were carried out in the real-world environment. During these flights, precautionary measures were taken, by wearing hard hats and carrying walkie talkies to avoid crashes by the workers at the training yard. A Non-disclosure Agreement (NDA) was signed by both parties, to make sure the data obtained and used in this research was secured.

Delimitations

AI is an ever-changing subject, and the associated human-centered design strategies with it are bound to evolve. Hence the solution to be designed considered current AI capabilities and human-centered design practices only. Along with that, this newly developed HC-XAI system was designed as a general model with respect to local electric company inspection procedures.

Conclusion

This chapter provides a high-level overview of how the research was carried out by touching on all the elements that were involved, such as background, significance, problem statement, research questions, and how the research was structured based on the research design, participant selection, and possible limitations that occur. In the coming chapters, a literature

review will help inform the research study, the detailed research design and methodology will be discussed, the detailed analysis will be presented, and detailed discussion and conclusion about the findings will be provided.

Chapter 2: Literature Review

Energy companies' use of AI in quality inspection continues to advance. However, many of these methods are being implemented using the Blackbox approach, and the AI lacks explainability and human-centeredness in those inspection methods. This led me to investigate posed research questions and investigate how the development of explainability and human centered AI can enhance and foster social innovation and sustainability. This section of the document presents research on general concepts of AI that are used in intelligent inspection systems, how AI is perceived through a technical and humanistic lens, and how HCAI has evolved. It will also explore the significance of explainable AI in inspection systems and how AI and sustainability are an important aspect to consider for the development of this HC-XAI system.

General Concepts of Artificial Intelligence

Artificial Intelligence (AI) was first used at the Dartmouth Conference in 1956 by a famous American computer scientist, John McCarthy. Although AI announced its arrival in the 1950s, it was not until recent times that it has become a household name and is being used by every individual, knowingly or unknowingly. As AI deals with mimicking cognitive functions for real-world problem solving, it helps researchers and developers build systems that learn and think like humans. Poole et al. (1998) termed this ability to possess such intelligence by a machine as Machine Intelligence. In contrast with human intelligence (Russell et al., 2010), this field revolves around cognitive science and computer science (Tenenbaum et al., 2011). Because of the shift caused by the COVID-19 pandemic, and the practical successes in Machine Learning (ML) and Deep Learning (DL) applications in recent times, people are looking for innovative ways to use AI in various industries, as a result of which AI now has huge interest in these

industries. On the flip side, in AI, there is always a strong connectedness to explainability; McCarthy (1960) proposed an early example in 1958, the advice taker "program with common sense" (p. 20). This was probably the first time AI developers brought up common sense reasoning abilities as AI's critical element. The latest AI developments have been increasingly used for many applications, and in daily lives for problem-solving using these AI models. According to Lake et al., (2017), more and more AI systems and their models should support explanation and understanding rather than just solving pattern recognition problems.

Different Machine Learning Approaches

ML is a field of AI that is used widely in a practical perspective in developing AI systems. According to Michalski et al. (1984), machines can learn automatically, based on previous data, to gain insights and knowledge that improve its learning behavior and ability to make predictions based on the new data. It faces challenges in terms of sensemaking in understanding the context given to it and making decisions under uncertain conditions (Holzinger, 2019). For these reasons, ML can be seen as a workhorse of AI. Its applications are being seen almost everywhere, throughout science, education, engineering, and business, which leads to more evidence-based decision-making, and makes life easier (Jordan & Mitchell, 2015). According to Abadi et al. (2016), due to the availability of large datasets and low-cost computation, there has been massive progress in ML developments. A machine can learn three different approaches that can be implemented in a real-world application based on the nature of the data and the problem at hand.

Supervised Learning

In this approach, the model is provided with lots of data that has been labeled, and trains the machine based on the data provided. The ML algorithm is designed so that it takes the input

collected and labeled to train the ML model to do a certain task. It is like teaching a child about a particular object and letting them learn over time to recognize that object in more nonprofessional terms. This is the process of training that happens in this approach of the ML algorithm. In this approach, the model trains itself to perform certain tasks. Some of the most-used algorithms are Classification and Regression.

Unsupervised Learning

Unlike the previous approach, the data fed to these algorithms are not labeled; instead, the machine looks for the patterns that it can find. This kind of approach is highly effective, especially when massive amounts of data are set, and humans cannot see a pattern. The most-used algorithms in these scenarios of unsupervised learning ML models are Clustering and Dimensionality Reduction.

Reinforcement Learning

This approach is quite the opposite of the two approaches discussed above. In this approach, no data is given to the algorithm; instead, the algorithm learns by itself using trial and error to achieve a clear objective. Alongside this, a reward system is implemented to penalize or reward, depending on the algorithm behavior that helps or hinders it from achieving the desired objective set. There are several examples of this kind of approach; one notable and popular example is Google's AlphaGo.

Those are some major concepts that give an overview of the concept of ML. Now let us dive into another advanced concept that is getting more popular these days—DL.

Powerline Inspection Using Deep Learning

DL is the ML family in which the models developed are using deep convolutional networks (Schmidhuber, 2015). Due to its capability of producing high-end results at human-

level performance, these are gaining high traction (LeCun et al., 2015). According to Vemula et al. (2020a), recent work by AVS Labs research was conducted on powerline inspections using DL, which classifies individual components of the power pole with a level of competence comparable to human perception. Another research work that came out of the same lab that leverages instance segmentation technique in detecting the individual components of powerline using Unmanned Aerial Vehicles (UAV's) (Vemula et al., 2020b). All these are well-illustrated examples of how well AI is lending its capabilities in real-world situations. These autonomous approaches emphasize that usable intelligence is difficult to achieve because there is a need for the model to learn from previous data in order to extract knowledge, generalize, and understand the context of which application domain the model operates (Bengio et al., 2009).

ML- and DL-based Intelligent Inspection Systems

During any industrial quality inspection, the detection of individual components for anomalies and maintenance is essential. Several researchers have studied the application of computer vision technologies for vision-based industrial inspection problems. Cusano and Napoletano (2017) have designed a visual recognition model for inspecting mechanical parts of an aircraft during its maintenance. Due to the increase in the use of deep neural networks, there are several cases that have used these deep convolutional neural networks (CNN) along with transfer learning to train the AI models in flower detection (Dias et al., 2018), and disease detection (Coulibaly et al., 2019), to name a few. Although there are many advances in other vision-related industrial detection tasks, like fire detection (Muhammad et al., 2019) and smoke detection (Filonenko et al., 2017), the detection subjects are amorphous when compared to the solid objects that are the focus of this dissertation. The anomaly detection system process mainly consists of two tasks: key component detection and detection of anomalies (Kang et al., 2018).

The first task's purpose is to localize and extract the target object features from the images with a complex background. In contrast, the second task focuses on identifying the anomalies and the exact location or positions of those components (Zuo et al., 2017). The inspection tasks vary based on different components that include insulator explosion (Gao et al., 2017; Nguyen et al., 2018; Yan et al., 2019; Yang, Huang et al., 2019), insulator missing (Adou et al., 2019; Nguyen et al., 2018), insulator swing angle (Gu et al., 2009b; Yang, Wang et al., 2019; Zhao et al., 2014), and snow and ice coverage (Gu et al., 2009a; Gu et al.,2009b). For intelligent inspection systems, these tasks of detection and localization of target objects are important and necessary (Zhao et al., 2012). In the electric power industry, insulators are crucial, and the faults that occur in these insulators lead to serious problems in power transmission systems (Park et al., 2017). Usually, regular maintenance and detecting anomalies are carried out using either walking patrol or helicopter assistance, which brings great risk to the lineman's safety (Prates et al., 2019). For these reasons, fully automatic autonomous vision-based inspection systems have received more attention in the electric power industry.

Different Perspectives of AI—Technological vs Humanistic

Since 1950, when Alan Turing proposed the Turing-Test (Crevier, 1993; Grudin, 2009), intelligent systems have evolved. During this evolution process of AI, two distinctive philosophical perspectives have emerged in how human-computer interactions are carried out (Grudin, 2009; Winograd, 1996). The philosophy that views AI from a technological perspective falls under "rationalistic," and the philosophy that views AI from a humanistic perspective falls under "design" (Winograd, 1996). These philosophical perspectives between science and the humanities have been going on for a little over a decade, even beyond AI development. During the early development of AI, in this philosophical divide, the technological ("rationalistic")

perspective was represented by John McCarthy, and the humanistic ("design") perspective was represented by Douglas Engelbart (Markoff, 2005; Winograd, 1996). Let us look in brief at what each perspective cares about and examine them through each of these lenses.

From a technological perspective, the term AI is surrounded by the theory and development of computer systems capable of mimicking human abilities and doing tasks that require human intelligence. The research related to this perspective focuses on mathematical and technological advancements like neural networks, statistical language, and ML, to create adaptive system mechanisms. Moreover, most humans are seen as "cognitive machines" (Winograd, 1996; Winograd et. al., 1986). In the humanistic perspective, AI research is mainly centered around a problem-solving tool, to advance human capabilities and improve their current conditions. This humanistic perspective mainly focuses on the interaction or involvement of humans with computers (Winograd, 1996; Winograd et. al., 1986) and sees human thought and physical embodiment as inseparable (Dreyfus, 1992). The main advantage of this perspective is that it allows us to align and cope with real-world complexities and human situations (Rittel et. al., 1973) and it has a unique approach or strength during the interactions of humans with the AI system. These two main design research areas relate to these perspectives, illustrated in Figure 1.

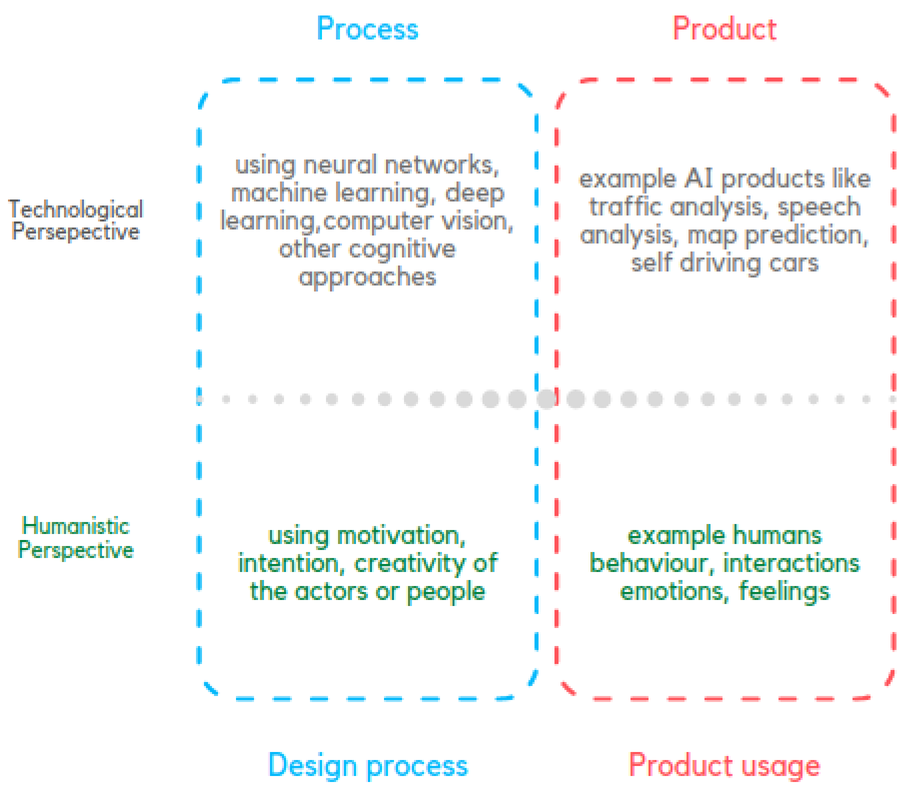
Evolution of Human-Centered Artificial Intelligence (HCAI)

HCAI is an effort that was started in recent years to bring the two most significant research fields together, Human-Computer Interaction (HCI) and AI. This emergence aims to place humans as the center in the design and development of the technologies that involve HCI and AI, which are intended to help humankind rather than pose a threat. According to Xu (2019), AI's role in the community is not to replace humans. Instead, its role is to augment human

capabilities to enhance their skills and increase productivity, which helps them make informed decisions.

Figure 1

Spectrum Illustration of Technological Perspective and Humanistic Perspective.



Unfortunately, most of the time the AI systems that were developed were projected as a threat to humans and created a false impression that these systems were going to replace humans, rather than augmenting humans, which is the main intent of these developments (the fourth industrial revolution). This idea of AI augmenting humans instead of replacing them is one of the key objectives behind the development of HCAI. Xu (2019) mentioned how important it is for HCAI solutions to be ethical, explainable, comprehensible, and useful (Xu, 2019). This study investigated how to integrate these values and propose a working framework that includes three factors in designing and developing HCAI-based technological solutions. There is a huge shift in

technology advancement, and how things are rapidly changing around the world, especially with AI. It is time for HCI and UX researchers to investigate the challenges related to human-AI interactions and address the methods and usability of these solutions. Most of the currently existing HCI and UX methods and usability solutions developed are not designed for AI systems. Amershi et al. (2019) provided a set of Human-AI Interaction Guidelines, given that humans are increasingly depending on and engaging with these AI systems in making important decisions based on algorithms and data. UX researchers must play a key role and invest heavily in adding end-user values throughout the AI development life cycle. This idea of a human-centered approach is not new. There are many user-centered practices, design labs, and co-creation methodologies (Mulvenna et al., 2017) pioneered by UX researchers over the years. Integrating HCI and AI development will produce a multidisciplinary approach by involving HCI sub-disciplines, such as human factors, psychology, and design.

Significance of Explainability AI in Inspection Systems

Due to the popularity of AI, a wide area of research has been carried out around producing algorithms that are focused on determining intelligent inspection systems using computer vision technologies and AI. As these algorithms focus more on the novel algorithms in carrying out inspection, there is a lack of explainability about why the system is giving those results when implemented. There is no explanation involved in why the system behaves in the way it behaves, which produces a lack of trust in those using those systems. The term explainability is as old as AI itself. In AI, reasoning methods were logical and symbolic during its developing days (Newell et al., 1958), and these approaches were successful in terms of space and practicality. One such example is MYCIN, and an expert system was developed in Lisp for detecting bacteria that cause severe infections and then recommending antibiotics (Shortliffe &

Buchanan, 1975). Due to the significant effort involved in maintaining the knowledge base, it was never used in clinical routines. These early AI systems could perform tasks based on logical inference of human recognizable symbols and provide some traceable inference steps that formed the basis for an explanation, and there is some early work available about it (Johnson 1994; Lacave & Diez, 2002; Swartout et al., 1991).

Moreover, AI's explainability in an intelligent system would enhance the needed trust factor that is lacking between humans and AI systems, especially in visual inspections in energy companies. Research trends focus more on building explainable AI (XAI) systems over the past two years. The current AI systems developed using ML and DL techniques were built as Blackboxes, where there is no explanation about why the system makes such a prediction. Therefore, there is an inherent tension that is created between ML performance and explainability. Even the best-performing methods, like DL, are less transparent, and the ones that provide a clear explanation, like decision trees, are less accurate (Bologna & Hayashi, 2017). In the current scenario in terms of an AI model, it is difficult to find and explain why such predictions are made or how the model parameters capture the underlying features of the trained mechanisms. One other constraint that humans have is their limited ability to visually assess or review explanations for a substantial number of axioms. This leads to the question of whether it is possible to deduce properties based purely on observations (Peters et al., 2017).

In the context of XAI, understanding, interpreting, or explaining are interchangeably used (Doran et al., 2017). Several interpretation techniques were applied in the past. One notable discussion by Lipton (2018) is on the "myth of model interpretability." The term "understanding" in XAI usually refers to the functional understanding of the model but not the low-level algorithmic understanding that seeks to characterize the model's Blackbox behavior in

terms of learning without knowing how inner learning and representations are formed. To differentiate between explanation and interpretation, Montavon et al. (2018) defines interpretation as the mapping of an abstract idea into a realm that humans can perceive and comprehend. Simultaneously, an explanation is defined as a collection of the features of those interpretable realms that are guided to produce a decision in any given example.

If these kinds of XAI systems complement inspection professionals, that can play a huge role in augmenting the human's role in the inspection process, which leads to a safer and quicker decision-making process and builds better trust and explainability in a more human-centric approach. Sometimes it is assumed that humans always explain their decisions, but it is not often the case due to various heterogeneous and vast information sources. Hence XAI calls for confidence, safety, security, privacy, ethics, fairness, and trust (Kieseberg et al., 2016), which brings usability and Human-AI interaction into a new and much more important focus (Miller, 2019).

AI Towards Sustainable Development

As the use of AI applications is on the rise in many fields, from autonomous vehicles (Bonnefon et al., 2016) to AI-powered healthcare solutions (De Fauw et al., 2018) and smart electrical grids (IEA, 2017), it is becoming more important to investigate how AI can be trustworthy and safe to use in critical decision making. Research was needed to focus on keeping these systems more robust, explainable, trustworthy, and assisting or augmenting humans in performing tasks, including keeping them updated on adversities like getting hacked (Russell et al., 2015). Research that investigates the safe integration of AI helps to understand the catastrophes that a systemic fault can enable in AI technology. The World Economic Forum (2018) raises concern over integrating AI in the financial sector. Due to this, it is essential to

raise concerns over the risks associated with AI systems in a society. In addition, numerous studies suggest that AI can potentially act as an enabler for many sustainable development goals and indicators. However, a fraction of these studies were conducted in a controlled laboratory environment based on limited datasets and using custom datasets for developing AI prototypes (Cao et al., 2014; Esteva et al., 2017; Gandhi et al., 2017). It is always a challenging task to evaluate the models in real-world settings and measure AI's impact in broader scales, both temporarily and spatially. While conducting controlled experimental trials for evaluating the real-world impacts of an AI system can depict a snapshot situation, the AI system is constrained to a known environment, which is not the case with society as it constantly changes. The requirements for AI also change, resulting in a feedback loop with interactions between society and AI.

Another aspect that needs to be brought to light is the resilience of society towards AI-enabled changes. There is a need for these novel AI methodologies to incorporate various points of view, like efficiency, ethics, and sustainability, before large-scale AI system deployments. For these reasons, research is essential and should aim to find out the reasons for the failure of AI systems by introducing a human-machine analysis tool (Nushi et al., 2018), with the aim of developing an accountable AI, by designing AI in a more human-centric way and by maintaining accountability and explainability to humans.

