Titre: Title: Effects of Automated Diagnostic Tools on Human Performance during Failure Management

Auteur: Author: Karine Ung

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Citation: Ung, K. (2025). Effects of Automated Diagnostic Tools on Human Performance during Failure Management [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/65821/

Directeurs de recherche: Advisors: Philippe Doyon-Poulin

Programme: Program: DR-Génie Industriel

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Effects of Automated Diagnostic Tools on Human Performance during Failure Management

KARINE UNG
Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie industriel
Avril 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Effects of Automated Diagnostic Tools on Human Performance during Failure Management

présentée par **Karine UNG**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Jean-Marc FRAYRET, président
Philippe DOYON-POULIN, membre et directeur de recherche
Moncef CHIOUA, membre et codirecteur de recherche
Bruno BLAIS, membre
Greg JAMIESON, membre externe

DEDICATION

To my parents, Ai Hour and Khai. To my partner, Thierry.

In loving memory of Waffle and Sasaki.

ACKNOWLEDGEMENTS

My first thoughts go to my research team. I am grateful to my research director, Philippe Doyon-Poulin, for believing in me and supporting me throughout this journey. His trust, patience, encouragement, availability, and insightful advice, along with his creative contributions, have been invaluable to this work. I extend my gratitude to my research co-director, Moncef Chioua, for his expertise, guidance, and patience in helping me navigate the fundamentals of chemical engineering. I thank Jean-Marc Robert, whose passion for ergonomics has shaped my perspective since my master's studies. His enthusiasm for the field has been inspiring.

I am also grateful to my jury committee for taking on the task of reviewing and evaluating this thesis. I thank Jean-Marc Frayret for assuming the role of president and Greg Jamieson and Bruno Blais for serving as jury members. It is an honor to have internationally recognized experts on my committee, and I strive to meet the high standards you uphold.

This research would not have been possible without the individuals who participated in the experiments. I appreciate their time and contributions, which provided the necessary data for this study. I also acknowledge the financial and institutional support of NSERC, IVADO, and Polytechnique Montréal, which made this research possible.

On a personal note, I am thankful to my family and friends for their unwavering support. To my parents, Ai Hour and Khai, who instilled in me the values of perseverance and lifelong learning. A special thanks to Thierry for his constant support and encouragement throughout the challenges of this work.

RÉSUMÉ

Les environnements de contrôle des procédés industriels utilisent des systèmes de gestion des alarmes pour aider les opérateurs à identifier et à traiter les conditions anormales. Dans les systèmes interconnectés, une seule faute peut déclencher plusieurs alarmes simultanément, entraînant une inondation d'alarmes, où le nombre d'alarmes dépasse la capacité de l'opérateur à traiter efficacement l'information. Les inondations d'alarmes peuvent ralentir les temps de réponse, augmenter la charge cognitive, réduire la conscience de la situation et compliquer l'identification des fautes, augmentant ainsi le risque d'erreurs opérationnelles et d'accidents. L'intelligence artificielle et l'automatisation contribuent à relever ces défis grâce à des outils de diagnostic automatisés basés sur l'IA, qui aident les opérateurs à interpréter les alarmes et à identifier la cause principale des fautes. Bien que ces outils puissent améliorer la performance des opérateurs et la prise de décision, ils présentent également des risques lorsque l'IA fournit des recommandations incorrectes. Cette thèse examine l'influence des outils de diagnostic automatisés basés sur l'IA sur la performance des opérateurs et la prise de décision en situation d'inondation d'alarmes, en mettant l'accent sur la conscience de la situation, la charge de travail, la confiance et l'impact des défaillances de l'IA.

Cette thèse s'appuie sur trois études, chacune publiée sous forme d'article distinct. La première étude a développé PER4Mance, un simulateur en accès libre et haute-fidélité basé sur le Tennessee Eastman Process. Ce simulateur a été conçu pour reproduire les conditions d'inondation d'alarmes et fournir un environnement contrôlé permettant d'étudier les interactions humain-machine. La deuxième étude a utilisé PER4Mance pour évaluer l'effet d'un outil de diagnostic automatisé sur la performance des opérateurs. Vingt participants ont pris part à cette étude, permettant le recueil de données sur la conscience de la situation, la charge de travail cognitive et l'oculométrie. Les résultats ont montré que l'outil a amélioré la conscience de la situation et réduit la charge cognitive, en particulier dans les scénarios de fautes complexes, où les alarmes seules ne fournissaient pas d'informations de diagnostic claires. Les données d'oculométrie ont indiqué que les opérateurs utilisant l'outil passaient moins de temps à parcourir la liste des alarmes et portaient davantage d'attention aux indicateurs de performance clés et aux recommandations diagnostiques. La troisième étude a analysé les effets des défaillances de l'IA sur la confiance, la dépendance, la charge de travail et la conscience de la situation. Vingt participants ont pris part à l'étude. Les résultats ont montré qu’en cas de défaillances de l’IA, leur performance était réduite, caractérisée par une diminution de la conscience de la situation, un taux de bonnes réponses plus faible et une augmentation des temps de réponse. Bien qu'ils aient déclaré avoir moins confiance dans le système d'IA, ils ont continué à suivre ses diagnostics incorrects, montrant une tendance à s'appuyer sur les recommandations automatisées.

Cette thèse contribue à l'avancement des connaissances sur (1) l'interaction humain-IA, en explorant les défis liés à l'intégration des outils d’aide à la décision basés sur l’IA tout en maintenant une supervision humaine, et en abordant les questions de biais d'automatisation et de masquage des fautes de l’IA ; (2) le rôle des outils d’aide à la décision basés sur l’IA dans la gestion des situations anormales, en montrant que leur efficacité dépend de la complexité des fautes, de l’expérience des opérateurs et du contexte opérationnel ; et (3) l’avenir des systèmes d’alarme, en questionnant le déclanchement d’alarmes comme principal moyen de transmission des informations, en raison du risque accru d'inondations d’alarmes dans les systèmes industriels, de transport, énergétiques et médicaux émergents et de plus en plus interconnectés. Les recherches futures devraient explorer l'automatisation adaptative, la formation dans des simulateurs haute-fidélité avec des opérateurs expérimentés et l’évaluation des outils d’aide à la décision basés sur l’IA afin d’améliorer la conception et l’intégration de ces technologies dans les industries de procédés et autres environnements à haut risque.

ABSTRACT

Industrial process control environments use alarm management systems to assist operators in identifying and addressing abnormal conditions. In interconnected systems, a single fault can activate multiple alarms at once, creating an alarm flood where the number of alarms exceeds an operator's ability to process information efficiently. Alarm floods can slow response times, increase cognitive workload, reduce situational awareness, and complicate fault identification, raising the likelihood of operational errors and accidents. Artificial intelligence and automation help address these challenges through AI-based automated diagnostic tools that assist operators in interpreting alarm patterns and isolating root causes. While these tools can enhance operator performance and decision-making, they also pose risks when AI provides incorrect recommendations. This thesis examines how AI-based automated diagnostic tools influence operator performance and decision-making during alarm floods, focusing on situational awareness, workload, trust, and the impact of AI failures.

The thesis was conducted in three studies, each published as a separate article. The first study developed PER4Mance, an open-source, high-fidelity process control simulator based on the Tennessee Eastman Process. The simulator was designed to replicate alarm flood conditions and provide a controlled environment for studying human-machine interactions. The second study used PER4Mance to assess the effects of an automated diagnostic tool on operator performance. Twenty participants took part in the study. Results showed that the tool improved situational awareness and reduced cognitive workload, particularly in complex fault scenarios where alarms alone did not provide clear diagnostic information. Eye-tracking data indicated that operators using the tool spent less time scanning alarm lists and directed more attention to key performance indicators and diagnostic recommendations. The third study examined the effects of AI failures on trust, reliance, workload, and situational awareness. Twenty participants contributed to the study. Results indicated that during AI failures, participants exhibited lower performance, as shown by reduced situational awareness, decreased accuracy, and slower response times. Although they reported lower trust and reliance on the AI system, they continued to follow its incorrect diagnoses, demonstrating a tendency to depend on automated recommendations.

This thesis advances knowledge on (1) human-AI interaction, examining the challenges of balancing AI-based decision support tools with human oversight, while also addresses automation

bias and automation failure masking; (2) the role of AI-based decision support tools in abnormal situation management, demonstrating that their effectiveness depends on fault complexity, operator experience, and operational context; and (3) the future role of alarms, questioning the use of annunciated alarms due to the increasing risk of alarm floods in emerging and more advanced industrial, transportation, energy or medical systems. Future research should investigate adaptive automation, training in high-fidelity simulators with experienced operators, and further evaluation of AI-based decision support tools to enhance the design and integration of these technologies in process industries and other high-risk environments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LISTE OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
AFC	Alarm Flood Classification
ALAP	Alarm Prioritization
ANOVA	Analysis of Variance
ANSI	American National Standards Institute
AOI	Area of Interest
ASM	Abnormal Situation Management
CCPS	Center for Chemical Process Safety
CD	Correct Diagnosis
DCS	Distributed Control Systems
DL	Deep Learning
EEMUA	Engineering Equipment and Materials Users Association
FDD	Fault Detection and Diagnosis
HMI	Human-Machine Interaction
ID	Incorrect Diagnosis
IEEE	Institute of Electrical and Electronics Engineers
ISA	International Society of Automation
KPI	Key Performance Indicator
ML	Machine Learning
NASA-TLX	Nasa Task Load Index
OOTL	Out-Of-The-Loop
PLC	Programmable Logic Controllers
P&ID	Piping and Instrumentation Diagram
RFC	Relative Fixation Count
RFD	Relative Fixation Duration
SA	Situational Awareness
SAGAT	Situational Awareness Global Assessment Technique
SCADA	Supervisory Control And Data Acquisition
TEP	Tennessee Eastman Process
WD	With Diagnosis

WOD	Without Diagnosis
WoZ	Wizard of Oz

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.1 Context

In process industry, alarm management systems notify operators of deviations from nominal conditions. Abnormal situations—such as equipment overheating, pressure buildup, or chemical imbalances—can vary in severity, impacting safety, efficiency, and the environment [1]. In complex and interconnected process systems, deviations can trigger multiple alarms simultaneously, leading to *alarm floods* [2]. These events, characterized by a high number of alarms in a short period, have been identified as contributing factors in several industrial incidents, as they can overwhelm operators' capacity to process information and take corrective action [3]. Notable examples include the Piper Alpha Oil platform, BP Texas City refinery and Buncefield oil depot.

The Piper Alpha Oil Platform Disaster (1988) accident began with a condensate pump failure, leading to a gas leak that triggered multiple alarms [4]. The overwhelming number of alarms made it challenging for operators to identify the critical threat. Subsequently, an explosion destroyed the control room, resulting in a loss of centralized command and hindering emergency response efforts. This sequence of events contributed to the loss of 167 lives.

The BP Texas City Refinery Explosion (2005) occurred during the start-up of the isomerization unit [5]. Operators inadvertently overfilled the raffinate splitter tower, leading to an overpressure and the release of a hydrocarbon vapor cloud. This vapor cloud subsequently ignited, resulting in a catastrophic explosion. Investigations revealed that operators were inundated with a high volume of alarms, many of which lacked prioritization based on severity. This alarm overload impeded the operators' ability to promptly identify and address critical warnings, delaying corrective actions. As process conditions deteriorated, the increasing number of alarms further overwhelmed operators, hindering effective management of the escalating situation. The explosion resulted in 15 fatalities and over 170 injuries.

The Buncefield Oil Depot Explosion (2005) accident began when a tank was overfilled with petrol due to a malfunctioning level gauge and an inoperative independent high-level switch, leading to the release of approximately 250,000 liters of fuel [6]. This resulted in a vapor cloud that ignited, causing a massive explosion and a fire that lasted five days. The malfunction of the tank control and alarm management systems was a key factor in the release. Operators were subjected to an overwhelming influx of alarms within a brief timeframe, which impeded their ability to discern and prioritize critical warnings. This alarm overload led to delays in implementing corrective measures, thereby compromising the effectiveness of the incident response. The explosion caused significant damage to the facility and surrounding areas.

Alarm floods occur in multiple sectors beyond industrial environments. In public transportation, automatic train control systems can generate thousands of alarms each week, making it more difficult for dispatchers to distinguish and respond to critical events [7], [8], [9]. In healthcare, the continuous activation of medical device alarms contributes to alarm fatigue, which can lead to slower response times and, in some cases, preventable harm [10], [11], [12], [13]. In aviation, frequent false or unreliable alarms reduce pilots' trust in automated systems, increasing the likelihood of overlooking essential warnings. Studies show that inaccurate alarms continue to be a challenge in flight operations [14], [15], [16]. Effective alarm management is necessary in various industries where an excessive number of alarms can interfere with timely and appropriate responses.

Advancements in connectivity and data integration have improved alarm flood management, enhancing fault detection and diagnostics [17]. Artificial Intelligence (AI)-driven systems can process extensive alarm data, identify patterns, and predict potential failures, thereby assisting operators in filtering non-critical alarms, prioritizing essential ones, and recommend diagnostics [18]. However, increased reliance on AI and automation introduces new challenges. AI systems improve operational efficiency but are also prone to failures [19]. When these failures occur, operators may experience errors in decision-making and declines in performance [20]. This thesis examines the effects of integrating an automated decision support tool on human performance, with a focus on automated system failures. The research aims to identify methods to improve human-AI collaboration and minimize operational risks in high-stakes environments.

1.2 Objectives

The objective of this thesis is to examine the effects of an automated decision support tool on human performance during an alarm flood episode, with a focus on automated system failures. To achieve this, the research aims to develop a simulator environment that represents a high-fidelity industrial process control system. This environment will allow for the simulation of faults and alarm flood scenarios, providing a controlled setting to study human interaction with automated decision support tools. The study also seeks to evaluate human performance when using an automated diagnostic support tool, determining whether it improves an operator's ability to manage alarm floods or has no measurable effect. Additionally, the research will investigate human-AI interaction during AI failures, analyzing how incorrect AI recommendations influence operator decision-making and performance.

1.3 Thesis organization

This thesis is made of three articles that are either published or under review in peer-review journals that constitutes the core of its scientific contributions. The thesis is structured as follows:

- **Chapter 1** (this chapter) introduces the research context, objectives, and overall organization of the thesis.
- **Chapter 2** provides a comprehensive literature review of the topics of alarm floods in industrial process control, alarm flood mitigation strategies, human-AI collaboration, and AI failures.
- **Chapter 3** outlines the general research approach, research objectives and summarizes the objectives of the three articles.
- **Chapters 4, 5, and 6** present the research findings in the form of three scientific articles.
- **Chapter 7** synthesizes the findings across the articles, discusses the contributions of this thesis, and acknowledges its limitations.
- **Chapter 8** concludes the thesis by highlighting recommendations for future research directions.

CHAPTER 2 LITERATURE REVIEW

This thesis examines the effects of automated diagnostic tools on human performance during abnormal situation management (ASM) in industrial settings. The research lies at the intersection of process alarm management and process control, AI and automated fault diagnosis, and human-machine interaction (HMI) and performance. To fully grasp the research objectives of this thesis, this literature review will explore key concepts within each of these domains, providing a comprehensive foundation for investigating the challenges and opportunities of human-AI collaboration in failure management scenarios.

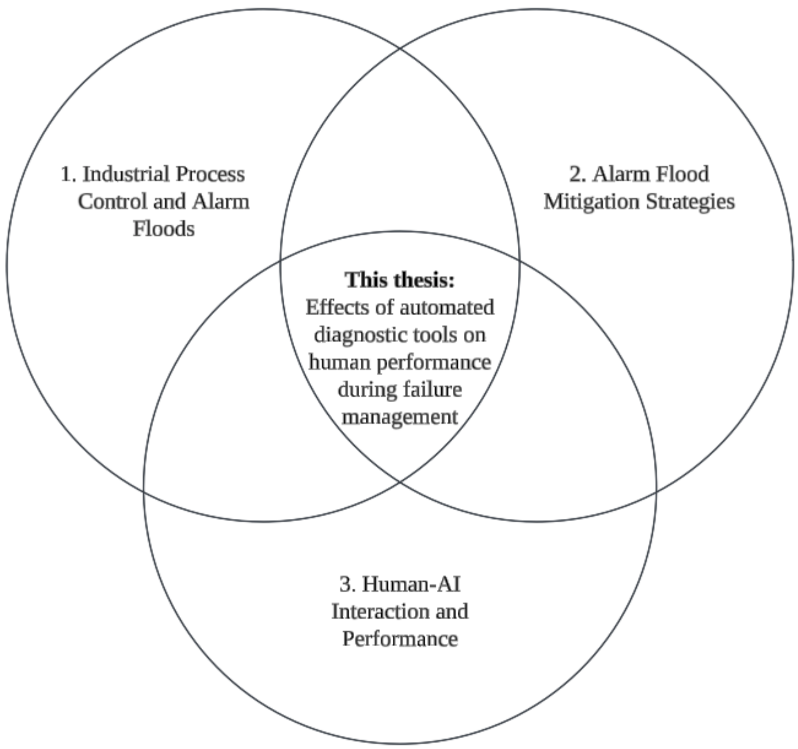

Figure 2.1 Break-down of the thesis' literature review in three main sections.

This literature review is structured into four main sections (**Figure 2.1**): The first section explores the foundational principles on alarm floods in industrial process control, such as faults, failures, and abnormal situations. The second section discusses alarm management, alarm analysis, and fault detection and diagnosis as alarm flood mitigation strategies. The third on Human-AI collaboration addresses key concepts between operator and automation interaction, including AI failures, situational awareness, trust and reliance, out-of-the-loop challenges, and cognitive workload. The fourth section reviews related past research on automated diagnostic tools during failure management.

2.1 Industrial Process Control and Alarm Floods

This section examines industrial plants, process control, and alarm management systems. It begins by describing the structure and function of an industrial plant, followed by an explanation of process control and the role of the operator in maintaining system stability. Then, it examines the factors that contribute to abnormal situations, including faults and failures. Finally, it discusses alarm management systems and alarm floods, outlining their functions, causes, and impact on operator performance. This foundational overview establishes the context for the challenges associated with alarm floods in industrial environments.

2.1.1 Industrial Process Control

Key definitions are needed to explain the industrial process control and alarm management system (**Figure 2.2**). An **industrial plant** is a facility where chemical, physical, or mechanical processes convert raw materials into finished products [21]. Examples include oil refineries, chemical plants, power stations, food processing facilities, pharmaceutical production sites, and automotive assembly plants.

Process control is the automated and manual regulation of industrial operations to optimize performance by monitoring and adjusting variables through real-time feedback and control mechanisms [22]. Industrial plant relies on process control systems to ensure efficiency, safety, and cost-effectiveness by monitoring and regulating variables such as temperature, pressure, flow, and chemical composition [23]. Effective process control enhances productivity, reduces waste, and ensures compliance with safety and environmental standards [21]. Automation and human oversight work together to maintain stability, minimize disruptions, and optimize performance [24].

Alarm systems identify deviations from normal operating conditions and notify operators to enable timely intervention [25]. Industrial facilities use these systems to monitor process control variables and signal when values exceed defined thresholds. While process control systems regulate operations, alarm systems provide notifications that support operator decision-making [3].

An industrial plant **operator** monitors and controls equipment and processes in an industrial facility to ensure safe and efficient operations [26]. Responsibilities include interpreting data from control systems, adjusting parameters, responding to alarms, and performing routine inspections. Operators work in sectors such as manufacturing, chemical processing, power generation, and oil refining, following operational procedures and regulatory standards [27].

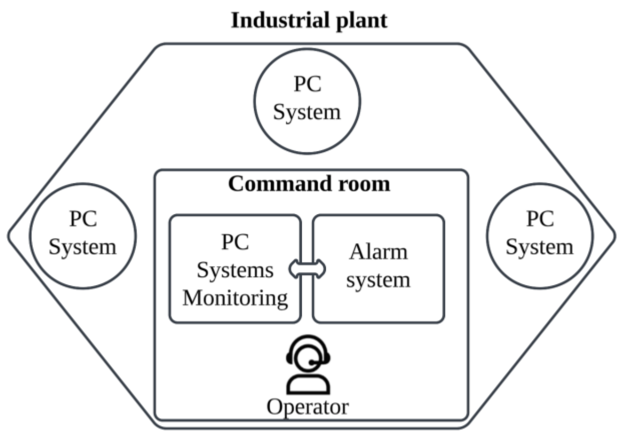

Figure 2.2 Alarm management system where operators monitor industrial control process (PC) and alarm systems in the command room.

2.1.2 Alarm Floods

Alarms play a crucial role in plant safety, acting as a safeguard to prevent **faults** from escalating into serious hazards or accidents [25]. Their primary function is to help operators maintain processes within normal operating conditions. By alerting operators to faults and **failures**, alarms ensure timely intervention to prevent operational disruptions [28]. To ensure clarity and precision in this thesis, it is essential to define key terms as follows:

- **Faults**: Unacceptable deviations of a process variable from its normal state. These deviations reflect the difference between a threshold value and a fault value and can lead to process malfunctions or failures. Faults may already exist in the process or arise at an unpredictable time, with varying rates of progression [29].
- **Failures**: Permanent interruptions in a system's ability to perform its required function under specified operating conditions. Failures are typically caused by multiple faults and result in the inability of the system to execute production or fulfill a demanded function [30].