How Literature Shaped This Study

From the above literature review, there are three primary takeaways. Those are that AI is growing faster and will be present everywhere, the need for designing an accountable and explainable AI is significant and bringing human-centeredness and explainability to the intelligent systems for quality inspection in energy companies will provide new way of doing things that fosters social innovation and sustainability. This dissertation investigates the problem

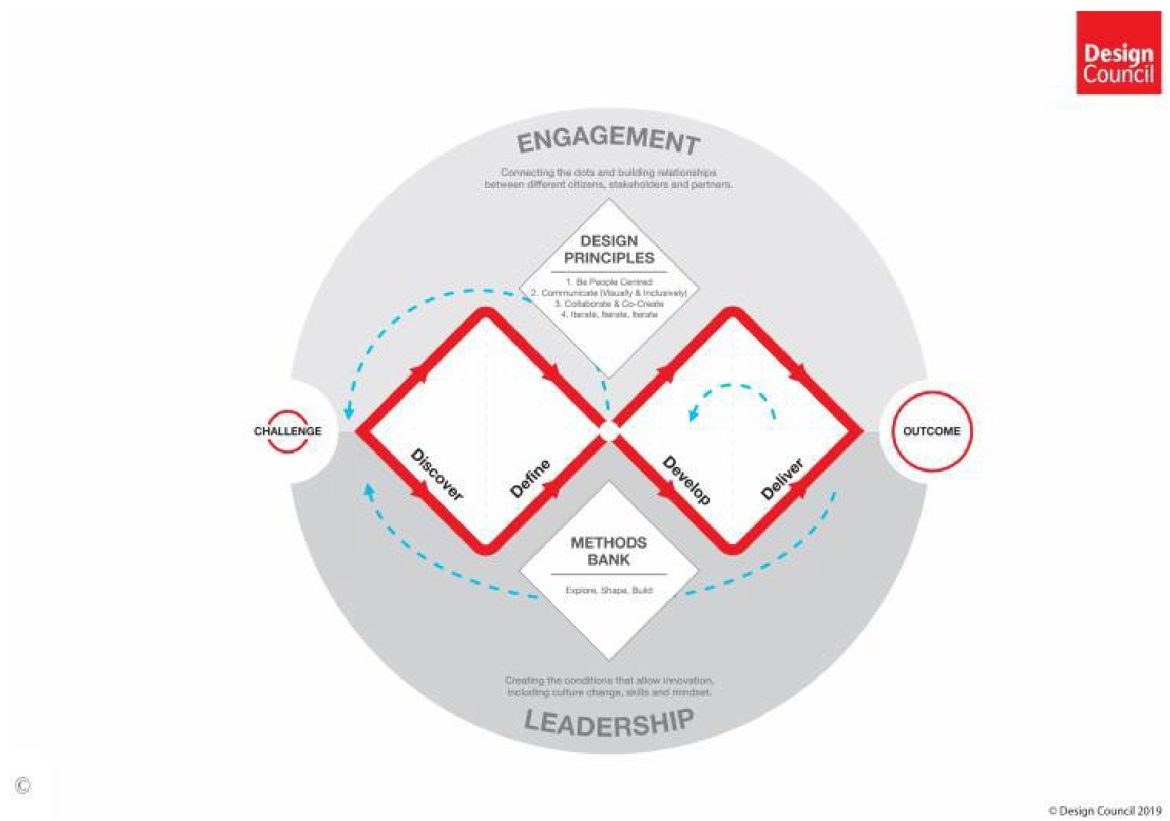
statement and investigates a way to answer the research questions. To obtain this goal, a proposed conceptual framework that guides in investigating the research questions is going to be discussed in the next section of this dissertation.

Chapter 3: Methodology

The methodology that inspired this research study is the framework for innovation that Design Council introduced in 2019. A modified version of this framework was used as a primary process to investigate the research questions. The core idea of this framework for innovation relies on the Design Council's double diamond methodology, where there are four phases: Discover, Define, Develop, and Deliver. Along with these, the framework for innovation includes the key principles and design methods that designers and non-designers need to use, and the working culture needed to achieve significant and long-lasting positive change (Figure 2).

Figure 2

Framework for Innovation



Note. Source: https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond

The core of this framework for innovation depends on the process of the double diamond methodology, which involves four phases:

- Discover: In this phase, it helps to understand what the actual problem is, rather than just assuming what it is. That involves speaking to key knowledge holders, spending time with affected groups, and learning more about the issues.
- Define: In this phase, the insights gathered from the discover phase can help me to define the challenge differently.
- Develop: In this phase, the second diamond is motivated to obtain different answers for the clearly defined problem by seeking inspiration from someone else and co- designing and developing from a different range of people.
- Delivery: This phase involves testing the solution on a small scale by rejecting those that do not work and improving the ones that do.

This double diamond process is not linear, as the dotted arrow in Figure 2 illustrates. Apart from these phases, four-core design principles exist in the framework for innovation, one of the main reasons this methodology was chosen for this study. Those are:

- Put people first: understanding the people using the service, their needs, strengths, and goals.
- Communicate visually and inclusively: create a shared understanding of the problem both for the people and the researcher.
- Collaborate and co-create: Work collaboratively and get inspired by what others are doing.
- Iterate, iterate, iterate: identifying errors and risks involved in the initial stages and iterate the prototyping process to build confidence in the ideas.

Along with those principles, this framework provided inspiration for this research study in the methods that it uses:

- Explore the challenges, needs, and opportunities that are involved during the process.
- Determine what shapes the prototypes, vision, and insights are.
- Build ideas, plans, and expertise towards solving the problem.

These are the three main reasons that this framework for innovation was modified for the current study, which involved designing an HC-XAI system that would help investigate the research questions and address the problem statement. The following sections provide a description of how this study was conducted by using the proposed modified framework for innovation to investigate the research questions.

To conduct this research study, a modified framework for innovation was proposed, which consisted of three stages: discover synthesis phase, redefine and prototyping phase, and implementation and evaluation phase. These three stages assisted the research study as a guiding process. Stages 1 and 2 (discover synthesis phase, redefine and prototyping phase) assisted in answering research question 1. Stage 3 (implementation and evaluation phase) was responsible for answering research question 1a. The proposed framework for this research study is presented in Figure 3.

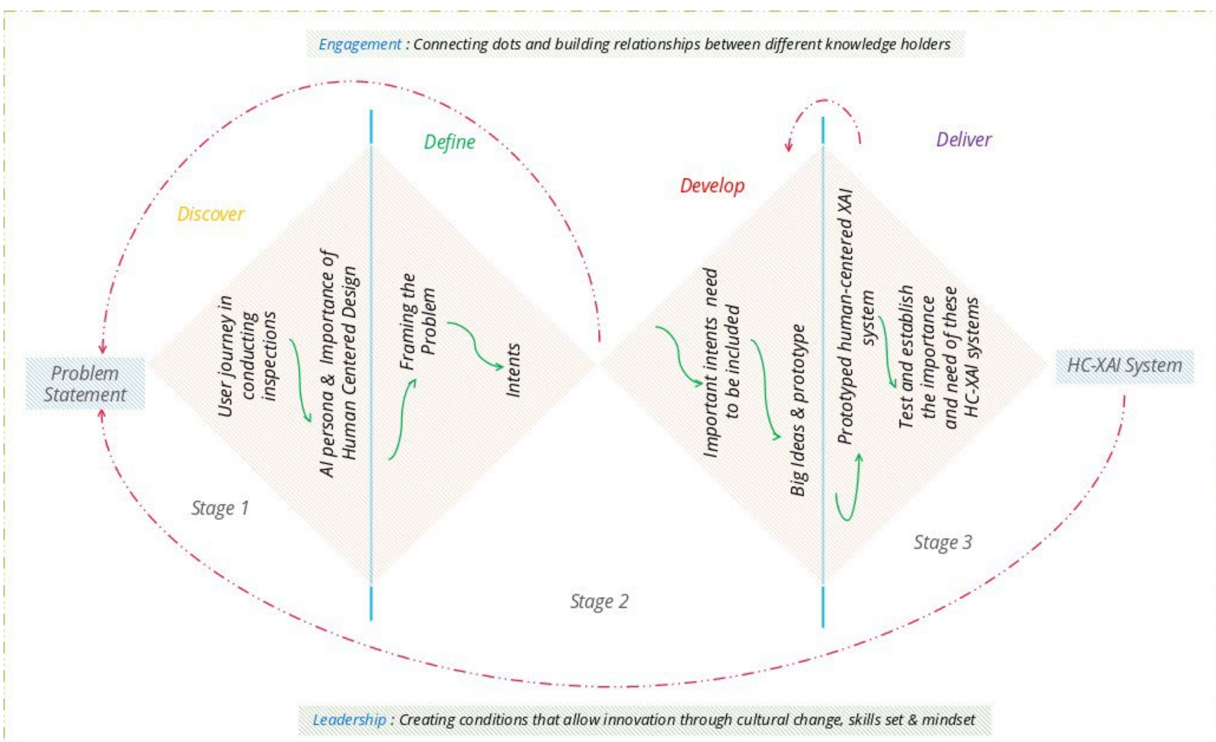
Stage 1: Discover User Needs (Discover Synthesis Phase)

Stage 1 of this research study involved interviewing individuals at two design firms (DEUS and Polytopal) and a local energy company (CPS Energy). The two design firms were selected because they are directly involved in working with the intersection of human-centered design and AI, and the energy company because I am currently working closely with the individuals involved in conducting regular daily inspections there. For this research study, one

industry professional from each design firm was recruited, and one from the energy company was recruited. These participants were recruited using email and phone, and interviews were conducted either in person or on Zoom, based on their preferences. Before moving to Stage 2, gathering insights from this Stage 1 was necessary. Three participants of the key knowledge holders were brought together, and ideation sessions were conducted individually to generate data. This data helped me to understand and gain insights in defining the problem—those three key knowledge holders were designers, AI developers, and inspection workers. The ideation sessions were conducted in a semi-structured manner. These ideation sessions were loosely based on the structure that is provided in Appendix A.

Figure 3

Modified Framework for Innovation



Modified Framework for Innovation : Proposed Framework

A constructivist approach was used to conduct research where there was a rich co-creative experience of participants, researchers, literature, and the data generated during this initial stage. For these reasons, the constructivist research paradigm was used in the data collection and analysis phase.

To briefly understand what constructivist grounded theory is, grounded theory must be understood. Grounded theory is a methodology focused on constructing theories to tell the issues embedded in society that are tied to humans, and was revolutionized by Glaser and Strauss (1967). This methodology suggests the researcher be a blank slate to attain theoretical sensitivity while conducting the research. Constructivist grounded theory, which Charmaz (2014) proposed, suggests that the theoretical sensitivity will be obtained through co-creating the experiences of participants, researchers, literature, and data. Since it is consistent with the researcher's epistemology and appropriate for this research, a constructivist grounded theory was used to collect and analyze data.

Ideation sessions were conducted to collect data from participants and each knowledge group holder individually. The data obtained from these ideation sessions was coded using constructivist grounded theory, where interviews were coded using open coding and from which the themes were identified from the data collected. This helped to synthesize the collected data and analyze the tasks in order to make critical decisions that helped to validate the hypothesized problem. Affinity mapping was used to determine patterns from coded data to generate themes. Affinity mapping is a process used to organize large amounts of data (concepts, ideas, and issues) into affinities (or groups) on their relationships created by Jiro Kawakita in 1960, an anthropologist. Adopting this method within this design ethnography was a beneficial inductive process to better understand the patterns within the data collected in this stage. The process of

analyzing the data using affinity mapping considered four essential elements:

- Step 1 - Generating ideas – Extracting concepts
- Step 2 - Shuffle and display ideas
- Step 3 - Sort ideas into groups
- Step 4 - Create header cards – Top-level descriptions of concepts

Once all these steps were completed, the affinity map itself had points that were going to be generated from each participant interview, such as motivations (goals, "if/then," context/surroundings, emotions, preferences) and behaviors (actions, "this/then"). Once these identified patterns were collected and grouped together, new patterns emerged. These patterns from the data collected and synthesized helped me to redefine the hypothesized problem and build a solid foundation for the next two stages. The template for analyzing and validating the data using affinity mapping is presented in Figure 4.

By the end of the stage, I better understood the problem in integrating human-centeredness and AI design, and developed an understanding of the significance of integrating AI in quality inspection. This stage was essential for this research to infuse human-centered design in an AI design that aims for human-machine collaboration. To truly design such systems by incorporating human-centered design practices in the AI development process, it was essential to determine where these three groups intersected or interacted with the proposed HC-XAI system.

Stage 2.1: Define Goals (Redefine Phase)

Once the data was collected and synthesized, the problem was redefined based on the findings collected from Stage 1. This redefined problem statement included research insights, intended audience, and the pain points discovered during Stage 1 research, and the way to solve the problem. This problem statement evolved with the system as the design became more

Figure 4

Affinity Mapping Template

Source: https://www.invisionapp.com/freehand/templates/detail/affinity-diagram-template

solidified. These findings helped focus on the redefined problem and on designing this HC-XAI system.

In this phase, the goal of what kind of system was going to be designed and developed was defined by keeping in mind the ethical aspects, and empathy for users and stakeholders, without ignoring the needs and requirements of the inspection workers and the businesses. All the pain points that exist were considered and converted to challenges in designing an HC-XAI system. With the redefined problem statement, the development mode in Stage 2.2 was reached. By the end of this stage, definitive goals to implement in the next sub-phase were provided, which was the perfect segway to developing ideas.

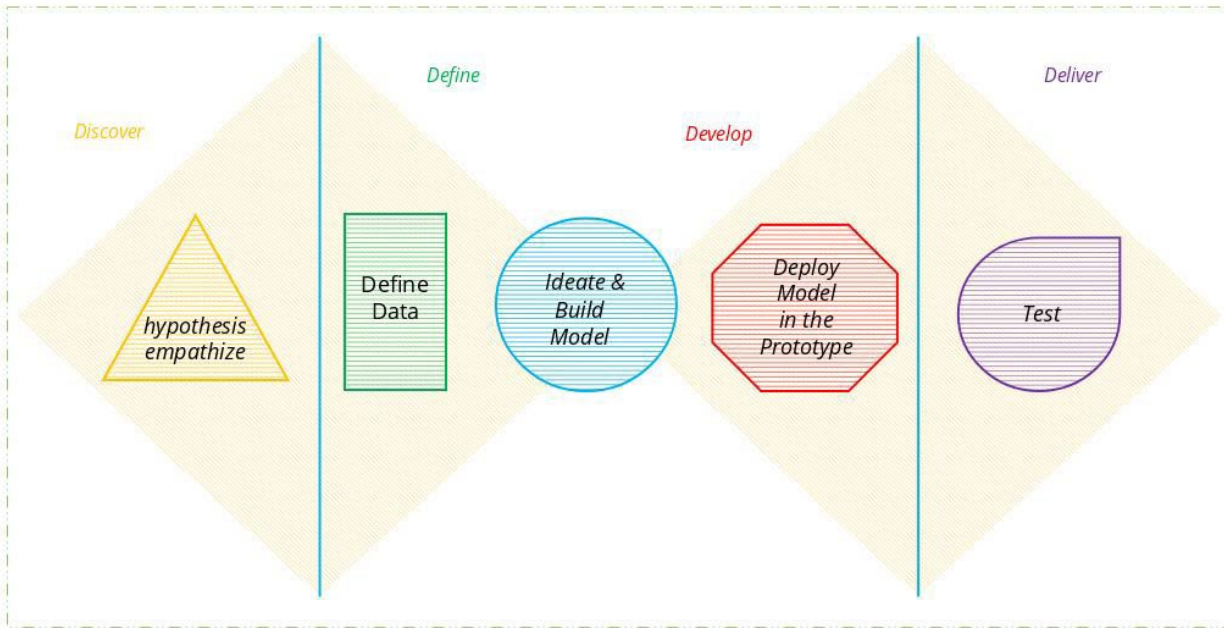
Stage 2.2: Develop Ideas (Prototyping Phase)

This stage of research had a redefined problem and the specific goals in designing the HC-XAI system to address. To address these goals and develop ideas in the prototyping phase, a feasibility mapping was generated based on what was achievable and what was not, and what were the immediate needs, long-term needs, and minor needs to be addressed. Based on these needs, the focus was on the immediate needs in developing the prototype. To develop the prototype, a blended pipeline that combines AI and Design pipeline in prototyping the HC-XAI system was proposed. See Figure 5.

For developing this prototype, ML and CV techniques were used in training the neural networks by following the proposed HC-XAI blended development framework. To train the neural networks, data was collected from the energy company based on immediate needs. Datasets were prepared to train the neural networks for the HC-XAI system to perform the operations and provide explainability in doing the tasks of augmenting human abilities in detecting anomalies during quality inspection. To train the ML model, it was necessary to be unbiased in annotating the data, one of the key elements in developing this system. Another key

Figure 5

Proposed HC-XAI Blended Development Framework



Modified framework for Innovation in development stage : Proposed HC-XAI System blended Pipeline (AI + Design)

element was adding an explainability factor to the HC-XAI system, in order to obtain trust. Then the research study moved on to the last and final stage, Stage 3.

Stage 3: Deliver Prototype (Implementation and Evaluation Phase).

Stage 3 of the research study corresponded to the implementation and evaluation of the prototype that was developed in the previous Stage 2.2., where the trained model was tested on whether it was spotting the differences between typical objects and anomalies during the inspection, in the process making decisions and providing understanding to the user and explaining why it was making decisions. These were crucial in validating how transparent and trustworthy the AI was in the loop with the human operator, and thus augmenting its capabilities. For validating the research questions, the usability testing method was employed, which helped me understand the factors related to the HC-XAI system. First, the study identified which part of

the HC-XAI system would be tested to conduct this validation. The test was then organized by scheduling the research planning meeting and inviting the persons interested in participating from CPS Energy. After that, users were recruited and informed by letting them know what participation would entail during this testing, and clarifying any logistical expectations involved. Once these were done, I ran the tests, and post-debriefs were conducted to record opinions and feedback.

In the last part of this validation, a collaborative synthesis meeting was conducted to discuss the issues observed or questions raised regarding the users' needs from this HC-XAI system. This validation method helped to iterate and incorporate anything that was missing during this HC-XAI system development design. This enabled or augmented the participants in making inspection routines and making sure the HC-XAI was explainable and understandable, which led to building trust between humans and AI. The research study investigated how this HC-XAI system could attain sustainability and inculcate social innovation. It also examined how such an HC-XAI system would make a significant difference by placing humans first and by collaboratively performing tasks with the use of technology at times where humans face risk in the inspection process and showing the difference of having an HC-XAI system.

Limitations and Ethical Considerations

This research study sought to achieve meaning-driven innovation by understanding the shifts in societal and cultural dynamics within the energy companies that I observed during the time I was working with the energy company. This approach of meaning and technology offered radical innovation, which was needed for the current world situation, where lots of things cause uncertainty. There were two main ethical issues that existed in this research study. One was data collection for training the model and bias in annotating that data. To address this known ethical

issue, data from CPS that was diverse in nature was gathered. Another ethical issue was the use of intellectual data from CPS Energy. For that, a non-disclosure agreement from CPS Energy was obtained. Finally, in attempting to design explainable AI, there was a possibility of getting trapped in the existing paradigms and being biased in conducting the research. These were the main potential limitations and ethical issues that were observed initially around conducting this research study.

Conclusion

This chapter provides an overall view of the methodology and how this research study was guided in data collection and analysis, and how the HC-XAI system would be tested. In the next two chapters, the developed system, lessons learned, and future research will be presented, along with conclusions.