- **Abnormal situation**: disturbances or deviations from normal operating conditions caused by factors like equipment failures, human errors, or external influences [31].
- **Alarms**: Notifications provided to operators to indicate equipment malfunctions, process deviations, or abnormal conditions that require immediate attention [32]. Alarm messages are directly associated with faults and are triggered when the process variable exceeds or falls below predetermined thresholds.

Alarm systems play a crucial role in alerting operators to abnormal situations in industrial processes. An **abnormal situation** arises when a process deviates from normal operating conditions, potentially leading to unsafe, inefficient, or environmentally hazardous outcomes [33]. These deviations can stem from equipment malfunctions, unexpected external influences, among other possible factors.

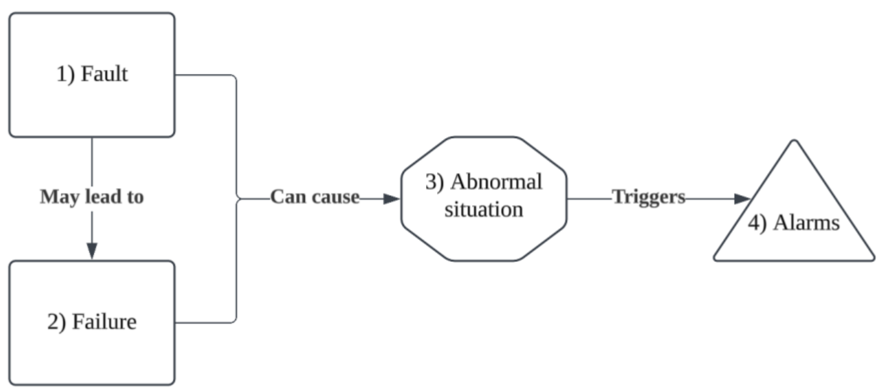

Figure 2.3 Faults and failures can cause abnormal situations which triggers alarms.

Here is a step-by-step diagram illustrating the relationship between faults, failures, abnormal situations, and alarms (**Figure 2.3**). We will use as an example a sudden rise pressure buildup from a clogged valve. This can be seen in process industries due to crystallization or chemical residue buildup. In this case, a blocked valve restricts fluid flow, disrupting normal operations (Step 1 fault). This obstruction causes pressure buildup in the system as fluid accumulates with no clear path (Step 2 failure). As pressure continues to rise, the system deviates from safe operating conditions, creating an abnormal situation that could lead to equipment damage or failure (Step 3 abnormal situation). To prevent further escalation, an alarm is triggered, alerting operators to the high-pressure condition so they can take corrective action before a critical incident occurs (Step 4 alarm).

Alarms help maintain plant safety by preventing faults from escalating into hazards. They assist operators in keeping processes within normal limits by signaling when action is needed. An alarm is triggered when a monitored variable crosses a predefined threshold, regardless of the situation's context. Alarms do not analyze patterns or diagnose problems—they simply respond to specific limits set in advance [34].

Each alarm message includes an operational procedure for corrective action [35]. In industrial processes, messages are prioritized based on the level of urgency of the operator's response [36], [37]:

- Warning-level alarms require immediate awareness and an immediate operator's response. These alarms are usually color-coded in red.
- Caution-level alarms require immediate awareness and a subsequent operator's response. They are usually color-coded in amber or yellow.
- Advisory-level alarms provide information but do not require a response. These alarms can be color-coded in cyan or white.

Operators follow predefined procedures to address warning- and caution-level alarms. Advisory-level alarms do not have procedure since no action is required [36], [37].

A major issue arises when a single fault triggers multiple alarms at once, overwhelming operators and making it difficult to isolate the root cause. This situation, known as **alarm flood**, happens when alarms are configured without considering how different system components interact. In large, interconnected industrial plants, disturbances can lead to multiple system failures, triggering a cascade of alarms and potentially causing an alarm flood. An alarm flood is defined as 10 or more alarms within a 10-minute period per operator [38]. It also refers to a situation in which the alarm rate exceeds the operator's capacity to respond effectively [39]. This operator overload limits the ability to investigate alarms and isolate the root cause of system disturbances [31]. As failures progress, new alarms accumulate without differentiating between causes, further complicating response efforts [40]. Previous accidents, introduced in chapter 1, have revealed that operators were overwhelmed by alarm floods that generated hundreds or thousands of process alarms, often requiring operators to silence them without addressing the root issue [41]. The large volume of alarms creates human factors challenges, primarily due to the limited capacity of human attention [7]. The gap between the amount of information provided and what operators can effectively process increases mental workload, contributes to human error, and reduces operational efficiency [42].

Alarm floods increase the risk of loss of control, accidents, potential loss of life, equipment damage, financial losses, and environmental harm [15]. The Engineering Equipment and Materials Users Association (EEMUA) has identified them as a major factor in catastrophic incidents [3]. One notable case is the 1994 Milford Haven explosion at the Texaco refinery in Pembroke, South Wales. A release of 20 tons of hydrocarbons from the flare header's knock-out pot led to a massive explosion, preceded by hundreds of alarms in the final minutes. The Health and Safety Executive’s investigation found that alarm overload can undermine safety by overwhelming operators rather than aiding decision-making [43]. In the 1998 Esso Australia gas plant explosion, a fractured lean oil vessel released gas, triggering hundreds of alarms. The excessive alerts desensitized operators, contributing to ineffective response. The explosion resulted in two deaths, eight injuries, and a two-week gas supply disruption in Melbourne [44]. These past incidents indicate that alarm floods often overload operators’ cognitive processing capacity, presenting challenges related to attention and decision-making [14], [45].

If we go back to the previous example of a clogged valve, the blockage can trigger multiple alarms simultaneously (**Figure 2.4**). This can trigger high-pressure alarms before the blockage, followed by a low-pressure and low-flow alarms, and potentially low temperature alarms if the restriction affects heat transfer. Operators may struggle to identify whether the primary issue is the valve blockage itself or its cascading effects on the system.

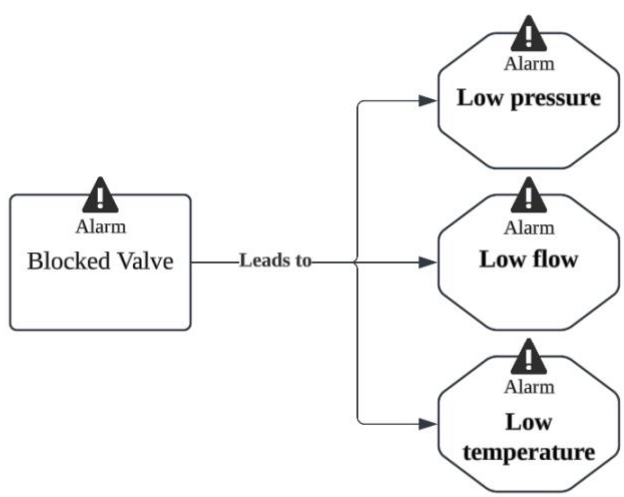

Figure 2.4 Blocked valve causing multiple alarms simultaneously.

Adjusting alarms to avoid unnecessary triggers while ensuring important ones are not missed is work-intensive. It requires fine-tuning thresholds and introducing deadbands [46], which prevent alarms from repeatedly triggering when values fluctuate near setpoints. Because alarms are based on thresholds, alarm floods can result from poor alarm design, system complexity, or both [47].

Alarm floods vary in their characteristics, including the rate at which alarms activate and the extent to which the alarms indicate the root cause of a fault or only its consequences [48]. This thesis investigates the impact of alarm floods based on their level of diagnostic clarity: "difficult" or "complex" alarm floods, where alarms do not explicitly isolate the root cause, and "easy" or "simple" alarm floods, where the root cause is clearly indicated.

Careful alarm management is necessary to keep alarms useful rather than overwhelming [49]. We investigate this topic in the next section.

2.2 Alarm Floods mitigation strategies

Building on the challenges of alarm floods, this section explores strategies to mitigate their impact. These strategies fall into three main categories: alarm management, alarm analysis, and fault detection and diagnosis (FDD). Each approach equips operators with tools to improve safety and performance by reducing alarm overload. Rather than competing, these methods complement each other, offering different ways to manage alarm floods effectively.

2.2.1 Alarm Management

Alarm management is a structured approach to optimizing alarm systems to ensure operators receive relevant alarms [50]. A high volume of alarms can overwhelm operators, reducing their ability to respond effectively. To address this issue, industries implement strategies such as dynamic threshold adjustments, alarm filtering, and rationalization. These methods help prioritize important alarms, reduce unnecessary alarms, and improve overall system performance [51], [52]. There have been multiple case studies that showed the application of alarm management methods to reduce alarm floods.

Effective alarm management minimizes unnecessary alarms, enabling operators to focus on urgent process deviations and reducing the risk of accidents and downtime [25]. A study on chattering alarms, which were repetitive alarm signals triggered in quick succession due to minor fluctuations in process variables, introduced an optimization-based method for determining dead-band values to reduce their frequency and improve efficiency and safety [53]. The method was evaluated using the Tennessee Eastman Process simulator, a standard benchmark for industrial systems, and compared with standard practices, demonstrating improved dead-band specification and operational performance. Reducing chattering alarms also lowered overall alarm volumes, aiding in the management of alarm floods, and improving operator response in industrial settings.

A clustering algorithm was applied to an ethylene cracking furnace to optimize alarm management systems and improve safety [54]. The implementation of the clustering–ranking algorithm in a refinery in northern China showed that the number of alarms generated by the distributed control systems (DCS) and those perceived by operators exceeded the manageable threshold set by EEMUA. This increase in alarms contributed to operator disturbance and reduced decision-making effectiveness. The proposed method provided a structured approach to alarm management, improving process operations and plant safety in the chemical industry.

Multi-agent systems were used in petroleum settings to help operators manage an alarm flood during emergencies [55]. A proposed multi-agent-based alarm management system synthesized process conditions during emergencies, assisting operators in interpreting and managing alarms. It incorporated reasoning, proactivity, communication, and adaptive behavior to improve alarm handling. The system achieved a suppression rate of up to 93.76%, allowing operators to focus on unexpected events rather than being overwhelmed by routine alarms.

Pattern mining was applied to a GE power plant dataset to reduce redundant alarms while maintaining safety and efficiency [56]. The method analyzed alarm logs to identify frequent patterns and correlations, enabling the removal of redundant alarms. It involved two steps: automatic pattern detection and expert analysis to validate correlations. The approach, tested on real power plant data, significantly reduced episodes of alarm flood without affecting operational performance.

Advanced strategies, including dynamic alarm management, alarm shelving, and predictive analytics, were used to improve refinery operations [57]. The study examined methods for reducing episodes of alarm flood and nuisance alarms, which contributed to operator fatigue. It analyzed the role of predictive analytics and machine learning in proactive monitoring and early issue detection. The research also addressed best practices such as alarm rationalization, prioritization, audits, philosophy documents, and operator training to manage alarm flood episodes and improve safety, efficiency, and regulatory compliance in refinery operations.

Finally, recent studies have consistently highlighted the need for future research to focus on the integration of emerging technologies, the ergonomic and human-centric design of alarm management systems, and adherence to industry standards such as ISA 18.2 for effective alarm management [58]. This includes incorporating artificial intelligence and machine learning for prediction and diagnosis [55], [59], as well as improving training, continuous monitoring, and operator feedback to enhance system effectiveness [57], [60].

2.2.2 Alarm Analysis

Alarm analysis is a data-driven method for evaluating alarm system performance [61]. By assessing historical and real-time data, it eliminates redundant and low-priority alarms, ensuring operators receive only relevant alerts [62]. The evaluation of alarm data reduces unnecessary activations, enhancing system clarity and allowing operators to focus on critical situations [63]. Alarm analysis also prevents episodes of alarm flood by identifying inefficiencies and recurring fault patterns [64]. Alarm analysis includes various methods such as statistical trend analysis, root cause analysis, event correlation, and predictive analytics. However, for the scope of this paper, the focus will be on methods that use historical data analysis and root cause isolation.

Previous papers have demonstrated that alarm analysis methods were successful alarm flood mitigation strategies. A study designed a semi supervised, data-driven method to classify episodes of alarm flood using historical data [63]. Their approach involved clustering, labeling alarms, and online early classification process. The method was evaluated using the Tennessee Eastman process (TEP) benchmark and an industrial alarm flood dataset. Results demonstrated a clustering reliability of 99.52%, highlighting the method's accuracy in detecting episodes of alarm flood at an early stage.

Another research introduced a method for classifying episodes of alarm flood using sequence mining and time series analysis to categorize floods based on historical data [64]. The apps operated in two stages: identifying whether a flood belonged to a new class and classifying it when a previous class provided a basis for comparison. A historical alarm classifier was integrated into fault detection and identification. A case study on an offshore oil-gas separation plant demonstrated that the algorithm successfully matched new alarm flood episodes with past instances of the same abnormal condition, enabling root cause isolation. The method achieved an average classification accuracy of 92.2%.

A research paper introduced a self-attention-based classifier with word embeddings to analyze historical alarm data [65]. The method was applied to the Tennessee Eastman process. The model achieved perfect accuracy and precision while being trained on a dataset of 2,000 alarm tags, significantly fewer than those required by existing methods. The training process was efficient, completing in approximately 22 seconds, demonstrating its potential for real-time applications in alarm-based root cause analysis.

A semi-supervised learning approach combined with case-based reasoning was used to analyze alarm flood episodes with minimal expert annotations [66]. The method consisted of two stages: offline learning, where historical alarm data was analyzed to identify patterns and build a case library, and online detection, where incoming an alarm flood was compared against stored cases to determine likely causes. Semi-supervised learning reduced the reliance on extensive labeled data, while case-based reasoning allowed the system to retrieve and update cases based on new alarm patterns. The approach was validated on real industrial alarm datasets, demonstrating its effectiveness in isolating root causes, reducing downtime, and improving operational decision-making.

Alarm analysis, such as historical data pattern matching and root cause identification, has proven effective in testing environments for managing and preventing alarm. A recurring recommendation from recent studies is to conduct more case studies to evaluate these methods in real-world scenarios. Expanding case studies could provide deeper insights into practical implementation, improving strategies for managing alarm floods and ensuring alarm systems operate efficiently.

2.2.3 Fault Detection and Diagnosis (FDD)

FDD is a proactive approach that detects, isolates, identifies, and diagnoses faults before they escalate into failures [67]. By continuously monitoring system parameters, FDD helps mitigate alarm floods by identifying anomalies early, preventing multiple alarms from being triggered by undiagnosed faults [68]. FDD employs methods such as model-based detection, data-driven analytics, signal processing, and machine learning to enhance fault recognition and resolution, preventing alarm overload, and improving response efficiency. The following section outlines the FDD loop (**Figure 2.5**) and the definition of key terms for the clarity of this thesis.

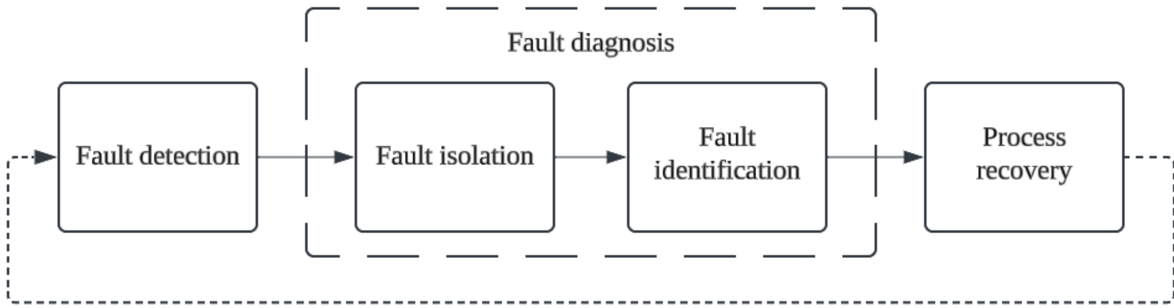

Figure 2.5 Steps of the Fault Detection and Diagnosis Loop

1. **Fault detection**: This step identifies deviations from normal operating conditions in a system by monitoring key parameters such as temperature, pressure, flow, and performance metrics. It involves analyzing real-time data to recognize anomalies that may indicate potential problems [68].
2. **Fault diagnosis**: This refers to the process of isolating and identifying faults in a system to determine their root causes and potential impacts, and whether it requires immediate intervention [30].
    - **Fault isolation**: Pinpointing the specific component, subsystem, or location within a system responsible for a detected fault [69].
    - **Fault identification**: Determining the fault's characteristics, severity, and underlying cause.
3. **Process recovery**: The method of restoring a system to normal operating conditions after a fault, failure, or disruption. It involves implementing corrective actions and adjusting system parameters. Recovery strategies may include automatic system reconfiguration, manual operator intervention, or backup system activation to maintain process continuity.

Once the fault has been diagnosed, operators follow standardized **operational procedures** to apply corrective actions [70]. Operational procedures are structured instructions that ensure the safe and efficient execution of tasks in industrial settings. According to the Center for Chemical Process Safety (CCPS), they outline the process, hazards, tools, protective equipment, and controls to help operators manage risks and verify process behavior [71]. They also guide troubleshooting, emergency shutdowns, and handling special situations like out-of-service equipment, ensuring consistency, compliance, and informed decision-making [32]. If fault detection, diagnosis, or the applied procedure fails to resolve the issue, the operator may need to restart the FDD loop.

Recent studies have demonstrated the applications of FDD in improving alarm flood management. A research paper examined the importance of early fault detection and diagnosis, particularly in managing alarm flood episodes [72]. The study introduced a novel algorithm to classify episodes of alarm flood by analyzing relationships between process variables and alarm data. The method was capable of handling large-scale plants with simultaneous alarms. Designed for online fault prediction, the algorithm was tested on the Tennessee Eastman Process system and a real industrial setting, demonstrating its effectiveness in early fault detection and diagnosis.

Another study developed a novel alarm dataset using a simulated nuclear power plant to evaluate fault detection and diagnosis (FDD) methods [73]. The researchers tested three Alarm Flood Classification (AFC) methods, which supported operators and automated systems in detecting, classifying, and diagnosing faults based on alarm patterns. The dataset contained controlled fault and alarm flood scenarios, allowing for a structured assessment of each method. One approach achieved 98% classification accuracy, demonstrating its effectiveness in fault identification. The findings showed that alarm flood classification improved fault detection and diagnosis in industrial systems.

Alarm Management, Alarm Analysis, and Fault Detection & Diagnosis (FDD) work together to enhance system reliability, operator efficiency, and alarm flood prevention (**Table 2.1**). Alarm Management ensures that alarms are structured and prioritized, Alarm Analysis continuously refines alarm settings based on performance data, and FDD proactively prevents faults from escalating into failures. For instance, during an alarm flood episode in an oil refinery, FDD can isolate the root cause, such as a pump failure, while alarm management filters out redundant notifications and highlights the most urgent alarms requiring immediate attention. At the same time, alarm analysis examines alarm trends, identifies patterns in recurring alerts, and helps refine alarm thresholds to prevent future episodes of alarm flood. This integration minimizes alarm fatigue, enhances decision-making, and strengthens industrial safety and efficiency [20].

Table 2.1 Summary of Alarm Management, Alarm Analysis and Fault Detection & Diagnosis with a focus on alarm floods and operator performance.

	Alarm Management	**Alarm Analysis**	**Fault Detection & Diagnosis (FDD)**
Definition	The systematic design, implementation, and maintenance of alarm systems.	The evaluation of alarm data to identify patterns, root causes, and system improvements.	A proactive method that detects, isolates, and diagnoses faults before they escalade.
Approach and Scope	System-level configuration and optimization of alarm settings. Prioritization, and suppression of unnecessary alarms.	Data-driven evaluation of alarm performance and root cause. Continuous evaluation of historical alarm data, data-driven insights into alarm frequency, trends, and impact.	Real-time monitoring and detection of real system issues.
Key Methods	Alarm rationalization, prioritization, suppression, shelving, dynamic thresholding.	Statistical trend analysis, root cause analysis, event correlation, predictive analytics.	Model-based detection, data-driven diagnostics, machine learning, signal processing.
Role in alarm flood mitigation	Prevent alarm floods through alarm rationalization, prioritization, and suppression.	Analyse alarm floods to identify root causes and optimize alarm thresholds.	Prevent alarm generation by detecting and diagnosing faults before they escalate into failures.

Although the reliability of these alarm flood mitigation strategies is not yet sufficient for real-world process control applications, they can isolate the root cause of alarm flood episodes within historical datasets and suggest it to the operator. While these approaches have shown to classify alarm flood episodes, they have not eliminated them, and their accuracy remain below 100%. Alarm floods remain a challenge, prompting industries to explore new strategies for improving alarm systems and mitigating their impact. Recent advancements have focused on artificial intelligence and machine learning to enhance FDD through more accurate fault prediction, real-time diagnostics, and adaptive alarm systems [18].