Chapter 4: Data Collection and Data Analysis

This chapter elaborates upon all the insights and information gained during primary research in the discover phase of this research, where the purpose was to place humans alongside AI and to answer the research questions on how to design this human-centered AI by making it explainable and fostering social innovation. The study presents how the data was collected through semi-structured interviews from the three participants, and talks briefly about their backgrounds. It then presents the details about how member checking was carried out to check the validity of the data captured during the interviews. After collecting the data, the study presents how the data analysis used computational grounded theory's three-step methodological framework. After obtaining the refined and confirmed categories from the computational grounded theory framework, the study presents how affinity mapping was used to group the categories with respect to the research questions and present the insights derived after the data analysis. A visual representation of how the entire process was carried out in this research stage is presented in Figure 6.

Discover Phase (Data Collection)

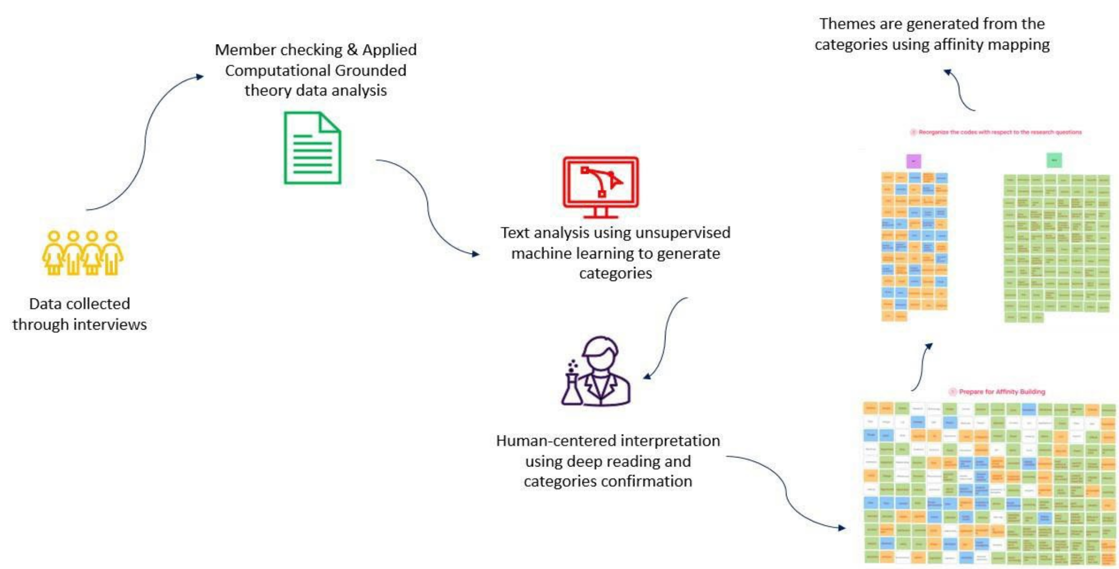
Semi-Structured Interviews with Energy Company and HCAI (Human-Centered Artificial Intelligence) Design Firms

The interviews conducted with participants provided a deeper understanding of the user's perspectives and perceptions regarding the research topic. These interviews also provided a guided pathway on what needed to be looked at in detail, expectations during the inspections, and what needs to be considered when designing a HC-XAI system. During this research phase, interviews were conducted with those associated with the inspections and those practicing or

advocating in developing HCAI solutions to address real-world problems. The interview guide used can be found in Appendix B and was used during the discover phase.

Figure 6

Overview of Data Analysis Carried Out in This Research



Since these interviews were exploratory, the participants were chosen from specific backgrounds that mainly covered three areas: AI, inspections in an energy company, and the human-centered AI domain. A total of three semi-structured qualitative interviews were conducted. Participants were selected based on their known ongoing involvement and knowledge in the specific fields. A brief introduction about the participants' backgrounds is presented below:

Participant 1 was the manager who oversees the entire inspection operation in a local electrical energy company, CPS Energy in San Antonio.

Participant 2 was the director of human-centered artificial development who advocates the importance of human-centered artificial intelligence and is responsible for working on projects

that heavily use human-centered AI to tackle socially challenged problems in DEUS.

Participant 3 was responsible for developing the guidelines and publishing the cookbook on designing human-centered AI in Polytopal.

The knowledge expertise of these three interviewees was crucial to this research, as they possess many years of research experience and are very active in those fields individually. Such types of experience helped me ask more profound questions and obtain different perspectives and thoughts on big ideas like explainability and human-centeredness in AI applications. This provided me with the ability to learn more about how these two big ideas could augment human intelligence, especially in energy companies. Conversations arose during the interviews regarding these big ideas and how there is a need for convergence of different technologies that can lead to sustainable and rational social innovation for companies who want to integrate AI systems to address real-world problems.

The Interviews

Primary Responses From Participant 1 in Key Areas: The first interview was with the manager at CPS Energy who oversees the entire operation of inspections for the energy company. When approached and asked how energy companies were using AI to conduct inspections, and the role of AI in the inspection industry, he said that they had started using drones and artificial intelligence to inspect in their company. He commented about how new ways of thinking could benefit regular inspections:

> The convergence of drone and artificial intelligence technologies to deliver the benefits of aerial inspections while enhancing worker safety…. With alternatives that keep humans on the ground, you can reduce risk. There is also an opportunity to reduce costs and time of inspections while improving the quality of the information gathered.

About this convergence, Participant 1 also said how it would help inspection workers reach out to the places where it is difficult to reach out. When asked about the amount of

reliability and trust in place when using AI during the inspection, the participant said that, with the AI systems that he encountered and tried with different inspection workers, they observed a phenomenon that workers expressed concern about the trust and explainability in using those systems during inspections. The participant also said that if there was a system to engage and augment how humans communicate, and to collaborate during the inspections, that would help augment their work more, reduce time and costs, be safer and would make the inspections more autonomous by augmenting human intelligence. When asked about the company policy in adapting these new and innovative modern technologies, he said that:

> Electrical companies are changing their standards all the time. The current standards for safety, inspection, and fixing are starting to work with modern options. The most modern of solutions today hinges on using drones and creating faster fixes than with normal solutions. As more energy companies start using drones, the industry standards will have to change, and early adoption is critical for the future of this type of work.

He added:

> As the sophistication of AI and drones increases, powerline inspections may become more autonomous—through a complete inspection revolution is unlikely in the short term…. It is more likely to be small deployments that grow steadily, resulting in incremental improvements in efficacy and reduced operations and maintenance costs. We might never get to full autonomy, but we are working toward augmented inspections that are safer, more effective, and less expensive.

He brought up the phenomenon of the Blackbox approach indirectly, which is very interesting, as is the context in which he mentioned the problem itself:

> Its human counterpart would be you asking a contractor to build you something. After close inspection of your request and doing some calculations, the contractor would send you an offer in dollars. The contractor, in this case, is the AI system, and the quote is comparable to the output. What went on in the contractor's mind is unknown to you and how they came up with that price is unclear. Sometimes just the output, or in this case, the quote, is all you need as an end-user. However, sometimes this is not sufficient, either because you need to re-explain the quote to another person or additional information to "believe" and "trust" the quote that the contractor gave you. To understand the output (quote), you need additional information like material cost and estimated man-hours. An explanation for why the quote is what it is. The same can be said for AI system outputs.

Primary Responses From Participant 2 in Key Areas: Interview 2 was with the director of human-centered artificial development at DEUS who operates at the intersection of design, AI, and business. DEUS is a firm located in the Netherlands that aims at developing human-centered AI applications. It also conducted valuable research on AI in practice and has publicly released a document discussing different AI elements that can bring value to an AI-driven business. When asked about what things are taken into consideration when designing AI for stakeholders, the participant stated that:

> The role of the designer/developer is to represent what is desirable for all the stakeholders involved and all the interacting parties. Having that empathy and understanding of who is going to be affected by it and where the ethical side of it plays a key role, and that is a responsibility you carry across the team.

The participant also mentioned that it is essential to assign equal responsibilities to three primary lenses when creating applications to impact ethics as a human: design, AI, and business. Another exciting thing that came out of the conversations was how psychology and ML play a vital role in creating systems that revolve around these three lenses. When asked about how AI systems were designed in DEUS, the interviewee stated that:

> Our goal is to revolutionize the way we think about, discuss, and create AI systems, starting with a vision of augmenting people and technology together and benefiting from information-rich displays that allow them to ask better questions and make more confident decisions. Human-Centered AI is the name given to this approach by a growing community of AI researchers (HCAI). Rather than eroding human agency, our purpose is to strengthen it.

Primary Responses From Participant 3 in Key Areas. Participant 3 was responsible for developing the guidelines and publishing the cookbook on designing human-centered AI at Polytopal, a firm that operates at the same intersection but has taken a slightly different approach. Polytopal is a strategic design studio in the USA that builds AI solutions for a positive human impact. They created a design language for AI, Lingua Franca, which the creators describe as a

standard set of techniques, frameworks, visuals, messaging, and overall design patterns that apply broadly to different kinds of AI to make it more usable, more trustworthy, and better aligned with people. The creators of Lingua Franca believe that design can bring clarity, intuition, and usability to these kinds of experiences, and that design is a lens we should use to make changes in the world, not just in designing toys and interfaces, but also in designing algorithms, business strategies, and policy frameworks. Participant 3 stated that:

> The notion of a design language does not exist in any field, other than UX or visual design. At the intersection of AI and design, people lack a shared vocabulary. Moreover, that is a huge part of getting people into the same imaginative space when they all share a vocabulary for what is happening, what they want to create, the processes that they want to use, etc.

Participant 3 also emphasized that the process of bringing together multiple fields, especially designers and data scientists, is very crucial:

> If you want to build AI, you should design it well. So, you should do user research and understand the means, create mock mock-up experiences. Question your assumptions, and then develop a plan. I mean, think about ethics, usability, interaction, and information architecture. It is like approaching it like a formal area of inquiry, not a technical, just a single kind of technology. So, what we found is that when you build AI - it is because of the complexity of dealing with the data and the kind of effect of that data on the user experience, you might create sort of unintended consequences.

Participant 3 stated that designing an understanding system is key to building more collaborative AI that augments humans and machines in the loop, and insisted that more emphasis should be placed on this:

> The main goal is to understand AI systems and the workings that operate within the models. Not only that but also make it understandable for those who need to collaborate with the AI system to create trust in the collaboration between human and machine.

Member Checking

Before analyzing the data collected, I conducted member checking to ensure that the validity of the collected data was what the participants said, and not my biased interpretation.

After cleaning the data collected from the three participants, a one-page synthesized summary of the data collected was sent to each participant via email for cross-checking to ensure that what was captured during the interview was the same as what they meant. With this synthesized member checking, the participants' knowledge and understanding were grounded within their experiences, as knowledge is socially constructed (Crotty, 1998; Gray, 2013; Hammersley, 1992; Snape Spencer, 2003). This synthesized member checking process was carried out for each participant and each was asked to comment if any information was not what they meant or intended. A synthesized document from CPS Energy company interview data is presented in Figure 7.

I generated the synthesized data using Natural Language Processing text mining to avoid biases when producing the synthesized data from the semi-structured interviews. This step allowed me to increase the validity of the data gathered during those semi-structured interviews with the participants. According to Freire (2000), member checking is one of the crucial parts of data analysis in any qualitative research, which provides validation and addresses the co-constructed nature of knowledge by providing participants with the opportunity to engage with, and add to, interview and interpreted data. After receiving confirmation from the participants about the data, I moved on to the next step of this research, conducting data analysis, and the detailed process is presented in the coming sections of this chapter.

Figure 7

Example of Company Data Collected for Member Checking

In need of a collaborative approach between humans and AI during the inspection

Improved inspections of power delivery infrastructure can help reduce the risk of outages by pinpointing equipment that needs to be repaired or replaced. Today's inspections require a lot of time, people, and equipment. "It is a combination of aerial inspections from helicopters and fixed-wing aircraft, climbing poles, using bucket trucks, and walking inspections. "Utilities use a different mix of those depending on the power line's importance, budgets, and timing."

Utilities can use helicopters to reach remote areas and cover more territory but sending inspectors into the air has safety risks. Utilities are exploring drone technologies to deliver the benefits of aerial inspections while enhancing worker safety. "With alternatives that keep humans on the ground, you can reduce risk," said Lewis. "There is also an opportunity to reduce costs and time of inspections while improving the quality of the information gathered."

With the emergence of relatively inexpensive drones Inspecting power lines becomes easier to manage with drone technology, simple as that.

- Drones enable collecting the needed data for identifying and mitigating risks in power distribution in advance
- Greatly reduce man-hours and costs by automating inspections, saving 30 – 50% of the cost and time when using drones to conduct power line inspections.
- Assess the condition and orientation of all components of cell towers with no need for workers to ascend to the height
- Inspections are drones from a safe distance while increasing efficiency due to data accuracy and reliability with real-time images, video feed, and zoom/thermal/4k capabilities that can be transmitted to a ground control station.
- Thermal and LIDAR can be used to aid in inspecting and monitoring the corridor for power lines and towers
- Higher-resolution visual inspections than ground-based inspections
- Towers remain functional during the inspection
- Drones enable collecting the needed data for identifying and mitigating risks in power distribution in advance
- Greatly reduce man-hours and costs by automating inspections, saving 30 – 50% of the cost and time when using drones to conduct power line inspections.
- Assess the condition and orientation of all components of cell towers with no need for workers to ascend to the height
- Inspections are done from a safe distance while increasing efficiency due to data accuracy and reliability with real-time images, video feed, and zoom/thermal/4k capabilities that can be transmitted to a ground control station.
- Thermal and LIDAR can be used to aid in inspecting and monitoring the corridor for power lines and towers
- Higher-resolution visual inspections than ground-based inspections
- Towers remain functional during the inspection

We have learned from our drone research that we need an automated system for processing the large number of images that drones can capture. CPS is compiling and curating thousands of images for AI developers in a similar project for distribution systems, who will use the images to train algorithms later this year. AI systems are, pretty much without exception, developed using machine learning methods. For now, it is essential to realize that machine learning can lead to an AI system that is something of a 'black-box' in terms of how it works. This 'black-box' phenomenon is observed in previous AI models tested because the more complicated an AI system gets, the less understandable its decision-making process becomes for humans.

It is also worth noting that an algorithm makes no mistakes. It will identify every faulty object, whereas a person might be tired or otherwise preoccupied and repentantly miss something important. This is especially true when inspection data is complex, as the margin for error is much smaller.

As AI gets more advanced, the days of analyzing data solely using human intelligence will be over. Anyone who does not have their artificial intelligence software will outsource.

As a result, companies need to ask themselves if they intend to develop their software in the future. If not, it might be a good idea to establish a partnership now. This will free up resources, which will allow you to build or improve other core competencies that are more beneficial to your focus area.
and artificial intelligence (AI)

Comments: Concisely, you captured the core idea of what we are going through and the missing piece.

Data Analysis

This research is mainly focused on how AI can augment humans in performing tasks. I used the computational grounded theory framework (Nelson, 2020) in conducting data analysis. This new way of conducting data analysis was done by converging grounded theory methods and computational textual analysis. This convergence addressed the significant issues in grounded theory, where generating categories can be subjective; like every other human being, researchers are subject to confirmation bias. The second major thing in grounded theory is that data analysis is not easily reproducible, and it is not easy to get the same person to code the same article in the same way twice. This is where the computational textual analysis comes into play, by adapting the latest developments in ML that would overcome the above two issues in grounded theory. This approach of converging these two methods for conducting data analysis not only helped me to avoid subjectivity and biases, but made the codes or categories easily reproducible, and faster.

A three-step framework was used to obtain the categories from collected data and bring together the best parts of these two methodologies. The detailed process of the three-step framework used in this research is presented below.

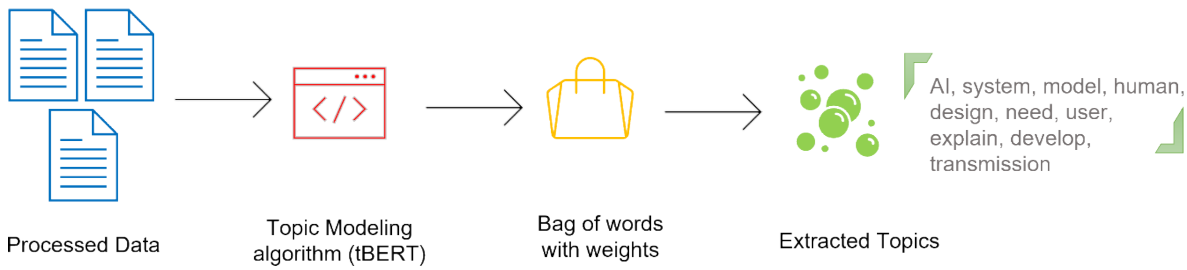
Human-Centered Computational Exploratory Data Analysis for Detecting Categories. In this first step, the data collected was analyzed using computer-assisted text analysis techniques, especially ML, which helped me explore the data by reducing its complexity and messiness into more interpretable and straightforward lists, or networks of words. There were different ways to identify categories across the data in ML. I used a topic-modeling algorithm to identify patterns across the data to uncover categories within a corpus.

The topic-modeling algorithm is one of the most popular ways of conducting unsupervised text classification. This topic modeling algorithm works in creating topic models by

analyzing the co-occurrence of words within the corpus and reducing a complicated corpus to simple, interpretable groups of words. Figure 8 shows the overall process of how the algorithm works.

Figure 8

Overview of How the Topic Model Creation Works Using tBERT



When the processed data was fed to the topic model tBERT (topic BERT) that was trained using an unsupervised technique, it allowed the model to extract meaningful patterns from text, first by generating a bag of words with weights from the data and then by producing the extracted topics. This process made it easier to glance over textual data and better understand the latent distribution of topics that live underneath the data. A sample bag of words based on the weights is shown in Figure 9, along with the topics.

When the tBERT algorithm was applied on the entire corpus, it produced extracted topics, as presented in Table 1.

These lists of words were weighted, and each list was a topic that helped me to summarize and visualize the corpus quickly. These topics forming the corpus helped me to detect thematic patterns across the documents. The primary intent of this step in analyzing the data is to perform the initial and entirely inductive analysis of the collected data. This helps me visualize the topics obtained that are exceptionally helpful and classify texts the same way every time,

making the classification step fully reproducible. Besides, this unsupervised text classification approach helped me to interpret the estimated categories rather than create the categories that helped me move away from the data and avoid biases. However, these algorithms provide me the ability to quickly summarize the text to obtain the categories or patterns from the data collected and make broad comparisons. It helped me look at the data differently and surprised me with the categories that were not thought of previously in categorizing the text data.