In this thesis, an automated fault diagnostic tool will be used to simulate alarm analysis by matching incoming alarms to a historical database in the Chapter 5 experiments. In the Chapter 6 experiments, the study will simulate the behavior of a FDD system that provides fault diagnostic recommendations to operators. This FDD system will also introduce AI failures, allowing an investigation into how operators respond to incorrect diagnostic outputs.

2.3 AI Failure in Process Control

Up to this point, we have provided an overview of industrial process control, the causes of alarm floods, and their impact on human performance and accidents. We then examined strategies for managing alarm floods, discussing their strengths and limitations. The focus now shifts to how emerging technologies, specifically AI and automation, enhance FDD as a tool for managing alarm floods and improve operator performance. To provide a foundation for this analysis, the section first presents an overview of AI and automation, followed by automation reliability and failures.

2.3.1 Artificial Intelligence and Automation

Artificial intelligence (AI) refers to the development of machines and systems capable of performing tasks that typically require human intelligence [74], [75], [76]. AI is designed to enable machines to perceive their surroundings, process information, and take actions to achieve specific objectives, often learning and improving from experience over time [77], [78]. This is made possible through advanced algorithms that can identify patterns, interpret data, and make decisions with minimal human involvement [79], [80]. AI encompasses a broad range of disciplines, including machine learning, which focuses on systems that adapt and evolve by analyzing data [81]; natural language processing, which allows machines to comprehend and produce human language [82]; computer vision, which enables the interpretation of visual inputs [83]; and deep learning, which has driven applications like facial recognition, voice synthesis and real-time language translation [84].

AI has become deeply embedded in the professional sphere, transforming how people work [85], [86], [87], [88]. AI is transforming industrial operations through advanced applications that enhance efficiency, reliability, and safety [89], [90]. One key application is predictive maintenance, where AI analyzes sensor data and historical trends to forecast equipment failures in chemical plants, allowing for proactive maintenance and reducing unplanned downtime [91]. Additionally, AI-driven process optimization improves energy efficiency in manufacturing by dynamically adjusting operational parameters based on real-time data [92]. In industrial control rooms, intelligent alarm management systems utilize AI to reduce episodes of alarm flood by filtering and prioritizing critical alerts, preventing operator overload [93]. Furthermore, AI-based anomaly detection enhances real-time process control and safety by identifying deviations from normal operating conditions, enabling quick corrective actions to prevent system failures or hazardous incidents [94], [95]. These AI applications are reshaping industrial environments by optimizing processes, reducing risks, and improving overall operational performance.

Automation involves the use of machines or systems to execute repetitive tasks based on predefined rules, ensuring precision and consistency [96]. It is particularly effective for structured, predictable processes that require minimal human intervention [97]. In industrial environments, automation enhances efficiency and safety by streamlining operations, reducing errors, and maintaining process reliability [98]. While often associated with AI, automation operates independently of learning algorithms, relying instead on programmed logic to perform specific tasks [99]. One key application is automated valve control in chemical processing plants, where programmable logic controllers (PLCs) [100], [101] and DCS [102], [103] regulate fluid flow, temperature, pressure and composition with precision, ensuring optimal process conditions. In manufacturing, conveyor belt and robotic arm automation streamline assembly lines, increasing production speed and reducing errors [104], [105]. Additionally, alarm suppression logic is implemented in industrial control rooms to filter out nuisance alarms, preventing operator overload and ensuring that only critical alerts are prioritized [106]. In pharmaceutical and chemical industries, batch process automation ensures precise ingredient mixing, temperature control, and reaction timing, improving product consistency and regulatory compliance [107], [108], [109]. These automation solutions help industrial facilities optimize operations, enhance safety, and improve overall productivity.

While AI and automation are often conceptually distinct [110], AI typically involving data-driven reasoning, and automation referring to rule-based functions, they frequently coexist and intertwine in industrial control systems. In practice, AI is not replacing automation but rather enhancing it, especially in the context of decision support and fault diagnosis. Many interactions between operators and AI-based systems occur through automated interfaces, making the distinction less visible from the operator's perspective. This is particularly true in environments like chemical plants, where AI-driven tools are embedded within existing automation architectures.

In this thesis, AI and automation are treated as components of a common human-machine ecosystem, where functional boundaries are less important than the cognitive consequences of system behavior. Failures, whether caused by automation logic or AI reasoning, can produce similar effects on trust, situational awareness, and performance. Therefore, although the thesis defines AI and automation separately for clarity, it draws extensively from the automation literature to inform its analysis of human-AI interaction. This approach is justified given that the operator's experience of support, disruption, or failure often emerges from the integrated behavior of both AI and automation systems.

2.3.2 Automation reliability

Understanding automation reliability is essential for studying how operators respond to decision-support tools, especially in high-risk environments where system performance influences trust and task management. In this context, reliability refers to the extent to which automated systems function correctly and consistently across operating conditions. This concept is particularly relevant when investigating automation failures and their effects on human performance.

In earlier literature, automation reliability was often defined as a performance ratio, described as how often a system produces correct outputs. However, more recent work recognizes that reliability is also shaped by user perception, system context, and the interaction between automation and human operators [111]. Reliability may refer not only to technical performance but also to the perceived consistency and dependability of the system during task execution.

Some researchers distinguish between performance reliability (accuracy of system outputs) and explanatory reliability (clarity of the system's reasoning or logic), both of which influence how operators trust and use automation [112]. This is especially relevant for AI-based systems, where recommendations may appear uncertain or change based on input conditions. Studies show that operators are more likely to lose trust in systems that fail unpredictably than in those that show consistently poor performance [113]. In dynamic environments, such as process control or emergency response, perceived reliability also shapes how and when operators intervene. When reliability is unclear or inconsistent, operators may disengage, delay responses, or misinterpret system outputs [114].

In this thesis, the reliability of a diagnostic tool is relevant because it influences operator behavior even when the system appears to function normally. Perceived reliability sets the conditions under which trust is formed, maintained, or degraded [115]. When reliability is violated—such as when a system provides an incorrect or missing diagnosis—the outcome is typically experienced as an automation failure [116], [117]. For this reason, automation reliability provides the foundation for understanding how failure conditions affect operator trust, workload, and situational awareness. In the next section, we will provide an overview of notable AI failures, examining their causes and consequences.

2.3.3 AI/Automation Failure

AI is increasingly used in chemical plants to improve efficiency, safety, and predictive maintenance [118], [119]. However, AI systems can produce inaccurate or unreliable results, leading to operational challenges [120]. AI-related failures in process industries include poor adaptation to new operating conditions [121], where AI struggles to respond to variations beyond its training data; cybersecurity risks [117], where AI-driven systems are vulnerable to hacking or data manipulation; and algorithmic errors in analyzing chemical reactions [122], which can lead to incorrect process optimizations.

AI-based decision-making tools can fail to perform its assigned function. Detection systems fail to identify targets, classifiers assign incorrect categories, decision aids produce inaccurate recommendations, and process automation deviates from expected operations [123]. An example is data bias, which affects predictive accuracy [124], [125]. If an AI model is trained primarily on data from normal operations without sufficient failure scenarios, it may fail to detect early signs of equipment degradation. This limitation can result in undetected malfunctions, increasing the risk of equipment failures and safety incidents. Another example of failure is sensor malfunctions and faulty data inputs [126]. AI systems depend on real-time sensor readings to regulate processes, but incorrect or inconsistent data—caused by calibration errors, physical damage, or interference—can lead to flawed assessments [127]. This may result in improper adjustments to operational parameters, reducing efficiency and potentially creating hazardous conditions.

AI failure does not always indicate an error in the system itself; it can occur when automation obscures important information, making problem detection more difficult for human operators [128]. Skjerve and colleagues examined how automation masked failures in a pressurized water reactor simulation, affecting operators' ability to diagnose issues [129]. In one case, a valve failure and a small leak developed in the letdown system, which regulates reactor pressure. The system's automatic controller adjusted its settings to compensate, concealing the leak and delaying operator detection. A second leak then occurred in the primary purification system, also within the letdown system. Automation responded by activating charging pumps to maintain pressure, further masking the problem. As the system continued to adjust for these changes, operators had to diagnose both leaks without direct indicators, complicating their ability to identify the underlying failures. In this scenario, both failures had similar symptoms and occurred within the same system. The automation system responded as designed by adjusting parameters, such as activating controllers and charging pumps, to stabilize pressure levels. While this maintained normal operations, it also concealed important warning signs, making it difficult for operators to identify the underlying issues. As automation continued to compensate, the crew struggled to differentiate between the leaks and determine the cause.

This demonstrates a challenge of automation masking. Although the system functioned correctly by maintaining stability, its actions hid failures that required operator intervention. By compensating for pressure fluctuations, automation reduced direct indicators of malfunction, delaying detection and response. In this case, automation operated as intended but had unintended consequences that affected human decision-making. Similarly, AI systems designed to optimize performance may prioritize maintaining output over providing diagnostic information. Instead of indicating potential failures, AI-driven adjustments can obscure developing issues, increasing the risk of delayed corrective action.

The definition of automation failure is contested, as it depends on whether the focus is on system performance alone or its broader impact on operations [130]. One narrow perspective defines automation failure as a malfunction, such as when biased data leads to incorrect predictions or sensor errors produce inaccurate outputs [131]. Another broader definition considers failure to include cases where automation functions as designed but creates unintended consequences, such as masking faults and reducing human operators' ability to diagnose issues [132]. Mumaw, Dekker and Woods, and Van Paassen and colleagues analyzed automation failures in commercial aviation [133], [134], [135]. They question restrictive definitions of automation system boundaries, which some designers use to argue that no failure occurred when automation operated within its predefined limits, despite not assisting operators as needed. Their perspective adopts a systems-based approach, defining failure based on the system’s functionality in an operational context rather than solely on whether it performed within technical constraints.

However, broadening the definition of automation failures excessively can diminish its precision [136]. When categories become too general, distinguishing between failure types becomes more challenging, limiting the classification's usefulness. An unclear definition may also encompass an excessive range of cases, making validation and assessment more complex. A well-defined classification enhances practical application and ensures accurate evaluation. This distinction raises questions about whether automation failure should be limited to technical errors or if it should also encompass situations where automation affects decision-making and operational oversight. The debate highlights the ongoing need for a broader understanding of automation failure, one that accounts for both system performance and its impact on overall safety and efficiency.