Figure 9

How Sample Topics are Generated From the Bag of Words With Weights

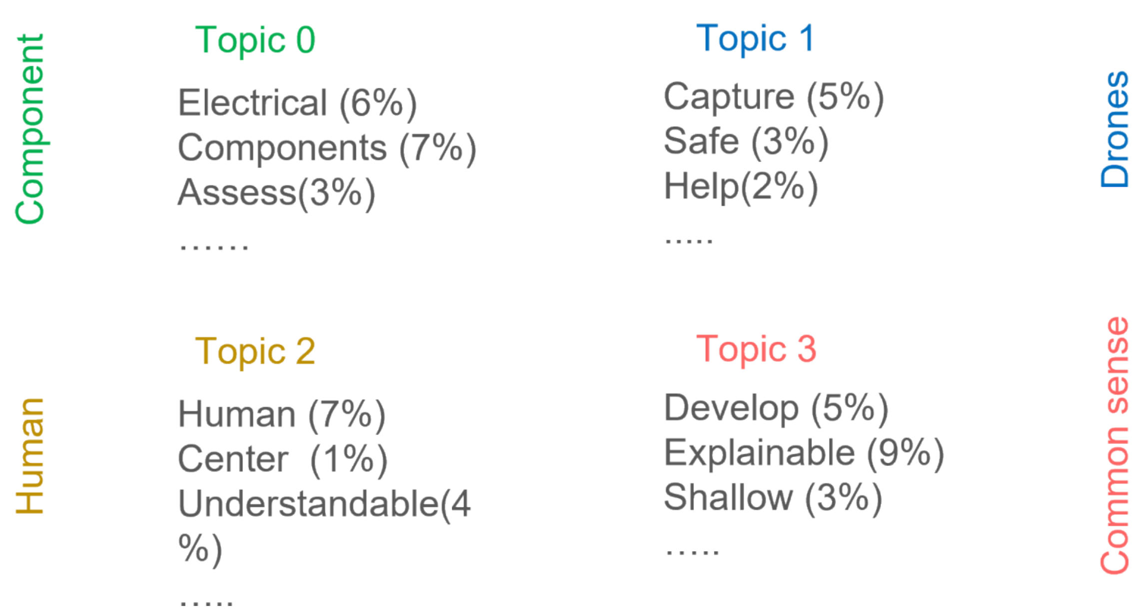


Table 1

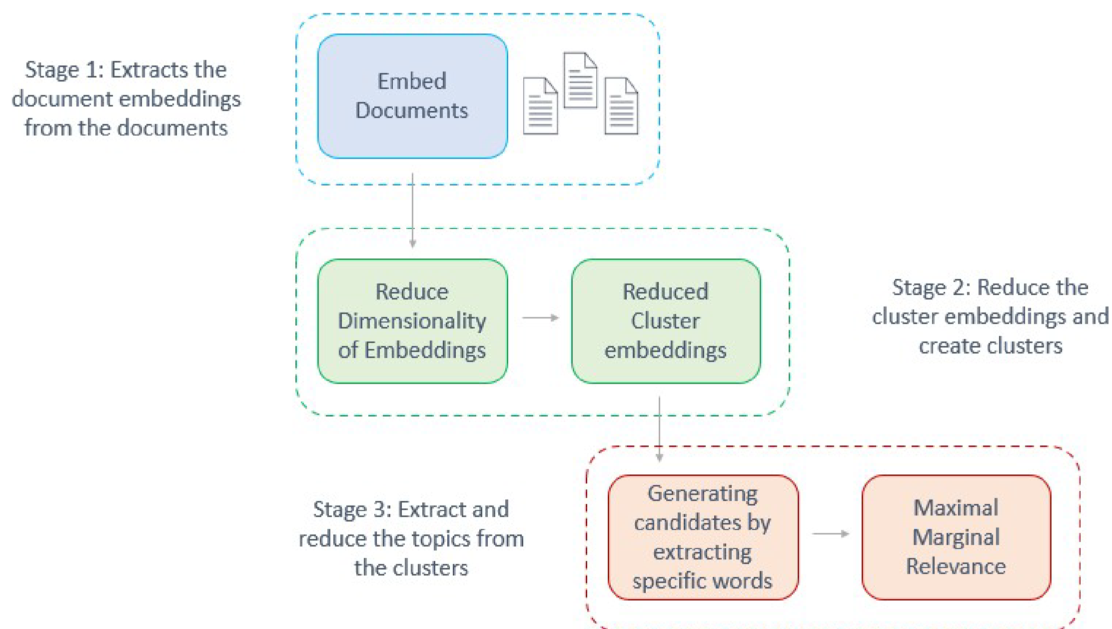
Sample List of Topics With Highest Weighted Words

Sample List of Topics with Highest Weighted Words						
AI	Solution	Computer	Difficult	Application	Implement	Robot
System	Thing	Kind	Make	Wrong	Assumption	Artificial Intelligence
Human	Control	HCAI	View	House	Rationalism	Center AI
Model	Build	Provide	Enable	Interface	Rationalist	Deal
Design	Research	Look	Identify	Failure	Action	Consider
Datum	Decision	Base	Cost	Powerline	Caregiver	Tower
People	Try	Different	Increase	Company	AI model	Mean
Need	Know	Process	Define	Long	Unintended Consequences	Software
User	Come	Lead	Today	Electrical	Design Process	Get
AI system	Image	Interact	Risk	Component	Study	Idea
Machine	AI	Unintended	Find	Instead	Place	Utility
Inspection	Output	Consequence	Fix	High	Set	Transmission
Work	Actually	Product	Say	Start	Invoice	Number
Go	Study	Drive	EPRI	Collect	Bit	Great
Drone	Machine Learning	Line	Believe	Distribution	Point	Developer
Use	Happen	Certain	Explainable	Real	Digital	Researcher
Time	XAI	Price	Imagine	Artificial	Modern	World
Problem	Example	Complex	Sort	Develop	Early	Goal
Way	Power	Good	Pattern	Improve	Available	Autonomy
Learning	Want	Train	Begin	Performance	Automate	Understandable
Car	Intelligence	Autonomous	Require	Insight	Accuracy	Quote
Create	Person	Box	Change	Explain	Visual	Give
Right	Information	Phase	Able	Level	Remain	Instance
Lot	Method	Approach	Future	Simply	Analysis	Question
Case	Black	Center	Feed	Situation	Advanced	Explainability
Experience	Think	Designer	Simple	Context	Ask	Self
Technology	New	Important	Reduce	Play	Deliver	See
Algorithm	Help	Community	Data	Social	Potential	Fake
Understand	Well	Safety	Allow	Voice	Additional	Bias

The above-mentioned process was used to create one topic model, using the tBERT algorithm. The tBERT algorithm is a topic modeling algorithm that uses transformers to create dense clusters and allows easy interpretation and visualization of the topics generated from it. This algorithm mainly consists of three stages: first, it extracts the document embeddings from the documents, then it reduces the cluster embeddings and creates clusters of semantically similar documents from this; and finally, it extracts and reduces the topics to create a topic representation in order to improve the coherence of words with Maximal Marginal Relevance. Figure 10 is a visual representation of this algorithm.

Figure 10

tBERT Topic Modeling Stages



To determine the best topic model for the collected data, I ran four topic models to determine which model would be the best for the data collected using this approach. Each topic model varied on the number of topics used to run the topic model algorithm on the corpus. After running these different topic models, I observed that the topic model with the smaller number of

topics combined multiple topics into one. I noticed this when running a more significant number of topics subsequently with the other given topic models. The lower topics model generated top-weighted words like power, want, intelligence, person, information, method, and understand. I found that these generated top-weighted words addressed the highlight of having the human intelligence needed in the powerline inspection. When compared to the same corpus with the higher topic model, where the topics of the top-weighted words were more than the previous model used, I observed words like enable, human, apply, lead, favor, human, understand, continuously, casual, long, and question, which were more detailed and specific. However, in this instance I was looking for more meaningful topics, more than to be more specific in the corpus. I ran the experiment by changing the topic models and changing the number of words to observe this phenomenon. After running various topic models, I observed that when changing the amount of the topics from lower to higher, there was an observable difference of the top-weighted words generated. I chose the topic model in between to have a meaningful insight. After that, I applied the same by applying sentiment analysis to obtain the properties and categories for the topics discovered, and generated the code dictionary with categories, properties, and topic dimensions. A sample code dictionary with category, property, and topic dimensions generated from the topic modeling is presented in Figure 11.

Figure 11

Coding Dictionary Generated From the Structural Topic Modeling

---Coding Dictionary---

CATEGORY	PROPERTY	DIMENSION
lead	rationalism	lead
...	...	favor
...	...	human
...	time	true
...	...	lead
...	...	new
...	developer	lead
...	...	drive
...	...	autonomous
begin	system	begin
...	...	autonomous
...	...	drive
...	design	enable
...	...	human
...	...	believe
...	approach	begin
...	...	autonomous
...	...	drive
drive	car	drive
...	...	increase
...	...	social
...	self	drive
...	...	autonomous
...	...	begin
...	algorithm	drive
...	...	get
...	...	true
enable	design	enable
...	...	human
...	...	believe
...	user	enable
...	...	human
...	...	lead
...	level	enable
...	...	human
...	...	apply
believe	design	enable
...	...	human
...	...	believe
...	technology	believe
...	...	start
...	...	human
...	datum	get
...	...	statistician
...	...	believe
learn	machine	human
...	...	autonomous
...	...	replace
...	algorithm	drive
...	...	get
...	...	true

enable	design	enable
...	...	human
...	...	believe
...	user	enable
...	...	human
...	...	lead
...	level	enable
...	...	human
...	...	apply
believe	design	enable
...	...	human
...	...	believe
...	technology	believe
...	...	start
...	...	human
...	datum	get
...	...	statistician
...	...	believe
learn	machine	human
...	...	autonomous
...	...	replace
...	algorithm	drive
...	...	get
...	...	true
...	time	true
...	...	lead
...	...	new
recognize	example	recognize
...	...	social
...	...	hide
...	designer	recognize
...	...	inspire
...	...	social
...	distinction	recognize
...	...	important
build	vehicle	come
...	...	improved
...	...	advanced
...	highway	come
...	...	improved
...	...	advanced
...	center	center
...	...	advanced
...	...	come
replace	machine	human
...	...	autonomous
...	...	replace
...	human	human
...	...	autonomous
...	...	replace
...	idea	human

...	...	reject
...	...	autonomous
bring	robot	human
...	...	humanoid
...	...	shake
...	approach	begin
...	...	autonomous
...	...	drive
...	food	old
...	...	live
...	...	deliver
help	satisfaction	able
...	...	experience
...	...	contribute
...	agenda	help
...	...	change
...	...	imagine
...	way	work
...	...	enable
...	...	lead
center	management	advanced
...	...	financial
...	...	compelling
...	vehicle	come
...	...	improved
...	...	advanced
...	highway	come
...	...	improved
...	...	advanced
understand	belief	understand
...	...	human
...	...	continuously
...	design	enable
...	...	human
...	...	believe
...	user	enable
...	...	human
...	...	lead
allow	learning	true
...	...	drive
...	...	allow
...	machine	human
...	...	autonomous
...	...	replace
...	design	enable
...	...	human
...	...	believe
include	human	human
...	...	autonomous
...	...	replace
...	machine	human
...	...	autonomous
...	...	replace

From the coding dictionary generated using the topic modeling algorithm to further explore the categories, I used standard text analysis and generated weights for the categories by

excluding the property and topic dimensions resulted in the categories with weights. A sample of the categories generated is shown in Figure 12.

Figure 12

Categories Generated, Along With Their Weights

---Categories with count---

CATEGORY	WEIGHT
AI	0.01008252164411196
system	0.006304110279982496
human	0.0051273430276540305
model	0.004034638579137598
design	0.0036984113642094646
datum	0.003530301756745398
people	0.0031100277380852314
need	0.0031100277380852314
user	0.003025972934353198
AI system	0.003025972934353198
machine	0.002689753719425065
inspection	0.002605698915693832
work	0.0024375893082289653
go	0.0021013700933008323
drone	0.002017315289568799
use	0.002017315289568799
time	0.0019332604858367656
problem	0.0018492056821047323
way	0.001765150878372699
learning	0.001765150878372699
car	0.001765150878372699
create	0.0016810960746406658
right	0.0016810960746406658
lot	0.0016810960746406658
case	0.0016810960746406658
experience	0.0016810960746406658
technology	0.0015970412709086323
algorithm	0.0015970412709086323
understand	0.0015970412709086323
solution	0.001512986467176599
thing	0.001512986467176599
control	0.0014289316634445658
build	0.0014289316634445658
research	0.0014289316634445658
decision	0.0014289316634445658
try	0.0014289316634445658
know	0.0013448768597125326
come	0.0013448768597125326
image	0.0013448768597125326
ai	0.0013448768597125326
output	0.0013448768597125326
actually	0.0013448768597125326
study	0.0013448768597125326
machine learning	0.0013448768597125326
happen	0.0012608220559804993
XAI	0.0012608220559804993
example	0.0012608220559804993
power	0.001176767252248466
want	0.001176767252248466
intelligence	0.001176767252248466
vision	0.0014671361502347417
autonomy	0.0014671361502347417
ai	0.0014671361502347417
self	0.0014671361502347417
car	0.0014671361502347417
well	0.0014671361502347417
rationalism	0.0014671361502347417
method	0.0014671361502347417
recognize	0.0014671361502347417
rationalist	0.0014671361502347417
action	0.0014671361502347417
caregiver	0.0014671361502347417
question	0.0011737089201877935
understanding	0.0011737089201877935
agency	0.0011737089201877935
consequence	0.0011737089201877935
contrast	0.0011737089201877935
automation	0.0011737089201877935
display	0.0011737089201877935
researcher	0.0011737089201877935
understand	0.0011737089201877935
learn	0.0011737089201877935
thinking	0.0011737089201877935
research	0.0011737089201877935
base	0.0011737089201877935
belief	0.0011737089201877935
relationship	0.0011737089201877935
developer	0.0011737089201877935
causal	0.0011737089201877935
sense	0.0011737089201877935
designer	0.0011737089201877935
interface	0.0011737089201877935
build	0.0011737089201877935
application	0.0011737089201877935
strategy	0.0011737089201877935
artificial	0.0008802816901408451
software	0.0008802816901408451
level	0.0008802816901408451
device	0.0008802816901408451
decision	0.0008802816901408451
get	0.0008802816901408451
daily	0.0008802816901408451
life	0.0008802816901408451
replace	0.0008802816901408451
meaningful	0.0008802816901408451
task	0.0008802816901408451
forward	0.0008802816901408451
bring	0.0008802816901408451
help	0.0008802816901408451
information	0.0008802816901408451
world	0.0008802816901408451
Vinci	0.0008802816901408451
strong	0.0008802816901408451
true	0.0008802816901408451
time	0.0008802816901408451
data	0.0008802816901408451
network	0.0008802816901408451
come	0.0008802816901408451
event	0.0008802816901408451
performance	0.0008802816901408451
improve	0.0008802816901408451
solution	0.0008802816901408451
right	0.0008802816901408451
experience	0.0008802816901408451
search	0.0008802816901408451
professional	0.0008802816901408451
setting	0.0008802816901408451
case	0.0008802816901408451
satisfaction	0.0008802816901408451
highway	0.0008802816901408451
clear	0.0008802816901408451
advanced	0.0008802816901408451
increase	0.0008802816901408451
management	0.0008802816901408451
old	0.0008802816901408451
disability	0.0008802816901408451
artificial intelligence	0.0008802816901408451
human agency	0.0008802816901408451
human control	0.0008802816901408451
AI system	0.0008802816901408451
center AI	0.0008802816901408451
user interface	0.0008802816901408451
early	0.0005868544600938967
public	0.0005868544600938967
professor	0.0005868544600938967
Maes	0.0005868544600938967
start	0.0005868544600938967
place	0.0005868544600938967
interest	0.0005868544600938967
different	0.0005868544600938967
value	0.0005868544600938967
crash	0.0005868544600938967
danger	0.0005868544600938967
hide	0.0005868544600938967
bias	0.0005868544600938967
economy	0.0005868544600938967
remain	0.0005868544600938967
direct	0.0005868544600938967
reject	0.0005868544600938967
idea	0.0005868544600938967
notion	0.0005868544600938967
responsibility	0.0005868544600938967
hand	0.0005868544600938967
humanoid	0.0005868544600938967
generation	0.0005868544600938967
grow	0.0005868544600938967

Overall, this step served me as the first reproducible category detection step, and had some weaknesses that were improved by steps two and three in the computational grounded theory process. At this stage, all the categories' topics and properties generated were completed by using unsupervised Natural Language Processing. These refined categories identified by topic modeling needed a guided deep reading by humans, which was carried out in the next step. A third step with additional computational techniques helped me to confirm the categories detected in step three, which will be presented in the coming sections of this chapter.

Human-Centered Interpretation Using Grounded Theory for Categories Detected. I implemented this step of data analysis in order to achieve three main things: confirm the plausibility of the patterns identified via an analysis of the computationally driven results in step one; add interpretation to the analysis; and potentially modify the identified patterns to better fit a human, and holistic, reading of the data using computational grounded theory.

From the results obtained in step one using computational analysis, I used computational guided reading to check the interpretations of the groups of words and categories produced in the previous step, which helped determine if those groups of words translated into complete sentences or arguments. Through this guided reading of the corpus, I was confident in interpreting the words with the topic distribution in the data. This process helped me not skip over essential passages because of fatigue or bias. Since this step involved humans reading the text along with the numbers, context in the previous step provided a meaningful, more traditional sociological and theory-informed approach.

In this step, the patterns from step one were taken and human-centered interpretation was applied using the traditional grounded theory approach, by associating the categories obtained from the related corpus data. A sample of data-driven categories and data interpretation back to the categories identified, using human-centered interpretation, is presented in Table 2.

Through this reading, more concrete redefined categories were obtained that brought the interpretation back to the data by duplicating the traditional approach to grounded theory, but with a computational twist. From steps one and two, the refined categories obtained were immediately reproducible, allowing other researchers to reproduce the computational portion of the analysis and be scalable. The interpretive portion helped translate the computational output generated in step one into more meaningful concepts that helped to draw more abstract

conclusions about the data. After identifying and redefining the categories from these two steps, a computational technique called supervised learning was used to confirm these categories in the last step of this computational grounded theory data analysis.

Table 2

Sample Data-Driven Categories Identified

Data-Driven Categories Identified	Data Interpretation Back to the Categories
Inspection	"There are a lot of available drone options, but in the case of using them to work with powerline inspection, specific models and expertise need to be considered. Electrical grids can push energy back against the drone and cause problems for the feeds, so selecting the suitable drone becomes important within this category."
Intelligence	"Artificial Intelligence can analyze and annotate large amounts of images in short periods, making your inspections effortless as well as very, very accurate"
Explainable	"For instance, look at convolutional neural networks that can contain hundreds of thousands of nodes (decision points) that interact on different levels, it can be difficult for a human to conceptualize the model and understand the output."
Components	"These are difficult to manage, and experts have to work within parameters that not only help fix electrical components but also provide safety measures of people that are doing the work."