There is no agreed-upon framework for defining automation failure, as different disciplines approach the concept with varying objectives [137], including human performance modeling, system design, testing, interface design, and failure analysis. The lack of clear definitions has led to uncertainty about which types of failures should be examined when studying human responses to automation mishaps [138]. Kanaan and Donmez observe that research in cognitive engineering often relies on a limited set of automation failure scenarios, which may not capture the full range of possible system disruptions [139]. Definitions that are too broad or too narrow reduce the ability of human performance models to explain operator interactions with automation. AI failure should be understood as a disruption that affects an entire system, influenced by operational conditions and the interactions between automated processes and human oversight. There is ongoing debate about the appropriate tools, frameworks, and models for defining and categorizing automation failure.

Skraaning and Jamieson have proposed a taxonomy of automation-induced human performance challenges, highlighting its relevance for studying human-automation interaction [138]. The taxonomy describes three categories of automation-induced human performance challenges: elementary automation failures; systemic automation failures; and human-automation interaction breakdowns. Elementary automation failures occur when specific automation components or functions fail in isolation. Some describe these as failures within the "support system." [131]. In contrast, systemic automation failures result from interactions among equipment, functions, and automated logic, disrupting system regulation and information processing. These failures cannot be attributed solely to a support system or an underlying system but emerge from the interconnectedness of system elements, where disruptions spread through automated processes and affect overall performance. Unlike these two categories, human-automation interaction breakdowns do not involve a loss of automation functionality. Instead, they result from a mismatch between automation design and human capabilities, limitations, and needs. In these cases, automation functions as intended but does not align with human decision-making processes, leading to errors or inefficiencies.

After the taxonomy was published, researchers have contributed by proposing revisions, comments, and expansions to improve the taxonomy and its applications. For instance, a review examines automation failures in surface transportation [139], [140], highlighting distinctions in operational context and user expertise. Industrial operators receive specialized training, while the general public often lacks formal instruction, influencing their interactions with automated systems. These differences affect the occurrence and management of automation failures. Modifying the taxonomy may be necessary to reflect these distinctions. Furthermore, it was acknowledged that the taxonomy does not address the automation/autonomy conundrum [134], [141], [142], [143], [144], which involves determining the appropriate level of automation while maintaining effective human oversight. Increased automation can reduce human engagement, affecting intervention capability in unexpected situations. Conversely, limited automation may place higher cognitive demands on operators, increasing the likelihood of errors. Finally, several commentaries highlight the need for automation failure scenarios in cognitive engineering research to reflect real-world conditions more accurately [19], [144], [145], [146], [147].

This thesis relies on the taxonomy of automation failure to examine key challenges. It will analyze user expertise and operational context, and study automation failure scenarios to align with real-world conditions. The objective is to improve understanding of these failures and assess their effects on system design and human performance.

Chapter 5 will examine an instance of elementary automation failure. The AI-based automated diagnostic tool does not provide an incorrect diagnosis but fails to provide a diagnosis altogether. This occurs when the system is unable to match incoming alarms to the historical dataset, which it relies on for fault identification. Due to limitations in data availability or processing, the system fails to recognize a match, preventing operators from receiving useful diagnostic information and hindering their response.

Chapter 6 will focus on a systemic automation failure, in which the AI applies incorrect logic when diagnosing a fault. This failure does not stem from a single component malfunction but from interactions among automated processes that influence decision-making. The AI processes available data and follows predefined diagnostic rules, but errors in its reasoning lead to incorrect conclusions. As a result, operators receive inaccurate information, which may delay or misdirect their response to the fault.

2.3.4 Routine-Failure Trade-off Model

When automation fails, the expectation is that human operators will take over and correct the issue [148], [149], [150]. However, this transition presents performance challenges, particularly when operators have had limited engagement with the task due to high levels of automation [151]. The Lumberjack Model [152], [153] describes a decline in task performance when automation fails beyond a certain threshold. This model, also known as the Routine-Failure Trade-off Model, explains how automation reduces workload during normal operations but can also lead to reduced operator engagement and skill retention [154], [155]. As a result, when automation fails, operators may struggle to intervene effectively. This can be illustrated with the example of a lumberjack who replaces an axe with a chainsaw [132]. The chainsaw allows for easier tree cutting, reducing physical effort. However, if the chainsaw stops working, the lumberjack may have difficulty using the axe efficiently due to a lack of practice. Similarly, operators who rely heavily on automation may experience skill degradation, making it more difficult to respond to failures. This concept can also be applied to AI systems that process large amounts of data [156]. If an AI model overfits, removes too much data during preprocessing, or prunes models excessively, it may lose important information needed for accurate decision-making, reducing its effectiveness in changing conditions [151].

Automation failures can be harder to detect and address when operators have limited involvement, reduced engagement, and a tendency toward complacency [157]. When automation assumes control of routine tasks, operators may have fewer opportunities to engage with the system, which can weaken situation awareness [19]. This reduction in awareness makes it more difficult to detect anomalies, assess system status, and anticipate future conditions [19]. Additionally, while automation typically reduces workload during normal operations, failures can cause a sudden increase in cognitive and physical demands [158], [159]. Operators may need to quickly process information, diagnose issues, and take corrective action under time constraints, which can lead to errors or delays in response [160]. Over time, frequent reliance on automation can also result in overreliance, where operators assume the system will function correctly and become less vigilant in monitoring its performance [161], [162], [163]. Given these challenges, research on automation failure examines task performance, workload, situation awareness, trust, and other cognitive factors to better understand how automation affects human decision-making.

2.3.5 Implications for Human Performance

AI systems are susceptible to failure in complex, dynamic environments where conditions evolve unpredictably [164]. While automation and AI-based fault detection and diagnosis offer significant advantages in safety-critical domains [17], [165], their limitations pose substantial risks, particularly during abnormal situations [166]. When AI fails, operators may struggle to regain situational awareness, especially if prolonged automation use has diminished their engagement with the system [167]. The consequences of undetected AI failures in high-risk environments can be severe, underscoring the necessity of maintaining human oversight even as automation becomes more advanced [168]. To provide deeper understanding of these challenges, the following sections will examine the interplay in human-AI collaboration with emphasis on the concepts of situational awareness, cognitive workload, trust and reliance, the out-of-the-loop phenomenon, and the stages of human information processing.

For the clarity of this thesis, we will definitions these key terms:

- **Endsley's Situational Awareness (SA):** Individual's or team's ability to perceive, comprehend, and anticipate relevant information in each environment to support effective decision-making and action [169]. The model emphasizes the role of tools and systems, such as AI, in supporting human decision-making by improving information processing and enhancing situational understanding.
- **Cognitive workload:** Mental effort required to perceive, process, and respond to information while performing a task. It is influenced by task complexity, time constraints, information density, and an individual's cognitive capacity [170].
- **Trust in the machine:** Degree of confidence an individual has in an automated system's reliability, accuracy, and effectiveness in performing its intended function [171]. Trust is

influenced by several factors, including the system’s past performance, transparency, predictability, and the user’s level of expertise.

- **Reliance:** The extent to which an individual depends on an automated system to perform tasks, make decisions, or support operations [172]. Reliance is influenced by trust, system reliability, and the perceived benefits of automation in reducing workload or improving performance.
- **Out-of-the-Loop phenomenon (OOTL)**: Cognitive state in which a human operator loses situational awareness and the ability to effectively intervene in an automated system due to prolonged disengagement or lack of direct interaction [166].
- **Stages of Human Information Processing**: A sequential model that describes how humans perceive, interpret, and respond to information [173]. Helps explain human cognitive limitations and the importance of designing user interfaces, automation, and training programs that align with human cognitive capacities.

Endsley discussed the impact of human-AI task delegation when examined through the stages of human information processing: perception, decision-making, and action execution [174]. At the **perception stage**, AI processes raw data into meaningful insights, enhancing SA by enabling proactive adjustments and anomaly detection. When AI effectively integrates and presents contextual information, it reduces cognitive workload, improves operator engagement, and minimizes OOTL effects. However, AI-generated data can be flawed due to inaccuracies, omissions, or misinterpretations, distorting an operator’s perception of system status. Over-reliance on AI may lead to operator disengagement. Trust in AI is dynamic—frequent errors undermine confidence, leading to under-reliance, while consistent AI performance may cause over-reliance and complacency.

During the **decision-making stage**, AI systems analyze complex data patterns to prioritize options. However, the transparency of AI processes significantly influences operator trust and SA. A lack of clear reasoning behind AI recommendations can lead to uncertainty, increased cognitive workload, and OOTL phenomena, where operators become disengaged and less effective in intervention scenarios. Conversely, when AI systems provide erroneous or opaque recommendations, operators may need to engage in additional verification processes, elevating cognitive demands and response times. This situation can degrade decision accuracy and SA, particularly under time-sensitive conditions.

At the **action implementation stage**, AI systems autonomously execute tasks such as adjusting machinery or initiating system shutdowns, thereby reducing human workload. However, over-reliance on automation can lead to the OOTL performance problem, where operators experience diminished SA and a decline in manual control skills. This phenomenon leaves operators less prepared to intervene effectively during automation failures, potentially compromising safety in critical environments.

2.3.6 Related work

Now that we have established an overview of FDD, the role of AI in these systems, and its impact on human performance, we turn to prior research examining their practical applications in alarm floods and other operational contexts.

AI-based FDD systems leverage machine learning algorithms to process extensive sensor data, recognize patterns, and detect faults in real time. Research indicates that AI-driven methods outperform traditional fault detection approaches in both speed and accuracy [175]. For instance, predictive models have successfully identified equipment failures early, allowing manufacturers to reduce downtime and improve productivity [84]. Chang and colleagues demonstrated that an AI-driven system for diagnosing operational issues in solar energy projects achieved a fault detection precision of 98.6% [176]. Similarly, another study reported that an AI-based system deployed in a chemical plant maintained a 98% agreement between predicted and actual anomalies over three months, significantly enhancing fault detection reliability [177]. Researchers applied deep learning to the Tennessee Eastman process, a benchmark for industrial chemical production, achieving 95.6% fault detection accuracy [178].

Beyond fault detection, AI-based FDD systems play a crucial role in alarm management and operator performance. Past research has investigated AI-assisted alarm systems within a high-fidelity ethylene manufacturing simulator, focusing on their impact on operator workload during episodes of alarm flood [179]. Using the NASA-TLX scale, the study evaluated three alarm management strategies—no rationalization, rationalized alarms, and smart alarms—under both manual and AI-based automation. Findings revealed that AI-enhanced smart alarms reduced operator workload and material losses during abnormal conditions, particularly during episodes of alarm flood. Similarly, researchers have introduced a proactive alarm reduction methodology for nuclear power plants, prioritizing alarms to reduce cognitive overload and improve situational awareness in alarm-heavy environments [180]. Testing with eight nuclear power plant operators demonstrated that SA improved significantly when proactive alarm reduction was applied, helping mitigate the effects of alarm floods.