Category Confirmation. From the above two steps, the categories were identified, and the identified categories were refined by interpreting the computational output through guided reading. After obtaining the refined categories in this step, tests using supervised text analysis and Natural Language Processing were conducted. To ensure that the identified categories were not an artifact of a specific algorithm and were not based on my biased interpretation through deep reading, the participants' data was coded by taking a small amount of data and using a

supervised ML algorithm, using Natural Language Processing to code the remaining amount of data from the participants. This allowed me to confirm the refined categories from the previous step and test that the categories identified were generalizable to the entire corpus and act as a reliability test to the grounded theory process.

This supervised ML algorithm relied on hand-coding text; this method could be applied to most patterns identified. With reliable categories identified in step one and backed up with expert interpretation from the deep guided reading in step two, and with refined categories confirmed using a supervised ML algorithm, I confirmed categories by building affinity maps and reorganizing and shuffling these confirmed redefined categories with respect to the research questions, and identified the themes.

Affinity Map Building From the Confirmed Categories. After obtaining the confirmed categories from this computational grounded theory three-step framework methodology, affinity mapping was used to visualize the categories obtained from the above method, and the affinity map was built using these categories. The process of affinity building is to organize the categories in one place and shuffle through and visualize them. During this process three themes were observed, each represented by a different color, where blue represents human-centered, green represents the rationalistic approach of conducting inspections, and orange represents explainable AI. The color coding facilitated mapping these themes with respect to the research questions. This shuffling of the categories based on the themes left some of the categories that did not fit in any themes as white labels. A visual representation of this affinity building, using the categories, is shown in Figure 13.

Reorganize the Categories Corresponding to the Research Questions. After building the affinity map, those grouped categories were organized based on the research questions and

arranged corresponding to the research questions that this research seeks to answer. After performing this step, I obtained the reorganized categories shown in Figure 14.

Conclusion

This chapter presented a detailed overview of how the data was collected, the participants' background, and how the data collected through semi-structured interviews were synthesized. It also presented how member checking was carried out and the steps that were involved in conducting data analysis, using a three-step computational grounded theory framework to obtain categories and themes. This chapter also presented how the process of affinity mapping was used in building the categories to obtain themes, and how these themes were mapped together to the research questions and helped to attain some key insights and big central ideas that enabled moving to the next stage of this research. The key insights and big central ideas from the data analysis will be presented in the next chapter.

Figure 13

Affinity Building From Refined Categories

Figure 14

Reorganizing the Categories With Respect to Research Questions

Chapter 5: Investigating the Research Question

This chapter will present how the research questions were investigated using Stages 2.1, Stage 2.2, and Stage 3 of the research study. The first part of this chapter discusses Stage 2.1 and Stage 2.2, which primarily deal with the redefine and prototyping phases of the HC-XAI design. I will present details of how this research study's primary research question was investigated through the prototyping phase (Stage 2.2) from a detailed proposed developmental framework that guided in designing HC-XAI. Then I will discuss Stage 3 of this research study, and how the developed HC-XAI design was evaluated by presenting the results from the AI model and validating it using usability testing. Overall, by the end of the chapter, each stage of the modified framework of innovation that helped in answering the primary research question of this study will be presented.

Stage 2.1: Redefine Phase

After finishing the Stage 1 data collection interviews with participants in the energy company and the two design companies, and analyzing the collected data using computational grounded theory, three critical insights emerged from the data analysis that act as a basis for this Stage 2.1 redefine phase. These emerged critical insights played an essential role in whether the original problem statement needed to be modified or not, before presenting critical insights from the Stage 1 research.

The first critical insight emerged from the first interview with the energy company, the significance of the explainability factor in AI systems that would enhance workers' safety, which provided a rationalist approach.

A second critical insight emerged that would help the firm to be conscious of trends in the competitive environment, prepare for a challenging future, and ensure that sufficient attention

is focused on the long-term by welcoming innovative technologies and to conduct inspections by converging AI and drone systems to perform faster inspections and make the process autonomous. Inspection workers could focus on more significant problems in identifying the faults and working on the solutions to fix the issues much quicker.

A third insight emerged from interviews with the design companies, who emphasized the need for a standard tool to augment AI and humans. Currently, tools are primarily tailored to individual companies and their design practices, making it essential to create a standard tool that all can use. Having a standard tool would assist those who are new to the field and may not have access to various tools and procedures. The interviewees also emphasized the ethics and bias factors involved in designing AI-driven products and services. HCAI that focuses on meeting user needs should consider how diverse and varied their end users can be. The ML algorithms cannot be trained to benefit one group of people while being incredibly harmful to others. It is essential to address socio-cultural implications as well.

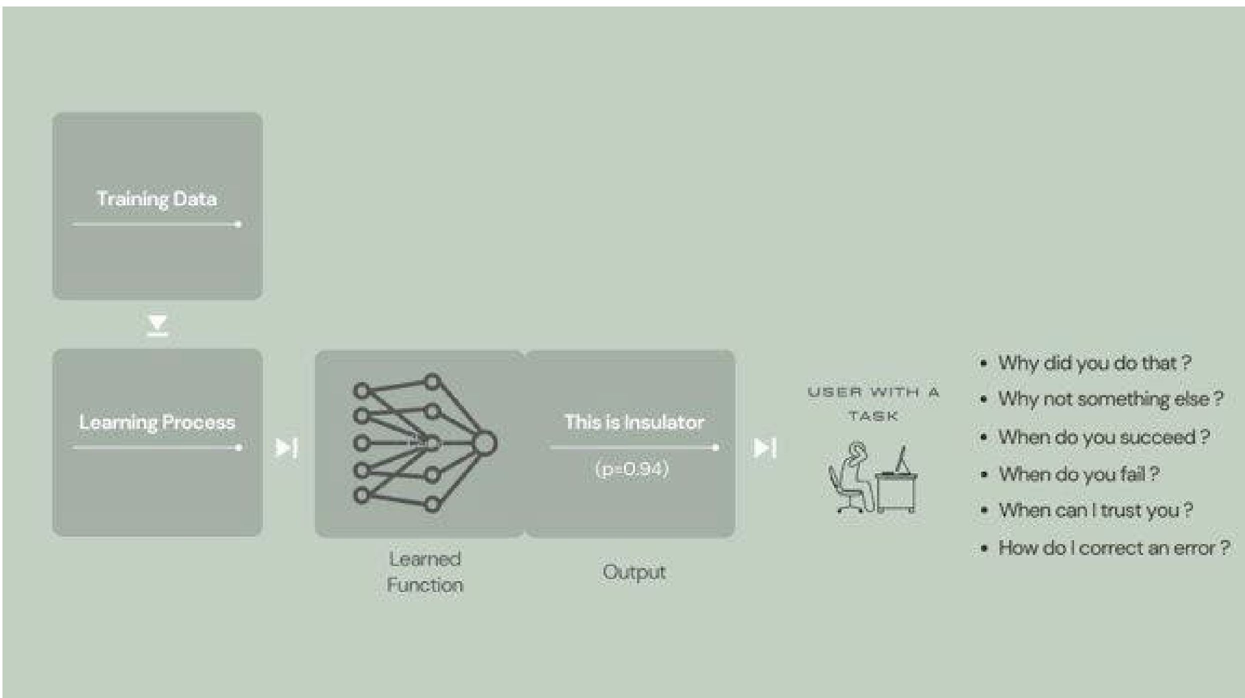
As mentioned, three critical ideas emerged from the data analysis, implying that there is lack of explainability with the intelligent systems that affects the trust factor between the systems and humans. To address this, a rationalistic approach is needed in looking into changing the process of conducting inspections for the long run by bringing an innovative approach through converging the fields of AI and drones Based on the three vital main insights, or themes, that emerged from Stage 1, I looked at the original problem statement that was posed at the start of this research. and did not find any need or necessity to redefine the problem statement initially assumed at the beginning of this research.

I kept the original problem statement and continued investigating answers to the research questions posed at the start of the research study. A visual illustration of the current problem

statement that defines what is currently happening while conducting inspections is illustrated in Figure 15. This visual illustration describes the current situation of the AI models used and the problem behind using those, training the AI model to use the data, and predicting the results based on the trained data. With this approach, there is a lack of explainability, reliability, and trustworthiness when used in the inspection industry for detecting anomalies.

Figure 15

Visual Illustration of the Problem Statement of the Current Problem

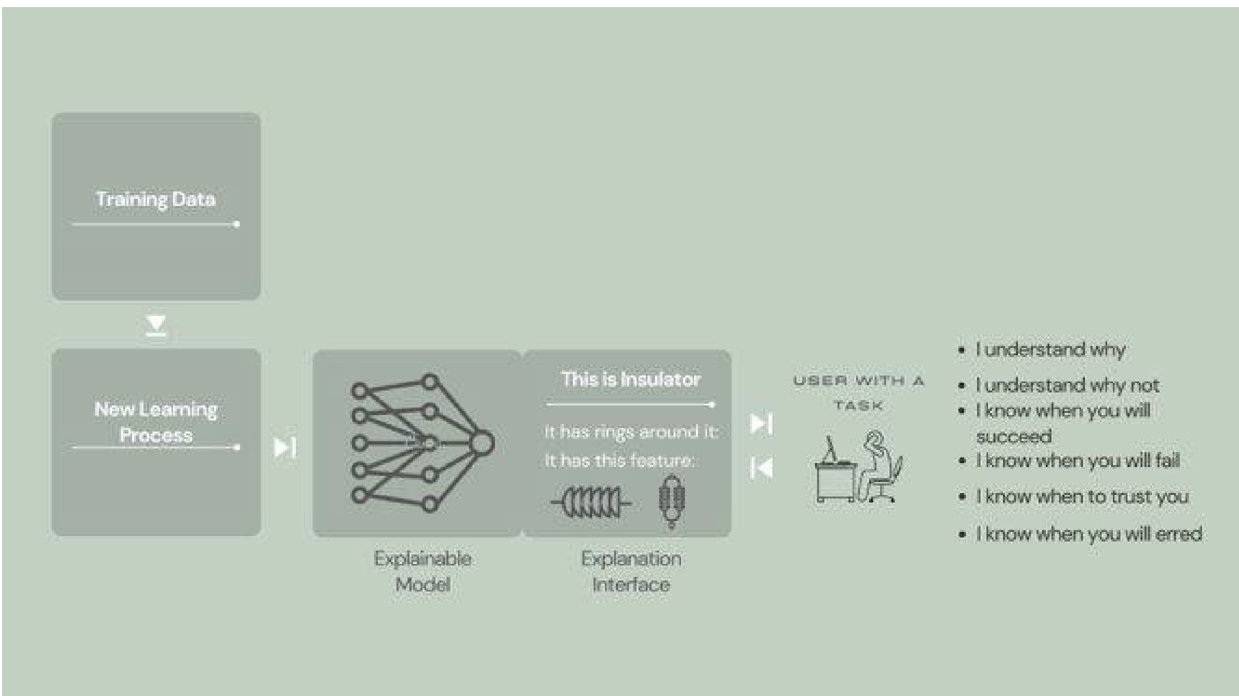


This led me to design a more explainable human-centered AI algorithm, and led to constructing an algorithm by incorporating two of these themes, explainability and human-centeredness, in an AI algorithm. The visual representation of how the proposed algorithm works to establish explainability while detecting anomalies is shown in Figure 16, and will be discussed

more in detail in the next stage of this research, the prototyping phase. In this phase, I discuss how the AI algorithm is constructed and how the dataset is obtained and annotated, and will present how the algorithm is used to train the proposed HC-XAI model. In the latter part of the chapter, I will also discuss how this HC-XAI system is implemented by integrating the model and evaluating it in Stage 3 of this research.

Figure 16

Visual Illustration of the Proposed HC-XAI Algorithm by Incorporating Explainability and Human-Centeredness



Stage 2.2: Prototyping Phase

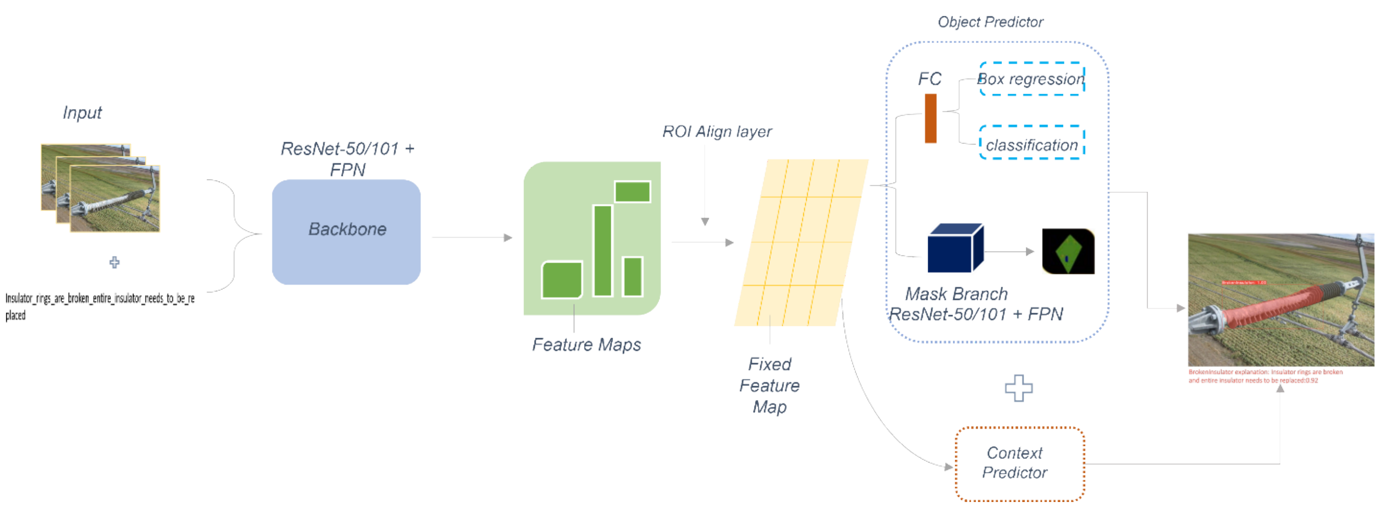
This stage of the research solely focuses on the themes uncovered during data analysis by keeping in mind those two themes, explainability, and human-centered AI, uncovered in Stage 1. I would like to build an HC-XAI system that incorporates these two themes in constructing the algorithm, which aims to investigate the following research question:

RQ1: How might we design a Human-Centered XAI (HC-XAI) system that augments human capabilities in conducting visual inspection for identifying anomalies?

To conduct further investigation in order to answer the research question, I presented a detailed development framework used to construct the explainable AI algorithm, which was integrated into the HC-XAI system in Stage 2.2 and Stage 3 of the research. Before getting into how the system is built using the development framework, first I am going to present the architecture on how this explainable AI algorithm is constructed. The overall architecture of this algorithm is presented in Figure 17.

Figure 17

Overall Architecture of the eXplainable AI algorithm

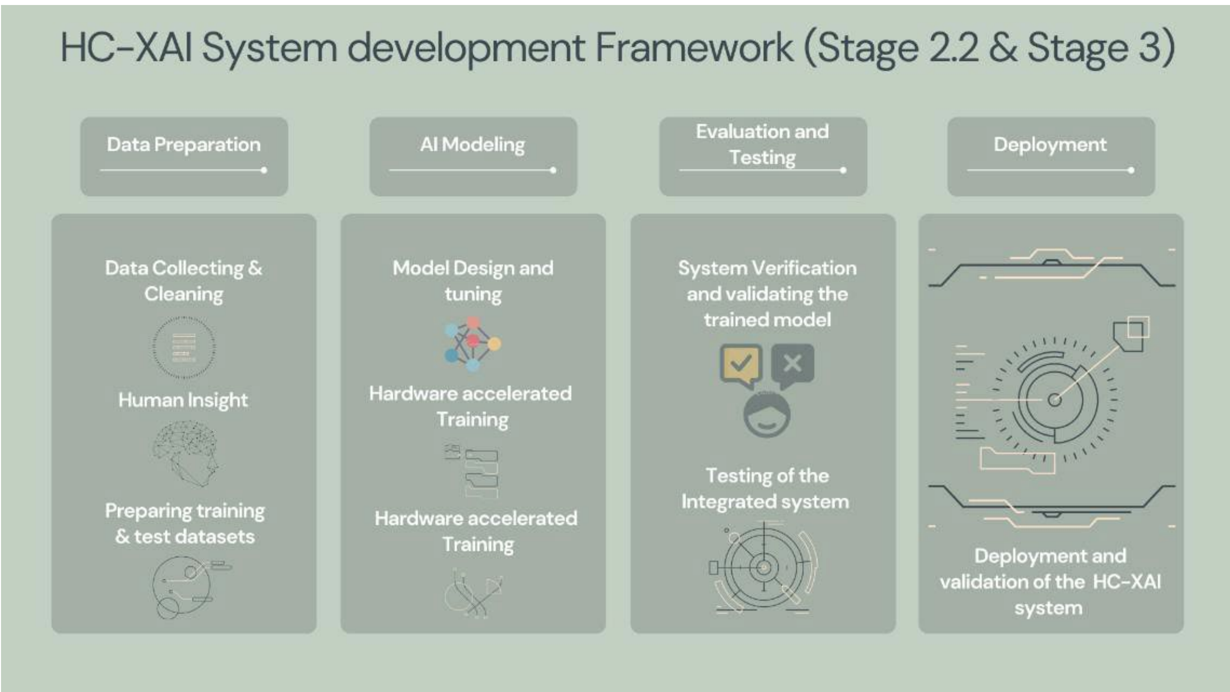


In this architecture the algorithm takes an image and context label as an input and generates a feature map from a Convolutional Neural Network (CNN) backbone. Then unlike Mask-RCNN, region proposal network is discarded and ground truth bounding boxes are directly utilized to extract the object-level representation with the ROI Align layer. Then two ROI features branch into two sibling predictors: an object predictor will take care of the object detection and masking of each class, while the context predictor layer will take care of the

prediction of the context label that is associated with the object class. Based on this architecture, the AI system will be developed with respect to the proposed development framework. The development framework for building this AI system is presented in Figure 18. Every stage of the framework provided an ability for the AI model to learn and improve itself over time. The detailed description of each stage is presented in the following sections of this prototyping phase.

Figure 18

Detailed Development Framework Used for Stage 2.2 and Stage 3 Research



Data Preparation

To develop the algorithm, I had to go through these three steps in this development Stage 2.2. They are the first to develop the algorithm and test a custom dataset. For this step, I had collected the data on the powerline components from CPS Energy. Then after collecting the data, I had to annotate the dataset from the curated list of data and create the dataset to train and test the datasets. After building this custom dataset by annotating the data, I developed the

algorithm as the last step and built an AI model by training the model using the custom dataset. After that, I tested the model's accuracy in terms of explainability and deployed it in an HC-XAI system to attain human centeredness. Each step carried out in this stage is presented below in detail.

Data Collection

The first step in developing the algorithm to investigate the research question was data collection. For this, I contacted CPS Energy to obtain the data needed. Since the data obtained from CPS Energy is intellectual property, I signed a Non-Disclosure Agreement with the energy company by agreeing that the data collected would not be used outside of this research and could not be used for personal purposes. After signing the agreement, I collected the data corresponding to the components, such as powerlines, poles, insulators and transformers. After collecting the data, I curated a list of image data used to construct the dataset, which is divided into two sets. One dataset, containing 80% of all data, was used as a training dataset, and 20% of the curated data served as a testing dataset. The sample type of data that I collected and used for data annotation is shown in Figure 19.

I also took care of the size of the images captured by drones. Since most of the data captured for this research were high quality, the resolution of the images was too high. This took a lot of time to train the data and caused memory issues. To overcome this, I reduced the size of the images and prepared the datasets to handle these issues while training the AI model. After reducing the size of the images and separating the entire dataset to train and test datasets, I moved to the next step in this prototyping phase.

Figure 19

Sample Data Collected

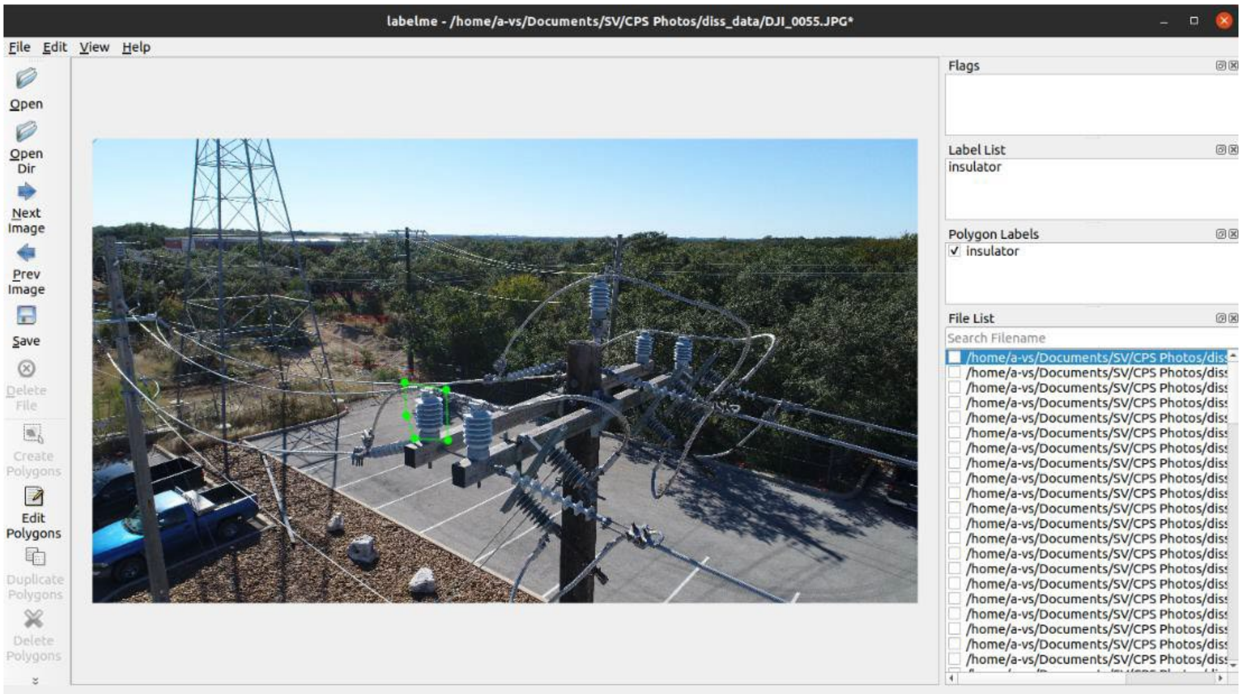
Annotation of the Data and Dataset Preparation

In this step of the prototyping phase, I used the cleaned datasets prepared in the previous step and used the data from the trained and test datasets created. From that, I used the LabelMe opensource software to annotate the data by selecting the anomalies and annotating the explanations and individual components used during inspection to determine whether the components were faulty. An example of how this was achieved using LabelMe is presented in Figure 20.

After finishing annotating the data in both the training and testing dataset folders, I then moved to the next step in constructing the HCAI algorithm, presented in the next section of this chapter.

Figure 20

LabelMe for Annotating the Data Collected



Human-Centered eXplainable AI (HC-XAI) algorithm (AI Modeling)

This section demonstrates how the HC-XAI neural network was created and trained using a custom powerline dataset. The main objective of this HC-XAI neural network was to detect the faulty components of a powerline and explain to the inspection worker why it has come to that prediction. This HCAI neural network's structure comprised three layers: input layer, hidden layer, and output layer. The input layer consisted of seven neurons that acted as inputs for the neural network. The number of neurons used in the hidden layers depended on the experiments that were conducted (1, 2, 3, 5 and 20) to measure the accuracy of the explanation generated and anomalies detected after the model had been trained for 120k epochs, based on trial and error. Out of each, the closest explanation generated and anomalies detected, with an accuracy that was best observed within the three hidden layers, were used in the model training to obtain the output

layer. After training, the testing of this model was observed on the validation dataset and it was observed that there was no considerable difference compared to the number of hidden layers used (2, 3, 5 and 20). For these reasons, three hidden layers were chosen for this HC-XAI neural network. Finally, a single output layer was used to capture the features from the neural network that included x, y, w, and h of the anomalies detected, along with explanations of the detected faulty components.

Stage 3: Implementation and Evaluation Phase

Implementation Phase

In the implementation stage, I presented the results obtained after this HCAI system was implemented, and the test results are presented in Figure 21. From the results, it was evident that the HC-XAI algorithm that was developed was performing as intended. The results suggest that the model identified the anomalies and explained why the model had come to that conclusion. This helped the inspection workers trust the HC-XAI system they were working with while conducting inspections. It also helped in augmenting the capabilities of the humans and made the job easier where inspections are needed, particularly in places where inspections are hard to conduct. In the next step, I evaluated the system with the actual users and evaluated the HC-XAI design.

Evaluation Phase—Usability Testing for Evaluating the System.

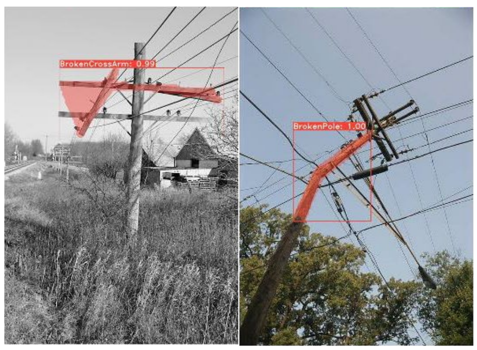
After implementing the HC-XAI system, I chose usability testing for evaluating the HC-XAI system, since it is the proven method for evaluating a system with real people and obtaining actionable insights that help create a better system. Before getting into the details of how usability testing was used in this research, a brief overview of usability testing is presented.

Figure 21

Results of Proposed HC-XAI Algorithm by Incorporating eXplainability and Human-Centeredness

BrokenInsulator explanation: Insulator rings are broken and entire insulator needs to be replaced:0.92

BrokenInsulator explanation: Insulator ring is missing and needs to be replaced:0.89



BrokenCrossArm explanation: Wooden crossarm is broken and needs replacement:0.99

BrokenPole explanation: Wooden crosspole is broken and need immediate attention:0.94

Usability testing is a process of testing a system with real people by observing and noting their interactions with the HC-XAI system. For this usability testing, research planning meetings were conducted with the initial participants from the three companies and were asked to use the HC-XAI system and provide feedback on the aspects of explainability, human-centerednessand what are the aspects that can be improved in the current HC-XAI system. Before starting the process of usability testing informed consent forms were given to each of the three participants and their observation and feedback were collected and recorded in validating the system. This approach allowed me to understand whether the design of the system developed in the prototyping phase was usable and intuitive, in order to conduct inspections using this HC-XAI

system. This approach also brought a holistic look at the participants using this HC-XAI system and helped explore intended and unintended uses of the system.

Several moderating techniques are available to gain insights from usability testing, of which I used the Concurrent Think Aloud (CTA) moderating technique. This moderating technique helped me to understand participants' thoughts during their interaction with the system. Before starting the session, I asked all the participants to sign an informed consent agreement, and explained what part of the HC-XAI system would be tested. The participants were asked to focus on system explainability and how this collaborative system would help them conduct inspections. I noted the participants' behaviors, comments, and suggestions about the task of HC-XAI system explainability, and how this collaborative system could assist in conducting inspections. After evaluating the HC-XAI system using this CTA moderating technique and recording the evaluation by note-taking sessions from the pilot system, I learned some key insights from the participants to improve the current system, which included:

- Users felt that the model expandability feature could be more intuitive by adding user feedback to update the explainability feature on the anomalies detected.
- Users felt that suggestions of what needs to be done when such failure is identified during the inspection were needed.
- Users felt that integration of this HC-XAI system with the drone would give more flexibility for the inspection workers.

These are some of the vital significant insights that evolved from the evaluation of the HC-XAI system with real people for whom the system was designed to make AI more collaborative and play a collaborative role in augmenting human intelligence. Apart from these, participants agreed that the HC-XAI system was more human-centric and explainable than the current system that they have worked with over the past year.

From these three stages (Stage 2.1 redefine phase; Stage 2.2 prototyping phase; Stage 3 implementation and evaluation phase) of this research study, I created a development framework that could be used in determining how one might design a HC-XAI system that augments human capabilities in conducting visual inspection to identify anomalies, the main research question of this research. Chapter 6 will discuss how this HC-XAI system will foster social innovation through a shared and collaborative approach, and how one might create value for an organization.

Conclusion

This chapter was mainly focused on answering the main research question of this study. In the process, it provided a detailed overview of how the remaining phases contributed to doing so, after finishing data analysis in Stage 1. Based on the results obtained from Stage 1, this chapter explains how the rest of the stages (Stage 2.1 redefine phase; Stage 2.2 prototyping phase; and Stage 3 implementation and evaluation phase) helped me to answer the main research question of this study—how might we design and construct a human centered explainable AI algorithm that not only detects anomalies but also provides explanations to the humans so they can take action when conducting powerline inspections? This chapter also provides insights from the implementation and evaluation phase about this HC-XAI system and how it can be improved. In the next chapter, I will present the findings from this HC-XAI system and discuss how this system will foster social innovation and sustainability through a shared and collaborative approach, thus augmenting human capabilities, and how one might create value for an organization, the sub research question of this study.

Chapter 6: Discussion, Findings and Future Work

This chapter focuses on presenting the problem which this study is based on and how the research questions, the findings, and interpretations from these findings create open discussion for further/future work. In this final chapter, I mainly focus on answering the sub-research question by presenting interpretation, findings, and discussions. The interpretation, findings, and discussion section of this chapter presents investigations on how this HC-XAI design will foster social innovation and sustainability, and the factors that need to be considered to create value in the HC-XAI design. Finally, I end the chapter by briefly presenting the findings from this research and then discussing future work I plan to do by creating a visual playbook that can guide common principles in designing human-centered explainable artificial intelligence based on a developmental framework and modified framework of innovation.

Discussion, Interpretation, and Synthesis

In this section of the document, I would like to talk about the findings from each stage of my conceptual framework and how these findings helped me in investigating the main research question and its sub-research question. The three key findings that emerged from Stage I and Stage II.1 of this research are as follows:

- Lack of explainability with current intelligent systems.
- Lack of human-centeredness in the AI systems.
- The need for a rationalistic approach by looking into changing the process of conducting inspections in the long run by bringing in a new, innovative approach.

This idea of AI augmenting humans instead of replacing them is one of the key objectives behind the development of HCAI. Xu (2019) mentioned how important it is for HCAI solutions to be ethical, explainable, comprehensible, and useful (Xu, 2019). The following three

key findings proves that how important it is to incorporate the factors of explainability to attain trust in AI systems and centered around humans when developing the intelligent systems that augments human's capabilities. Especially developing these XAI system calls for confidence, safety, security, privacy, ethics, fairness, and trust (Kieseberg et al., 2016), which brings usability and Human-AI interaction into a new and much more important focus (Miller, 2019). These kinds of XAI systems complement inspection professionals, that can play a huge role in augmenting the human's role in the inspection process, which leads to a safer and quicker decision-making process and builds better trust and explainability in a more human-centric approach in conducting powerline inspection.

These three takeaways functioned as a building block for me to move into the next stage of this research study. These findings were compared with the initial problem statement that I assumed when I started the research study. I observed that there was not much of a difference in the problem, which then helped me to move to the next stage of the study by taking these findings and investigating the main research question of the study, which is:

RQ1: How might we design a HC-XAI system that helps augment human capabilities in conducting visual inspection for identifying anomalies?

This led me to the next stage of my research study by taking the pain points from the previous stage and then heading over to the prototyping, implementation, and evaluation phases. First, I will discuss the prototyping phase in which a HC-XAI system development framework is proposed to address the first two findings that came out in Stage I, which is lack of explainability and human-centeredness in AI Systems. To address these, the proposed development framework consisted of four phases, Data Preparation, AI modeling, Evaluation and Testing, and Deployment. The first two of these four phases took place in the Prototyping and implementation

phase, in which data was collected from CPS Energy and annotated to determine the faults that were determined by the inspection workers and explanations to the dataset were added. Once the data was annotated it was divided into two sets, one a training set into which 80% of the data went, 10% fell to testing, and the remaining 10% fell into evaluation datasets. After preparing the datasets for the model, an explainable AI algorithm was developed in Python to train the AI model to detect anomalies by providing explanations and complete instance segmentation on the anomalies detected. The algorithm used to train the model is presented in a pseudo form in Figure 22.

After prototyping and implementing the algorithm, the AI model was trained on 400K iterations to make sure the model was behaving as intended and tested for every 100K steps to evaluate if the model was training properly. After taking into considerations the learning rate and the number of times the model was trained, I moved to the next phase of the development framework, the evaluation phase, in which I took the trained model and evaluated it using Concurrent Think Aloud (CTA) moderating technique usability testing. While evaluating the model I specifically evaluated the aspects of explainability and human-centeredness. In particularly focusing on system explainability and how this collaborative system helps in conducting inspections by observing the participant's behavior, comments and suggestions were recorded in Zoom from the participants. From this phase I observed three main insights that emerged that helped in answering the main research question (RQ1) of this study, as follows:

1. Model explainability feature could be more intuitive by adding user feedback to update the explainability feature on the anomalies detected.

Figure 22

Code Snippets of the HC-XAI algorithm

```
# ---------------------- DATASETS ---------------------- #

dataset_base = Config({
    'name': 'Base Dataset',

    # Training images and annotations
    'train_images': './data/coco/images/',
    'train_info':   'path_to_annotation_file',

    # Validation images and annotations.
    'valid_images': './data/coco/images/',
    'valid_info':   'path_to_annotation_file',

    # Whether or not to load GT. If this is False, eval.py quantitative evaluation won't work.
    'has_gt': True,

    # A list of names for each of you classes.
    'class_names': COCO_CLASSES,

    # COCO class ids aren't sequential, so this is a bandage fix. If your ids aren't sequential,
    # provide a map from category_id -> index in class_names + 1 (the +1 is there because it's 1-
indexed).
    # If not specified, this just assumes category ids start at 1 and increase sequentially.
    'label_map': None
})
 Powerline_COCO_CLASSES = ("CrossArm", "ElectricPowerPole", "Insulator", "Primarywire",
 "Transformer", "BrokenCrossArm", "BrokenInsulator", "BrokenPole")

 Powerline_COCO_LABEL_MAP = { 0:  1,  1:  2,  2:  3,  3:  4,  4:  5, 5:  6,  6:  7,  7:  8}

 Powerline_dataset = dataset_base.copy({
   'name': 'VS-Powerlinedetection',
   'train_info': '/home/sv/Documents/CPS/RPDS-master/powerlinedataset/train/trainval.json',
   'train_images': '/home/sv/Documents/CPS/RPDS-master/powerlinedataset/train/images/',
   'valid_info': '/home/sv/Documents/CPS/RPDS-master/powerlinedataset/val/testval.json',
   'valid_images': '/home/sv/Documents/CPS/RPDS-master/powerlinedataset/val/images/',
   'class_names': Powerline_COCO_CLASSES,
   'label_map': Powerline_COCO_LABEL_MAP
 })
  # ---------------------- TRANSFORMS ---------------------- #

  resnet_transform = Config({
      'channel_order': 'RGB',
      'normalize': True,
      'subtract_means': False,
      'to_float': False,
  })
```

```
# ---------------------- ACTIVATION FUNCTIONS ---------------------- #

activation_func = Config({
    'tanh':    torch.tanh,
    'sigmoid': torch.sigmoid,
    'softmax': lambda x: torch.nn.functional.softmax(x, dim=-1),
    'relu':    lambda x: torch.nn.functional.relu(x, inplace=True),
    'none':    lambda x: x,
})
    # ---------------------- BACKBONES ---------------------- #

    backbone_base = Config({
        'name': 'Base Backbone',
        'path': 'path/to/pretrained/weights',
        'type': object,
        'args': tuple(),
        'transform': resnet_transform,

        'selected_layers': list(),
        'pred_scales': list(),
        'pred_aspect_ratios': list(),

        'use_pixel_scales': False,
        'preapply_sqrt': True,
        'use_square_anchors': False,
    })
            resnet50_config = base_config.copy({
                'name': 'resnet50',

                'backbone': resnet50_backbone.copy({
                    'selected_layers': list(range(1, 4)),

                    'pred_scales': base_config.backbone.pred_scales,
                    'pred_aspect_ratios': base_config.backbone.pred_aspect_ratios,
                    'use_pixel_scales': True,
                    'preapply_sqrt': False,
                    'use_square_anchors': True, # This is for backward compatability
                }),
            })
    resnet50_Powerline_config = resnet50_config.copy({
        'name': 'resnet50_Powerline',
        # Dataset stuff
        'dataset': Powerline_dataset,
        'num_classes': len(Powerline_dataset.class_names) + 1,

        # Image Size
        'max_size': 416,
    })
```

```
resnet101_backbone = backbone_base.copy({
    'name': 'ResNet101',
    'path': 'resnet101_reducedfc.pth',
    'type': ResNetBackbone,
    'args': ([3, 4, 23, 3],),
    'transform': resnet_transform,

    'selected_layers': list(range(2, 8)),
    'pred_scales': [[1]]*6,
    'pred_aspect_ratios': [ [[0.66685089, 1.7073535, 0.87508774, 1.16524493, 0.49059086]] ] * 6,
})

resnet101_gn_backbone = backbone_base.copy({
    'name': 'ResNet101_GN',
    'path': 'R-101-GN.pkl',
    'type': ResNetBackboneGN,
    'args': ([3, 4, 23, 3],),
    'transform': resnet_transform,

    'selected_layers': list(range(2, 8)),
    'pred_scales': [[1]]*6,
    'pred_aspect_ratios': [ [[0.66685089, 1.7073535, 0.87508774, 1.16524493, 0.49059086]] ] * 6,
})

resnet101_dcn_inter3_backbone = resnet101_backbone.copy({
    'name': 'ResNet101_DCN_Interval3',
    'args': ([3, 4, 23, 3], [0, 4, 23, 3], 3),
})

resnet50_backbone = resnet101_backbone.copy({
    'name': 'ResNet50',
    #'path': 'resnet50-19c8e357.pth',
    'path': 'yolact_resnet50_Powerline_25678_282462_interrupt.pth',
    'type': ResNetBackbone,
    'args': ([3, 4, 6, 3],),
    'transform': resnet_transform,
})

resnet50_dcnv2_backbone = resnet50_backbone.copy({
    'name': 'ResNet50_DCNv2',
    'args': ([3, 4, 6, 3], [0, 4, 6, 3]),
})
```

2. Participants felt there should be a suggestions system of what needs to be done when such failure happens during the inspection.
3. Integration of this HC-XAI system with a drone would provide more flexibility for inspection workers.