These findings highlight the role of AI-driven FDD and alarm systems in enhancing fault detection, reducing operator workload, and improving SA during alarm floods. These results are particularly relevant to this thesis, which investigate the impact of an AI-based fault diagnostic tool in a chemical plant simulator during alarm flood scenarios.

CHAPTER 3 RESEARCH APPROACH

3.1 Research problem

Modern process industries, such as chemical plants and power generation facilities, depend on FDD systems to monitor operations, identify anomalies, and support decision-making. The emergence of AI-driven diagnostic tools has enhanced fault detection and analysis, improving plant efficiency and reducing operational downtime [175], [176], [178]. However, these AI-based FDD systems are fallible, as they are susceptible to misdiagnoses, false alarms, and undetected failures [84], [115], [133], [177]. In safety-critical environments, such inaccuracies can lead to erroneous operator decisions, heightened cognitive workload, and reduced SA.

A critical challenge in AI-based FDD lies in its integration with alarm management systems and understanding its impact on human performance, particularly during abnormal situations. While process alarm analysis algorithms have been refined using large datasets, their effectiveness remains largely unvalidated in high-fidelity simulations with real operators. Furthermore, limited research has examined how human operators interact with imperfect AI-based diagnostic tools in realistic operational settings.

Current studies on human-AI collaboration in fault diagnosis primarily emphasize algorithmic performance rather than human factors, such as trust, reliance, and SA during alarm floods. Although some human-in-the-loop experiments exist [176], [179], [180], few investigated how AI reliability influences operator decision-making in high-pressure environments. A major concern is the potential for OOTL effects, where excessive reliance on AI diminishes an operator's ability to intervene effectively when automation fails. This thesis addresses these gaps by assessing how AI-based diagnostic tools affect operator performance and decision-making during alarm floods in a high-fidelity process control simulator.

3.2 Research objectives

This thesis aims to investigate the impact of an imperfect AI-based FDD tools on operator performance and decision-making during alarm floods using a high-fidelity process control simulator. The research objectives are structured into three key steps:

1. Development of a high-fidelity simulator for research studies

- Design and develop a high-fidelity process control simulator that realistically replicates industrial fault scenarios and alarm flood episodes.

- Validate the simulator's effectiveness in simulating a fault detection and diagnosis system within a chemical plant operational context.

2. Investigating whether an automated diagnostic tool supports human performance during alarm flood episodes

- Assess the reliability of an automated FDD tool in assisting human operators during alarm flood episodes and fault diagnosis scenarios.
- Analyze whether the AI-based automated diagnostic support tool enhances or hinders operator performance in high-stress environments.
- Complete a human-in-the-loop case study using the chemical plant simulator.

3. The effects of an imperfect AI-based automated diagnostic tool on human performance during failure management

- Examine how AI reliability influences operator trust and reliance.
- Investigate the impact of an imperfect AI on situational awareness (SA) and cognitive workload, particularly in high-pressure failure management situations.
- Identify conditions that contribute to out-of-the-loop (OOTL) effects, where excessive reliance on AI impairs operator intervention capabilities.
- Identify strategies to optimize fault diagnostic tool, design for human-AI collaboration, and enhance operator training programs for safety-critical industrial environments.

3.3 Research methodology

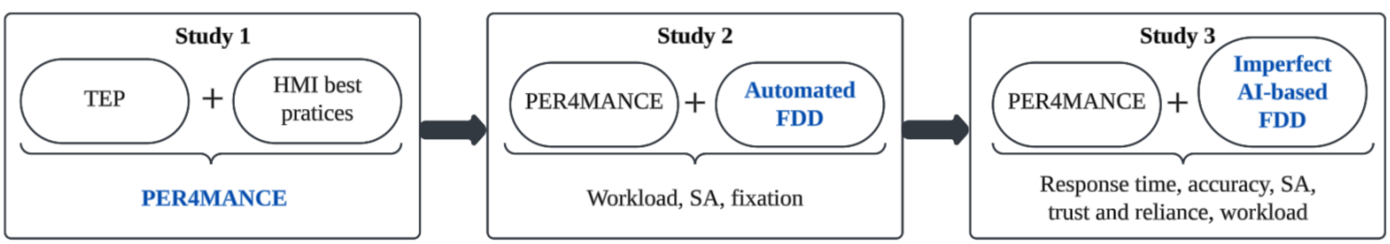

Figure 3.1 Overview of this thesis's three studies

This research follows a structured methodology across three studies, each building on the previous findings to examine human-AI interaction in abnormal situation management (**Figure 3.1**). Each study forms the basis of an article. The first study focuses on the development of PER4Mance, a chemical process simulator. Built using the Tennessee Eastman Process and designed according to

HMI best practices and industry standards, the simulator provides a controlled environment for studying episodes of alarm flood and AI-based fault diagnosis.

The second study evaluates an automated diagnostic support tool incorporated within PER4Mance to assess its impact on operator performance during alarm flood episodes. The study measures situational awareness, cognitive workload, and eye-tracking data to determine whether the tool improves performance.

The third article expands on the findings of article 2 by examining how AI failures affect operator response time, decision-making accuracy, situational awareness, trust, and reliance. It investigates whether operators recognize AI misdiagnoses or rely on incorrect recommendations, particularly under high cognitive workload and alarm flood episodes. The study provides insight into human-AI interaction when AI reliability varies.

This progression moves from system development to human performance assessment and finally to the impact of AI failures, ensuring a structured investigation into AI-assisted decision-making in industrial environments.

3.3.1 Study 1: Development of a human-machine simulator environment

Objective: Design a realistic and working chemical plant simulator based on the Tennessee Eastman Process (TEP) using HMI guidelines and industry standards.

3.3.2 Study 2: Impact of an automated fault diagnostic tool on human performance

Objective: Investigate the difference in operator performance with and without the use of an automated fault diagnostic tool during an abnormal situation.

3.3.3 Study 3: Impact of an AI's misdiagnosis during an alarm flood episode on human performance

Objective: Investigate the impact on operator performance during a process abnormal situation when the diagnostic tool fails (provides a misdiagnosis).

3.4 Thesis Hypothesis

Based on the literature review conducted, we formed three hypothesis that we will investigate in the following studies. Chapter 7 will evaluate whether these hypotheses were confirmed based on our main findings.

H_1: The implementation of a diagnostic tool enhances operator performance during episodes of alarm flood.
H_2: The impact of a diagnostic tool is significantly greater when managing difficult alarm flood episodes compared to easier ones.
H_3: Operators are capable of identifying incorrect AI diagnoses during episodes of alarm flood.

CHAPTER 4 ARTICLE 1 : PER4MANCE PROTOTYPING ENVIRONMENT FOR RESEARCH ON HUMAN-MACHINE INTERACTIONS FOR ALARM FLOODS MANAGEMENT: THE CASE STUDY OF A CHEMICAL PLANT PROCESS CONTROL

This article was published in the Proceedings of the Human Factors and Ergonomics Society Annual Meeting on 27 October 2022. doi:10.1177/1071181322661248

Karine Ung (Polytechnique Montréal), Omar Nemer (Polytechnique Montréal), Aswin Krishna (Indian Institute of Technology Guwahati), Moncef Chioua (Polytechnique Montréal), Philippe Doyon-Poulin (Polytechnique Montréal)

4.1 Abstract

Alarm floods are dangerous because the quantity of alarms triggered is too numerous for operators to reliably implement the right corrective action. Process operators of complex systems, such as chemical plants or nuclear power production, are faced with alarm management systems that can be better built in consideration of human capabilities and limitations. Developing human-machine interfaces (HMIs) that better support operators is critical for ensuring the safe and reliable operation of critical systems and processes. The research team has developed an accessible and adaptable prototyping environment dedicated for research on alarm management and human-machine interactions in the process industry. The method used was to build on the Tennessee Eastman Process (TEP) simulator and incorporate Human-Machine design guidelines. The results are an open-sourced prototyping environment that incorporates data from a real chemical plant and integrates true alarm data and thresholds. At the end of this article, we share the Github link to the entire MATLAB, Simulink and App Designer files of PER4Mance: a prototyping environment for research on human-machine interactions for alarm flood management.

4.2 Introduction

Process alarms play a significant role in maintaining a chemical plant's safety by providing a layer of protection in preventing the occurrence of faults from escalating into process hazards. Alarms aim at helping the process operators keep the plant within normal operating conditions. They provide an indication to the operators that their action is required to fix a fault or to prevent an undesired consequence. Throughout the years, the number and frequency of alarms have increased with technology. In the days of pneumatic controls, installing a new alarm had significant financial costs. The addition of an alarm with mechanical panels required adding light indicators and connecting them (hydraulically) to the sensor. As the number of alarms grew during plant operations, it reached a point where there was no longer any space available on the dashboard to add new alarms [181]. With the use of computer-based control systems, alarms became digital and the operator can configure them by defining thresholds for triggering the alarm [40]. Therefore, adding new alarms no longer had any financial costs or need for additional equipment. Furthermore, with the discovery of each new fault, alarms were added to the alarm system. As a result, the number of alarms has continued to increase over the years to a point where alarms could no longer be handled effectively [48]. It is common in a process control plant to have well over thousands of alarms per day, a number exceeding the recommended maximum manageable rate of 300 alarms per day [20].

Detrimental effects of alarm floods on safety and performance are documented in several application domains. In public transportation, automatic train control systems generate alarms to notify train dispatchers of the presence of faulty circuits. The rate of alarms can sometimes reach 8,000 per week and cause the dispatchers to become desensitized to the alarms [7]. In healthcare, the constant alarms from blood pressure machines, ventilators, heart monitors, etc., can cause health professionals to "tune out" the sounds. Alarm desensitization has been highlighted as a widespread problem in hospitals and many alarm-related deaths and injuries have been reported over the past few years [45]. Finally, in the aviation sector, the occurrence of unreliable alarms has shown to foster mistrust and complacency in airline pilots. Studies have shown that alarm-related problems frequently occur across flight operations and that false and incorrect alarms remain a significant concern in aviation [14]. Research on alarm flood mitigation can be useful in chemical process control as well as across multiple other domains.

According to the Abnormal Situation Management (ASM) Consortium, petrochemical plants suffer one major accident every three years on average [51]. An important number of these incidents reported were due to poor performance of alarm systems, resulting in plant damages, loss of production, and environmental incidents.