These findings from the evaluation phase of the development framework suggested that the system is explainable and human centered, as the system learns from human input and collaboration by saving the data that it has never seen to the local drive, which helped to answer the RQ1 of this research study. AI's explainability in an intelligent system would enhance the needed trust factor that is lacking between humans and AI systems, especially in visual inspections in energy companies. Sometimes it is assumed that humans always explain their decisions, but it is not often the case due to various heterogeneous and vast information sources. Hence XAI calls for confidence, safety, security, privacy, ethics, fairness, and trust (Kieseberg et al., 2016), which brings usability and Human-AI interaction into a new and much more important focus (Miller, 2019).

The last finding obtained at the end of Stage I of this research study is the need for this HC-XAI system to use a rationalistic approach by looking into changing the process of conducting inspections for the long run by bringing a new innovative approach and fostering social innovation and sustainability. The investigation of this finding was guided by using the following sub-research question:

RQ1a: How might this HC-XAI design foster social innovation and sustainability through this shared and collaborative approach?

What is Social Innovation?

Before answering the sub-research question, I would like to talk about what innovation is and how it can be defined. Social innovation is a term that is used to describe new products or services, or new combinations of social practices that are aimed at meeting emerging or previously neglected societal needs (Caulier-Grice et al., 2012). However, the concept of innovation addresses not only having new, but also relatively new, patterns of action. As there

are several definitions of innovation, I would like to define innovation, simply, as new ideas that work. Social innovation can be defined as new ideas that work for social goals, which can be further narrowed down as innovative ideas or activities and services that are predominantly developed and diffused through organizations whose primary purposes are social (Akrich et al., 2002). This approach ultimately differentiates social innovation from business innovation. In contrast, business innovation is mainly motivated by profit maximization, and social innovation is primarily aimed at producing replicable programs or models in an organization for the benefit of the society, to name a few - 3D-printed homes, liquid nano clay that can grow crops in deserts (Sutton, 2020).

How Social Innovation Happens (Stages of Innovation)

The story of change comes with the interaction between the innovators and the environment within which they are working, in which new ideas must secure support if they are to survive. On the other hand, social change depends on the alliances between the bees and trees, where bees are the small organizations, individuals, and groups who have new ideas that are mobile, quick, and able to cross-pollinate. The trees are the more prominent organizations—governments, companies, or big NGOs where there is a lack of creativity, but where they are generally good at implementation and have the resilience, roots and scale to make things happen. Each is dependent on the other, and most social change comes from these alliances between leaders and groups. An idea must pass through several stages of innovation to foster social innovation. Some organizations have developed formal creative methods, such as Edward de Bono’s Six Hats, which are now widely used (De Bono et al., 1970). The different methods used by IDEO design company and What If consultant are all aimed at freeing groups to think laterally and detect new patterns. The simulation of creative ideas can take place from

conversations that happen from other people by collecting the good ones and incorporating them and by eliminating the bad ones. All innovation processes include taking promising ideas and testing them in practice (Mulgan, 2006). He also discusses in his article that the process of social innovation occurs in three stages, where a creative or innovative idea should pass through: Generating ideas by understanding needs and identifying potential solutions; developing, prototyping, and piloting ideas; and assessing, then scaling up and diffusing the good ones.

Generating Ideas by Understanding Needs and Identifying Potential Solutions

This is the first stage of social innovation, identifying needs that are not being met and developing some ideas of how they can be met in relation to a problem (Mulgan, 2006). Sometimes needs are glaringly obvious, like hunger, homelessness, or disease. However, sometimes needs are less obvious or not recognized, like the need for protection from domestic violence or racism, and it takes campaigners and movements to name and define these. Needs come to the fore in many ways—through angry individuals and groups, campaigns, and political movements, as well as through careful observation. In this research, needs were carefully observed during the time I worked closely with the energy company, over a year. During that time I developed empathy and had a better understanding of the needs that the energy company was having during inspections, and started off with a presumption of the problem statement, which is one of the effective methods for cultivating social innovation. This presumption of the problem statement led me to look for positive deviants of what might be possible if an HC-XAI innovative solution was brought to inspection workers. In this way, I generated ideas by putting into research questions the needs that were tied to new possibilities of exploring the capability of adding explainability and human centeredness.

Developing, Prototyping, and Piloting Ideas

The second phase of any social innovation process involves taking those promising ideas and testing them out in practice (Mulgan, 2006). Few plans survive their first encounter with reality wholly intact; however, it is through action that they evolve and improve. Social innovations may be helped by formal market research or desk analysis, but progress is often achieved more quickly by turning the idea into a prototype or pilot and then galvanizing enthusiasm. Social innovations are often implemented early. Because those involved are usually highly motivated, they are too impatient to wait for governments or professions to act. This experience of trying to make their ideas work speeds up their evolution, and the power of example then turns out to be as persuasive as written argument or advocacy.

In this case, I carried out research to investigate how we might design a Human-Centered XAI (HC-XAI) system that augments human capabilities in conducting visual inspections for identifying anomalies. In the process of investigating the possibility of having this kind of innovative idea, I conducted research in two stages by following the framework of innovation. The first stage involved discovering whether the initial presumption of the problem was true; if so, what were the needs and opinions of the practitioners related to inspection? This analysis made me realize that the problem statement was accurate and I continued to investigate the research question, which was developing the HC-XAI system by prototyping the system and piloting the idea by evaluating and validating the system using usability testing of the ideas in protected conditions, halfway between the real world and the laboratory. An essential virtue of quick prototyping is that innovations often require several tries before they work. The first outcomes are invariably flawed, and in the social field, parallel methods were being developed to crystallize promising ideas so that they could be tested quickly.

Assessing, Then Scaling Up and Diffusing the Good Ones

The third stage of the social innovation process comes when an idea is proving itself in practice and can then be grown, potentially through organic growth, replication, adaptation, or franchising (Mulgan, 2006). Usually, innovations spread in an S curve, with an early phase of slow growth amongst a small group of committed supporters, then a phase of rapid take-off, and then a slowing down as saturation and maturity are achieved. In this research, after developing, prototyping, and piloting the explainable algorithm, I learned key insights of where the HC-XAI system could be improved to scale up and incorporate new critical insights obtained from the usability testing, and offered the potential of growth without too much managerial responsibility. This type of growth is most attractive for social innovators.

This research was carried out by following the stages of how social innovation happens. When comparing the research stages that were completed in investigating the main research question, I identified the needs of an energy company by understanding the needs of inspection and identifying possible solutions that might develop new and innovative ways by incorporating an explainable AI system that augments human capabilities. This idea then moved to the next stage of research by moving to prototype, implementation, and evaluation phases of research. I was able to show that an explainable AI system can be developed that is human-centric, by proving that a new innovative idea or service that is motivated by meeting the goal of social need can be predominantly developed and diffused through organizations whose primary purposes are social.

Sustainability

When it comes to sustainability, first, one must look at the goals of sustainable development that are multi-faceted and spread across many spheres of human life (Cobbinah et

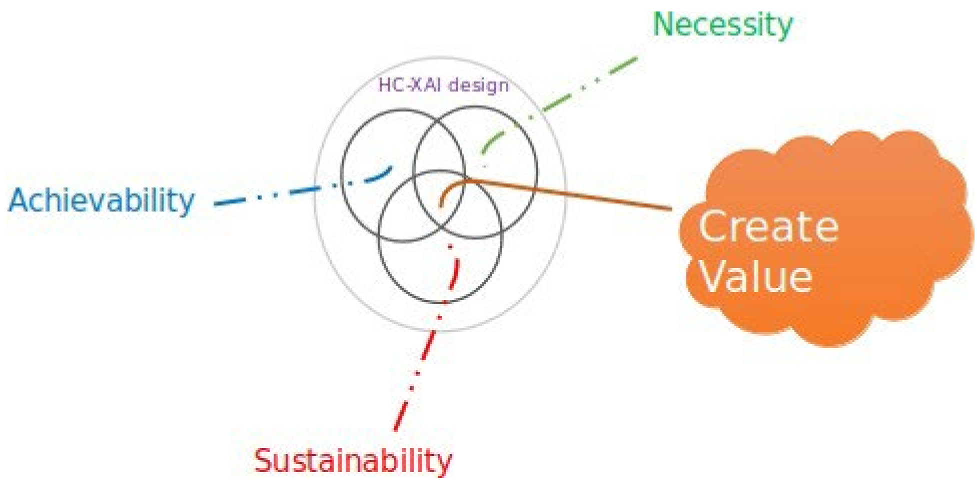
al., 2015; Moore, 2015). The ultimate purpose of achieving sustainability is to improve humanity's socio-economic wellbeing by creating an environment conducive for citizens to develop their full potential and thus live productive lives (Moore, 2015). Investigating the research question and being able to design the HC-XAI algorithm that augments humans reduced the risk of committing mistakes and enhanced productivity, and allowed more focus on important tasks. This encompassed an integrated and somewhat intertwined development goal, such as protecting the natural environment, promoting education, production, consumption, and the wellbeing of citizens (Addison et al., 2015; Coscieme et al., 2021). It can be assumed that the HC-XAI system that is developed to investigate the main research question will foster social innovation and sustainability. I also learned one more thing during the journey of this research, which was that it was significant to consider these three main factors whenever one is designing an HC-XAI system, in order to create value to the designed HC-XAI system. Those are necessity, achievability, and sustainability. The visual representation of the value creation, with three main factors that need to be considered when designing an HC-XAI system, is shown in Figure 23.

Recommendations for Future Work

Despite much interest in the area of XAI, evaluating these explainable models is still a topic that has not been solved and requires further research. Unlike ML models where there is a ground truth that can be used for model evaluation, there is currently no commonly agreed definition of what constitutes a right explanation and what are the properties that an explainable model should satisfy. Because of this lack of standardization assessment techniques of these XAI, where most of the explanations primarily comes from humans assuming that humans know what an accurate explanation would or should look like in explaining things. This is occasionally

Figure 23

Three Factors to Consider In Designing an HC-XAI System to Create Value



true but not true all the time, so when it comes to this research study as well, the explanations that the model predicts are mostly from the inspection workers. This is the major limitation in producing XAI models for any industry.

Also, according to van Wynsberghe (2021), there are two branches that need to be considered when measuring sustainability in AI. Those are AI for Sustainability (AI4good, AI4Climate) and Sustainability of AI (reusable data, reduce carbon emissions from training AI) which are in their infancy stages. To address the sustainability of AI, these two branches should be addressed simultaneously.

In terms of future work, especially after conducting the usability testing in Stage 3 of this research study, I found some critical insights that the initial pilot HC-XAI design should be improved upon, which allows redesigning and improving in the future. I also want to build a visual codebook by providing common design principles to construct more human-centered explainable AI systems that augment humans, instead of creating confusion or threats because of the development of AI.

Summary and Conclusion

This chapter presented how the research questions were answered and discussed the research findings. In this chapter, I discussed how the data analysis from Stage 1 helped me look at the problem statement. From the data analysis stage, the evolved themes played a significant role in Stage 2.1 of this research study, where I had a chance to revisit the original problem statement. The redefine phase (Stage 2.1) allowed me to look at the presumed problem statement with the derived themes that evolved. After this redefine phase, I used the same problem statement, since there was not much alteration needed in terms.

I then moved on to the next phase of this research study, the prototyping phase (Stage 2.2). This prototyping phase was critical to this research study, because it was used to investigate one of the research questions: how we might design HC-XAI systems that augment human capabilities in conducting visual inspections to identify anomalies. In this stage, I proposed a common development framework for designing an HC-XAI system, which helped me answer this study's main research question. The proposed development framework provided common steps to design any HC-XAI system that consists of various phases, including the data preparation phase, AI modeling phase, evaluation phase, and testing and deployment phase. Of these four phases in the development framework, two, data preparation and AI modeling, were carried out in the prototyping phase of this research study. The evaluating and testing phase and the deployment phase were carried out in Stage 3 of this research study.

These phases helped me to answer the main research question of this study, which led to investigating the sub-research question: how this HC-XAI design might foster social innovation and sustainability through a shared and collaborative approach. To answer this sub-research question, I investigated what innovation is and the different types of innovation, then discussed

how social innovation happens and the different stages that it possesses, by mapping each stage of social innovation against current research study phases. This helped me in answering the sub-research question, and I learned that when designing any HC-XAI system, three key factors need to be considered to create value—necessity, achievability, and sustainability - to achieve rationalist or incrementalist strategies or innovation. These factors bring value to any HC-XAI design. They will foster social innovation by following the modified framework of innovation, by mapping how social innovation happens in stages, following the stages in the current research study.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowiz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*. https://doi.org/10.48550/arXiv.1603.04467

Addison, T., Niño-Zarazúa, M., & Tarp, F. (2015). Aid, social policy and development. *Journal of International Development*, *27*(8), 1351-1365. https://doi.org/10.1002/jid.3187

Adou, M. W., Xu, H., & Chen, G. (2019, October). Insulator faults detection based on deep learning. In *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)* (pp. 173-177). IEEE. http://doi/10.1109/ icasid.2019.8925094

Akrich, M., Callon, M., Latour, B., & Monaghan, A. (2002). The key to success in innovation part I: The art of interessement. *International jJournal of Innovation Management*, *6*(02), 187-206. https://doi.org/10.1142/S1363919602000550

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh. J., Iqbal. S., Bennett, N. P., Inkpen. K., Teevan. J., Gil, K. R., & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3290605.3300233

Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc. https://doi.org/10.1561/2200000006

Bologna, G., & Hayashi, Y. (2017). Characterization of symbolic rules embedded in deep DIMLP networks: A challenge to transparency of deep learning. *Journal of Artificial Intelligence and Soft Computing Research*, *7*. https://doi.org/10.1515/jaiscr-2017- 0019

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573-1576. https://doi.org/10.1126/science.aaf2654

Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2014, March). Detecting price manipulation in the financial market. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (pp. 77-84). IEEE. http://dx.doi.org/10.1109/CIFEr.2014.6924057

Caulier-Grice, J., Davies, A., Patrick, R., & Norman, W. (2012). *Defining social innovation*. A deliverable of the project: "The theoretical, empirical and policy foundations for building social innovation in Europe" (TEPSIE), European Commission–7th Framework Programme, Brussels: European Commission, DG Research, 22. The Young Foundation.

Charmaz, K. (2014). *Constructing grounded theory*. Sage.

Cobbinah, P. B., Erdiaw-Kwasie, M. O., & Amoateng, P. (2015). Rethinking sustainable development within the framework of poverty and urbanisation in developing countries. *Environmental Development*, *13*, 18-32. https://doi.org/10.1016/j.envdev.2014.11.001

Coscieme, L., Mortensen, L. F., & Donohue, I. (2021). Enhance environmental policy coherence to meet the sustainable development goals. *Journal of Cleaner Production*, *296*, 126502. https://doi.org/10.1016/j.jclepro.2021.126502

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., & Traore, D. (2019). Deep neural networks with transfer learning in millet crop images. *Computers in Industry*, *108*, 115-120. https://doi.org/10.1016/j.compind.2019.02.003

Cusano, C., & Napoletano, P. (2017). Visual recognition of aircraft mechanical parts for smart maintenance. *Computers in Industry*, *86*, 26-33. https://doi.org/10.1016/j.compind.2017.01.001

Crevier, D. (1993). Expert systems as design aids for artificial vision systems: A survey. *Intelligent Robots and Computer Vision XII: Algorithms and Techniques*, *2055,* 84-96. https://doi.org/10.1117/12.150131

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.

De Bono, E., & Zimbalist, E. (1970). *Lateral thinking*. Penguin.

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, *24*(9), 1342-1350. https://doi.org/10.1038/s41591-018-0107-6

Dias, P. A., Tabb, A., & Medeiros, H. (2018). Apple flower detection using deep convolutional networks. *Computers in Industry*, *99*, 17-28. https://doi.org/10.1016/j.compind.2018.03.010

Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable AI really mean? A new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*. https://doi.org/10.48550/arXiv.1710.00794

Dreyfus, H. (1992). Heidegger's history of the being of equipment. In H.L. Dreyfus and H. Hall (Eds.) *Heidegger: A Critical Reader.* Blackwell.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://dx.doi.org/10.1038%2Fnature21056

Filonenko, A., Hernández, D. C., & Jo, K. H. (2017). Fast smoke detection for video surveillance using CUDA. *IEEE Transactions on Industrial Informatics*, *14*(2), 725- 733. https://doi.org/10.1109/TII.2017.2757457

Freire, P. (2000). *Pedagogy of the oppressed*. Continuum.

Gandhi, N., Armstrong, L. J., & Nandawadekar, M. (2017, April). Application of data mining techniques for predicting rice crop yield in semi-arid climatic zone of India. In *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (pp. 116-120). IEEE. https://doi.org/10.1109/TIAR.2017.8273697

Gao, F., Wang, J., Kong, Z., Wu, J., Feng, N., Wang, S., Hu, P., Li, Z., Huang, H., & Li, J. (2017, December). Recognition of insulator explosion based on deep learning. In *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* (pp. 79-82). IEEE. https://doi.org/10.1109/ICCWAMTIP.2017.8301453

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine Publishing Company.

Gu, I. Y., Sistiaga, U., Berlijn, S. M., & Fahlström, A. (2009a). Intelligent video surveillance for detecting snow and ice coverage on electrical insulators of power transmission lines. In *International Conference on Computer Analysis of Images and Patterns* (pp. 1179-1187). Springer. https://doi.org/10.1007/978-3-642-03767- 2_143

Gu, I. Y., Sistiaga, U., Berlijn, S. M., & Fahlström, A. (2009b). Online detection of snow coverage and swing angles of electrical insulators on power transmission lines using videos. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 3249-3252). IEEE. https://doi.org/10.1109/ICIP.2009.5413984

Grudin, J. (2009). AI and HCI: Two fields divided by a common focus. *AI Magazine*, 30(4), 48-48. https://doi.org/10.1609/aimag.v30i4.2271

Gray, D. E. (2013). *Doing research in the real world*. Sage.

Hammersley, M. (1992). *What's wrong with ethnography?* Routledge.

Holzinger, A. (2019). Introduction to MAchine Learning & Knowledge Extraction (MAKE). *Mach. Learn. Knowl. Extr.*, *1*(1), 1-20. http://dx.doi.org/10.3390/make1010001

IEA, S. (2017). International Energy Agency, 2016. *Key World Energy Statistics, ed*.

Jenssen, R., & Roverso, D. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems*, *99*, 107-120. https://doi.org/10.1016/j.ijepes.2017.12.016

Johnson, W. L. (1994, August). Agents that learn to explain themselves. *AAAI* (pp. 1257-1263). https://www.aaai.org/Papers/AAAI/1994/AAAI94-193.pdf

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kang, G., Gao, S., Yu, L., & Zhang, D. (2018). Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning. *IEEE Transactions on Instrumentation and Measurement*, *68*(8), 2679- 2690. https://doi.org/10.1109/TIM.2018.2868490

Katrasnik, J., Pernus, F., & Likar, B. (2009). A survey of mobile robots for distribution power line inspection. *IEEE Transactions on power delivery*, *25*(1), 485-493. https://doi.org/10.1109/TPWRD.2009.2035427

Kieseberg, P., Weippl, E., & Holzinger, A. (2016). Trust for the doctor-in-the-loop. *ERCIM News*, *104*(1), 32-33.

Lacave, C., & Díez, F. J. (2002). A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review*, *17*(2), 107-127. https://doi.org/10.1017/S026988890200019X

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16001837

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31-57.

Markoff, J. (2005). *What the dormouse said: How the sixties counterculture shaped the personal computer industry*. Penguin Books.

McCarthy, J. (1960). *Programs with common sense* (pp. 300-307). RLE and MIT Computation Center.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1984). *Machine Learning: An Artificial Intelligence Approach (Symbolic Computation)* (1983 ed.). Springer. https://doi.org/10.1002/zamm.19850651113

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Mulgan, G. (2006). The process of social innovation. *innovations*, *1*(2), 145-162.

Muhammad, K., Khan, S., Elhoseny, M., Ahmed, S. H., & Baik, S. W. (2019). Efficient fire detection for uncertain surveillance environment. *IEEE Transactions on Industrial Informatics, 15(5),* 3113-3122. https://doi.org/10.1109/TII.2019.2897594

Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1-15. https://doi.org/10.1016/j.dsp.2017.10.011

Moore, H. L. (2015). Global Prosperity and Sustainable Development Goals. *Journal of International Development*, *27*(6), 801–815. https://doi.org/10.1002/jid.3114

Mulvenna, M., Zheng, H., Bond, R., McAllister, P., Wang, H., & Riestra, R. (2017, November). Participatory design-based requirements elicitation involving people living with dementia towards a home-based platform to monitor emotional wellbeing. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2026-2030). IEEE. https://doi.org/10.1109/BIBM.2017.8217972

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development*, *2*(4), 320-335. https://doi.org/10.1147/rd.24.0320

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research, 49*(1), 3-42. https://doi.org/10.1177%2F0049124117729703

Nushi, B., Kamar, E., & Horvitz, E. (2018, June). Towards accountable ai: Hybrid human-machine analyses for characterizing system failure. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 6,* 126-135. Retrieved from https://ojs.aaai.org/index.php/HCOMP/article/view/13337

Park, K. C., Motai, Y., & Yoon, J. R. (2017). Acoustic fault detection technique for high- power insulators. *IEEE Transactions on Industrial Electronics*, *64*(12), 9699-9708. https://doi.org/10.1109/TIE.2017.2716862

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Poole, D, Mackworth, A, & Goebel, R. (1998). *Computational intelligence: A logical approach.* Oxford University Press.

Prates, R. M., Cruz, R., Marotta, A. P., Ramos, R. P., Simas Filho, E. F., & Cardoso, J. S. (2019). Insulator visual non-conformity detection in overhead power distribution lines using deep learning. *Computers & Electrical Engineering*, *78*, 343-355. https://doi.org/10.1016/j.compeleceng.2019.08.001

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169. https://doi.org/10.1007/BF01405730

Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence*. Prentice Hall.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, *36*(4), 105-114. https://doi.org/10.1609/aimag.v36i4.2577

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, *23*(3-4), 351-379. https://doi.org/10.1016/0025-5564(75)90047-4

Snape, D., Spencer, L. (2003). The foundations of qualitative research. In Ritchie, J., Lewis, J. (Eds.), *Research practice: A guide for social science students and researchers* (pp. 1–23). Sage.

Sutton, A. (2020, July 27). *50 social impact innovations that might save the world*. Board of Innovation. https://www.boardofinnovation.com/blog/50-social-impact-innovations-helping-to-save-the-world/

Swartout, W., Paris, C., & Moore, J. (1991). Explanations in knowledge systems: Design for explainable expert systems. *IEEE Expert*, *6*(3), 58-64. https://doi.org/10.1109/64.87686

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279-1285. https://doi.org/10.1126/science.1192788

Nguyen, V.N., Jenssen, R., & Roverso, D. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems*, *99*, 107-120. https://doi.org/10.1016/j.ijepes.2017.12.016

Winograd, T. (Ed.). (1996). *Bringing design to software*. ACM.

Winograd, T., Flores, F., & Flores, F. F. (1986). *Understanding computers and cognition: A new foundation for design*. Intellect Books.

World Economic Forum. *The New Physics of Financial Services – How artificial intelligence is transforming the financial ecosystem*. (2018, August 15). https://www.weforum.org/reports/the-new-physics-of-financial-services-how-artificial-intelligence-is- transforming-the-financial-ecosystem.

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*(3), 213-218. https://doi.org/10.1007/s43681-021-00043-6

Vemula, S., & Frye, M. (2020a). Mask R-CNN powerline detector: A deep learning approach with applications to a UAV. In *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)* (pp. 1-6). IEEE. https://doi.org/10.1109/DASC50938.2020.9256456

Vemula, S., & Frye, M. (2020b). Real-time powerline detection system for an unmanned aircraft system. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 4493-4497). IEEE. https://doi.org/10.1109/SMC42975.2020.9283354

Vemula, S., Marquez, S., Avila, J. D., Brouwer, T. A., & Frye, M. (2021). A heterogenous autonomous collaborative system for powerline inspection using human-robotic teaming. In *2021 16th International Conference of System of Systems Engineering (SoSE)* (pp. 19-24,). IEEE. https://doi.org/10.1109/SOSE52739.2021.9497501

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *Interactions*, *26*(4), 42-46. https://doi.org/10.1145/3328485

Yan, B., Chen, Q., Ye, R., & Zhou, X. (2019). Insulator detection and recognition of explosion based on convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, *17*(02), 1940008. https://doi.org/10.1142/S0219691319400083

Yang, L., Huang, X., Zhang, Y., & Nie, T. (2019). Technology of actual wind deviation monitoring for suspension insulator strings based on improved edge detection. *IEEE Access*, *7*, 170157-170167. https://doi.org/ 0.1109/ACCESS.2019.2954935

Yang, Y., Wang, L., Wang, Y., & Mei, X. (2019). Insulator self-shattering detection: A deep convolutional neural network approach. *Multimedia Tools and Applications*, *78*(8), 10097-10112. https://doi.org/10.1007/s11042-018-6610-4

Zhao, J., Liu, X., Sun, J., & Lei, L. (2012, July). Detecting insulators in the image of overhead transmission lines. In *International Conference on Intelligent Computing* (pp. 442-450). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978- 3-642-31588-6_57

Zhao, Z., Liu, N., Xi, M., & Yan, Y. (2014, June). The orientation angle detection of multiple insulators in aerial image. In *Chinese Conference on Image and Graphics Technologies* (pp. 294-304). Springer. https://doi.org/10.1007/978-3-662-45498-5_33

Zuo, D., Hu, H., Qian, R., & Liu, Z. (2017, July). An insulator defect detection algorithm based on computer vision. In *2017 IEEE International Conference on Information and Automation (ICIA)* (pp. 361-365). IEEE. https://doi.org/10.1109/ICInfA.2017.8078934

Appendices

Appendix A

Interview Questions for Ideation Sessions

- Introduction., Explain the scope of the thesis and research question
- In the realm of AI and Intelligent systems, where do you see yourself on the spectrum?
- Concerning having no explanations, do you think that this current approach of models that having inability to explain causes mistrust between AI and humans?
- Why do current AI models need to provide Explanations, and will it help in inspection systems? Do you think human centered intelligent systems are missing? What are the key steps taken to bridge the gap in your company?
- What improvements can be brought about the current methodology?
- Who do you think will benefit the most from such a collaboration (between AI human- centered systems in inspection industry)?
- What does ideation with respect to AI look like? How does one identify opportunities?
- When it comes to bringing HC-XAI systems what are the factors that matters most in designing for critical uses. What is the desired educational path?
- What case studies with respect to HC-XAI design shed light on this intersection?
- What if your company doesn't use machine learning right now, and doesn't have the in- house expertise of Google, Amazon, or Facebook? What are some examples of improving existing AI related products?
- For developers who aren't yet working with human centered AI systems, how will your framework help them in their everyday work?
- What expertise does this require? There's already the ongoing debate about whether Explainable AI is necessary or not. Will it bridge the gap

between humans and AI if these are developed?

- What about the other way around? Would they benefit from learning about related design and UX considerations for data scientists already working with machine learning?

Appendix B

IRB Letter

UNIVERSITY OF THE INCARNATE WORD

October 12, 2021

To: Mr. Srikanth Vemula

From: University of the Incarnate Word Institutional Review Board, FWA00009201 Srikanth:

Your request to conduct the study titled Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approachto Bridge the Gap Between Humans and AI was approved by expedited review on 10/12/2021. Your IRB approval number is 21-10-002. You have approval to conduct this study through 10/12/2022.
The stamped informed consent document is uploaded to the Correspondence section in the Research Ethics Review system. Please use only the stamped version of the informed consent document.

Please keep in mind the following responsibilities of the Principal Investigator:

1. Conducting the study only according to the protocol approved by the IRB.
2. Submitting any changes to the protocol and/or consent documents to the IRB for review and approval prior to the implementation of the changes. Use the **IRB Amendment Request** form.
3. Ensuring that only persons formally approved by the IRB enroll subjects.
4. Reporting immediately to the IRB any severe adverse reaction or serious problem, whether anticipated or unanticipated.
5. Reporting immediately to the IRB the death of a subject, regardless of the cause.
6. Reporting promptly to the IRB any significant findings that become known in the course of the research that might affect the willingness of the subjects to participate in the study or, once enrolled, to continue to take part.
7. Timely submission of an annual status report (for exempt studies) or a request for continuing review (for expedited and full Board studies). Use either the **IRB Study Status Update** or **IRB Continuing Review Request** form.
8. Completion and maintenance of an active (non-expired) CITI human subjects training certificate.
9. Timely notification of a project's completion. Use the **IRB Closure** form.

Approval may be suspended or terminated if there is evidence of a) noncompliance with federal regulations or university policy or b) any aberration from the current, approved protocol.

If you need any assistance, please contact the UIW IRB representative for your college/school or the Office of Research Development.

Sincerely, Mary Jo Bilicek
Research Compliance Coordinator
University of the Incarnate Word
(210) 805-3565
bilicek@uiwtx.edu

Appendix C

Email to Potential Participants

Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI

Researcher: Srikanth Vemula
Department: Dreeben School of Education (PhD Candidate) Phone: (210) 283-5047
Email: vemula@uiwtx.edu

Dear Sir or Ma’am,

I am a PhD candidate who is currently in dissertation stage of my research on the topic of Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative approach to bridge the gap between Humans and AI.

I am sending this email to you as I feel your participation in the study would be extremely valuable. In your role as a Practitioner at DEUS/Polytopal/CPS, you have insights and knowledge that will enhance the scope of my research in building and testing the explainable AI model and use of the human-centered intelligent systems in energy industry.

For the purposes of my study, I will be facilitating a Qualitative Study. The study will explore your unique perceptions at a DEUS/Polytopal/CPS concerning the development of explainable models and how human centered explainable AI systems will foster social innovation. This study will also focus on your perceptions on possible barriers involved in creating those human-centered AI systems and how these systems can foster social innovation and establish a collaborative approach between humans and AI systems.

If you agree to be a part of this study, you will be asked to participate in method of data collection through individual interviews. During these Ideation sessions, you will be asked a series of open- ended questions. All interviews will be facilitated via the zoom pla tform, with a timeline of 60 minutes. Time dedicated to the interviews could be shorter or longer depending on the individual.

Once the initial set of data is collected, you will be asked to participate in a follow up interview. I truly believe that your participation will significantly contribute to this study, and I am hopeful that you will accept this invitation to participate. If you are willing to accept, please respond via email at vemula@uiwtx.edu. Once I receive your confirmation of acceptance, I will provide you with the appropriate Informed Consent Letter.

Thank you for your consideration, and I look forward to working with you in the upcoming

months. Sincerely,
Srikanth Vemula
PhD Candidate
Concentration: Social Innovation and Adult education Emphasis: Social Innovation
Dissertation Chair: Dr. Alison Buck

Appendix D

Letter of Cooperation from CPSE

Letter of Cooperation

Jose G Leandro
Coordinator EDS Maintenance Program 2 | DSO Reliability CPS Energy | 10830 Nacogdoches Rd.
San Antonio, Tx 78217 | MD:36.01.01 Date: 7 October 2021

Dear Mr. Vemula,

Based on my review of your research proposal, I give you my permission to facilitate the study entitled "Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI." As a part of this study, I provide my permission for you to do this study of an Artificial Intelligence (AI) algorithm that can identify damaged and malfunctioning equipment with explanations from the visual images. CPSE will provide datasets for training of AI models and test the HC-XAI system during their Continued flight operations training at the Training Yard and along energized powerlines within the CPS Energy.

We understand that our organization's responsibilities include permitting the recruitment of personal participation. We reserve the right to withdraw from the study at any time if our circumstances change.

I understand that I am authorized to approve research in this setting and that your plan complies with the organizations policies.

The parties acknowledge and agree that CPS Energy has no further obligation to provide financial support under this Agreement. I understand that the data collected will remain confidential and will not be provided to anyone outside of the researcher and the faculty/staff at the University of the Incarnate Word.

Sincerely,

Jose G Leandro
Coordinator EDS Maintenance Program 2 | DSO Reliability CPS
Energy | 10830 Nacogdoches Rd.
San Antonio, Tx 78217 | MD:36.01.01

Appendix E

Informed Consent Document for CPS

Subject Consent to Take Part in a Study of:

Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI.

University of the Incarnate Word

Authorized Study Personnel:
Researcher: Srikanth Vemula,
PhD Candidate Dreeben School
of Education (PhD Candidate)
Phone: (210) 283-5047
Email: vemula@uiwtx.edu

Faculty Advisor: Dr.
Alison Buck Phone:
210.422.4568
Email: mbuck@uiwtx.edu

Key Information: Your consent is being sought for a research study facilitated at CPS. The proposed study seeks to collect data from purposively selected key participants to identify importance of human centeredness in Intelligent systems and the role of explainable AI in constructing a sense of trust and collaborative environment between humans and AI systems. If you agree to participate in this study, the project will involve the following:

- Procedures will include the participants and the researcher to complete two individual interviews. Each interview will have a pre-determined set of questions and will last approximately 60 minutes in length.
- The meeting will take approximately one hour. During that time the researcher will make sure sufficient data is collected.
- There are no risks associated with this study.
- You will not be compensated for your participation.
- Your participation is voluntary, and you can decide not to participate at any time.

Invitation:

You are invited to volunteer as one of the subjects in the research project named above. The information in this form is meant to help you decide whether to participate. If you have any questions, please feel free to ask.

Why are you being asked to be in this research study? You are being asked to be in this study because the researcher feels that you will be able to provide in-depth information on your personal perceptions and experiences in regards to the AI inspection systems and the need for human-centered explainable systems to bridge the gap between humans and AI.

Appendix F

Informed Consent Document for DEUS

Subject Consent to Take Part in a Study of:

Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI.

University of the Incarnate Word

Authorized Study Personnel:
Researcher: Srikanth Vemula,
PhD Candidate Dreeben School
of Education (PhD Candidate)
Phone: (210) 283-5047
Email: vemula@uiwtx.edu

Faculty Advisor:
Dr. Alison Buck
Phone:
210.422.4568
Email: mbuck@uiwtx.edu

Key Information: Your consent is being sought for a research study facilitated at DEUS. The proposed study seeks to collect data from purposively selected key participants to identify importance of human centeredness in Intelligent systems and the role of explainable AI in constructing a sense of trust and collaborative environment between humans and AI systems. If you agree to participate in this study, the project will involve the following:

- Procedures will include the participants and the researcher to complete two individual interviews. Each interview will have a pre-determined set of questions and will last approximately 60 minutes in length.
- The meeting will take approximately one hour. During that time the researcher will make sure sufficient data is collected.
- There are no risks associated with this study.
- You will not be compensated for your participation.
- Your participation is voluntary, and you can decide not to participate at any time.

Invitation:

You are invited to volunteer as one of the subjects in the research project named above. The information in this form is meant to help you decide whether to participate. If you have any questions, please feel free to ask.

Why are you being asked to be in this research study? You are being asked to be in this study because the researcher feels that you will be able to provide in-depth information on your personal perceptions and experiences in regards to the AI inspection systems and the need for human-centered explainable systems to bridge the gap between humans and AI.

Appendix G

Informed Consent Document for Polytopal

Subject Consent to Take Part in a Study of:

Human-Centered Explainable Artificial Intelligence (XAI) for Anomaly Detection in Quality Inspection: A Collaborative Approach to Bridge the Gap Between Humans and AI.

University of the Incarnate Word

Authorized Study Personnel:
Researcher: Srikanth Vemula,
PhD Candidate Dreeben School
of Education (PhD Candidate)
Phone: (210) 283-5047
Email: vemula@uiwtx.edu

Faculty Advisor:
Dr. Alison Buck
Phone:
210.422.4568
Email: mbuck@uiwtx.edu

Key Information: Your consent is being sought for a research study facilitated at Polytopal. The proposed study seeks to collect data from purposively selected key participants to identify importance of human centeredness in Intelligent systems and the role of explainable AI in constructing a sense of trust and collaborative environment between humans and AI systems. If you agree to participate in this study, the project will involve the following:

- Procedures will include the participants and the researcher to complete two individual interviews. Each interview will have a pre-determined set of questions and will last approximately 60 minutes in length.
- The meeting will take approximately one hour. During that time the researcher will make sure sufficient data is collected.
- There are no risks associated with this study.
- You will not be compensated for your participation.
- Your participation is voluntary, and you can decide not to participate at any time.

Invitation:

You are invited to volunteer as one of the subjects in the research project named above. The information in this form is meant to help you decide whether to participate. If you have any questions, please feel free to ask.

Why are you being asked to be in this research study? You are being asked to be in this study because the researcher feels that you will be able to provide in-depth information on your personal perceptions and experiences in regards to the AI inspection systems and the need for human-centered explainable systems to bridge the gap between humans and AI.