One of the most famous incidents in the field of alarm management is the Milford Haven incident at the Texaco refinery in Pembroke, South Wales, in July of 1994. A massive explosion resulted from 20 tons of flammable hydrocarbons being released from the knock-out pot on the flare header, leading up to hundreds of alarms being triggered. The Health Safety Executive's investigation report [43] identified the concern that alarms can overwhelm the operator, and instead of improving safety, can have the opposite effect and contribute to the incident.

This example illustrates that a fault can affect multiple related systems and trigger an overwhelming number of alarms. An alarm flood is defined as 10 or more annunciated alarms in a 10-minute period per operator [38]. In ISA-18.2 it is stated as: “A condition during which the alarm rate is greater than the operator can effectively manage [182].”

Alarm floods are troublesome because the quantity of alarms triggered is too numerous for operators to manage, making it difficult to implement the right corrective action. A fault can lead to a cascade of alarms, or multiple faults can occur during the same time period. Both scenarios can lead to an alarm flood, without any alarm differentiation between the separate faults. This phenomenon can affect hundreds or even thousands of alarms, with many unnecessary and redundant alarms resulting from the same root cause being enunciated and displayed to the operator. The discrepancy between the amount of information presented and the amount of information to which individuals can effectively manage leads to increased workloads, human error, and decreases in efficiency [42]. Despite improvements in alarm rationalization and prioritization processes, alarm floods are still a significant issue in abnormal situation management [31]. In alarm flood situations, one of the only responses available to the operators is to acknowledge and silence the alarms [41].

ANSI/ISA-18.2 Management of Alarm Systems for the Process Industries and the EEMUA 191 Human-Machine Interfaces (HMI) are standards providing guidelines for alarm systems management in process control plants. However, alarm systems built using these standards still need to be tested in a safe environment with human operators prior to the implementation in real operating industrial processes [36].

There are existing prototyping tools or simulation environments available for HMI test, but with limited availability. For instance, the company Corys [183], provides high-fidelity and dynamic simulators. Their simulator has been previously used in a human-in-the-loop study which investigated the impact of alarm management system design i.e. alarm rationalization, on the process operator’s workload [179]. However, the simulator comes at a financial cost that limits its accessibility to the public. Other researchers code their own simulator [184], but their simulator and its codes are not made available to the general public.

Alarm systems designed according to safety considerations provide the primary source of warning for operators when it comes to abnormal situations. Still, to the best of our knowledge, there has been no freely available and open-source process control simulator environment that has been developed to provide a platform for research on human-machine interactions during alarm floods. Following the approach of Simonson et al. 2022 [179], we developed a human-machine prototyping environment that can be used as a research tool to investigate alarm flood management in a process control environment. We aimed at creating an environment that can promote the study of the impact of machine learning-based decision support systems to guide the operator during periods of alarm floods, what we'll call the "diagnostic tool". The next section presents the development method, followed by validation results, discussion and conclusion.

4.3 Method

4.3.1 Step 1 - Tennessee Eastman Process (TEP)

The first step in creating the prototyping environment was to use the Tennessee Eastman Process (TEP) simulator to represent a chemical process control [185]. The TEP is a realistic simulation of a chemical process that runs on MATLAB [186]. It consists of five main process units: a reactor, a separator, a stripper, a compressor and a condenser (**Figure 4.1**).

The process has a total of eight different chemical components identified as A through H. These components consist of three gaseous reactants, A, D, and E that are fed to the reactor, which contains a small amount of inert gas B. There is also the gaseous reactant C that is fed directly into the stripper. Liquid products G and H exit the stripper base and are transferred to subsequent units and cells. The primary objectives of the process are to maintain the specified ratio of G/H in the product and maintain the specified product rate during normal operation and process disturbances. There is also a liquid by-product F which is purged from the TEP. The operator can manipulate 12 input variables and monitor 41 output variables. The TEP simulator also has 20 pre-defined fault scenarios [187]. The process control community has used TEP extensively as a benchmark to compare the performance of control strategies, but has received little attention as a user-facing simulator [188].

4.3.2 Step 2 - TEP alarm dataset

As the TEP simulator did not comprise of alarms embedded in its program, the second step of the tool development consisted of adding an alarm dataset to the prototyping environment. We used the work from the IEEE TEP Alarm Management Dataset [189], where the authors identified the TEP variables with their alarm high and low threshold values. We programmed their alarm thresholds into our tool, so that the alarms are triggered at the correct threshold limits. Therefore, whenever a variable's actual value crosses the high or low threshold, the respective alarm is triggered.

4.3.3 Step 3 - Real-Time Data Exchange

The next step involved creating a real-time data exchange link between the TEP simulator and our prototype. By adding a scope block, the prototyping environment is able to locate the variables and read the data from Simulink [186]. We added a single scope block to the default configuration of the TEP at the output block of the variables. This enabled us to read the data of the variable outputs from our prototype. Furthermore, by adding the additional scope blocks to all the variables, we managed to capture the data generated by the simulator during its execution, and were able to display them in real-time on our prototype. In addition to reading the data, this also allowed us to make input changes to the variables during the simulation. It was therefore possible for operators to change the manipulated variables, i.e. the valves opening and setpoints, while the environment was running.

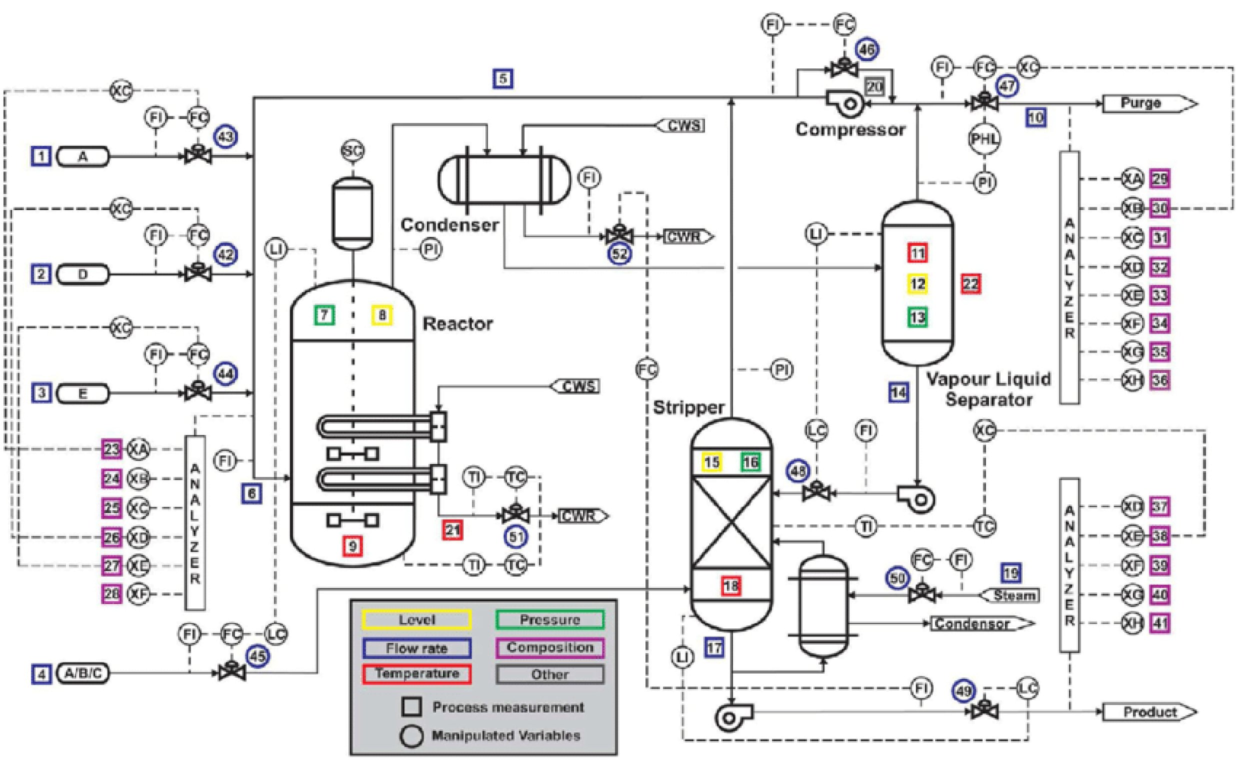

Figure 4.1 Piping and Instrumentation Diagram of the Tennessee Eastman Process [190]

4.3.4 Step 4 - Human-Machine Guidelines

Having the TEP simulator with its alarm dataset defined and the data exchange established, we were ready to design the tool's interface on MATLAB App Designer. We followed human-machines guidelines of ISA-101 which provides a design model called High-Performance (HP) HMI to design an interface that allows operators to detect, diagnose and correct efficiently dynamic operations in a process control environment [191]. More specifically, the standard provides guidelines on how to display information when developing the prototype's interfaces.

The presentation of the data should be done in a hierarchical manner across four levels. The first level is the most important and should present a global view of the whole process. It is also where information about the most critical equipment should be displayed. The second level is dedicated to the subunits of the system, with each subunit having its own view presenting more detailed information on its operating conditions than on the first level. The third level is an even more specific view of a particular piece of equipment of a subunit. Finally, the fourth level contains any other useful information that can help the operators make their diagnosis. For levels one and two, we identified the tasks the operators need to be able to perform, and defined the relevant variables. We omitted levels three and four because all the information that was identified as relevant during our analysis could be transmitted within the first two levels. Then, we defined the format for each variable (e.g., graphs, trends, thermometers, lists, etc.) depending on their context.

Following these HP HMI principles, we were able to identify where to present the 41 variables and their format, the key performance indicators (KPIs), the alarms, the diagnostic tool and the controllers for the manipulated variables. We will present them in the following section.

4.3.5 Step 5 - Implementation

We chose to use App Designer [186] as the development tool because it is an extension of Matlab, which was required for the TEP simulator to function properly. Since these three modules are under the same working environment, communication and data exchange was running properly. Moreover, the App Designer tool offers a library of objects (graphs, gauges, etc.) ready to use that can be dragged and dropped onto the interface.

4.4 Results

The prototyping environment consists of two interfaces, one interface open per computer monitor simultaneously. The first interface (**Figure 4.2**) represents the system overview, containing the global and critical information showing the system's health status (level 1). This overview interface has a panel on the top that provides the key performance indicators of the system. These are the inputs' flow rate, their concentration to the reactor, production rate, quality of the G and H products, production cost per hour and finally the concentration of the chemical components at the output, including the purge and the products. In this same section, on the right, we have the diagnostic tool which displays a solution when a fault occurs. The user of the prototype can choose to provide a correct solution, an incorrect solution or no solution at all. In the middle section, we integrated a diagram representing the logical flow of the TEP system from left to right so that the operators have a global view of the process.

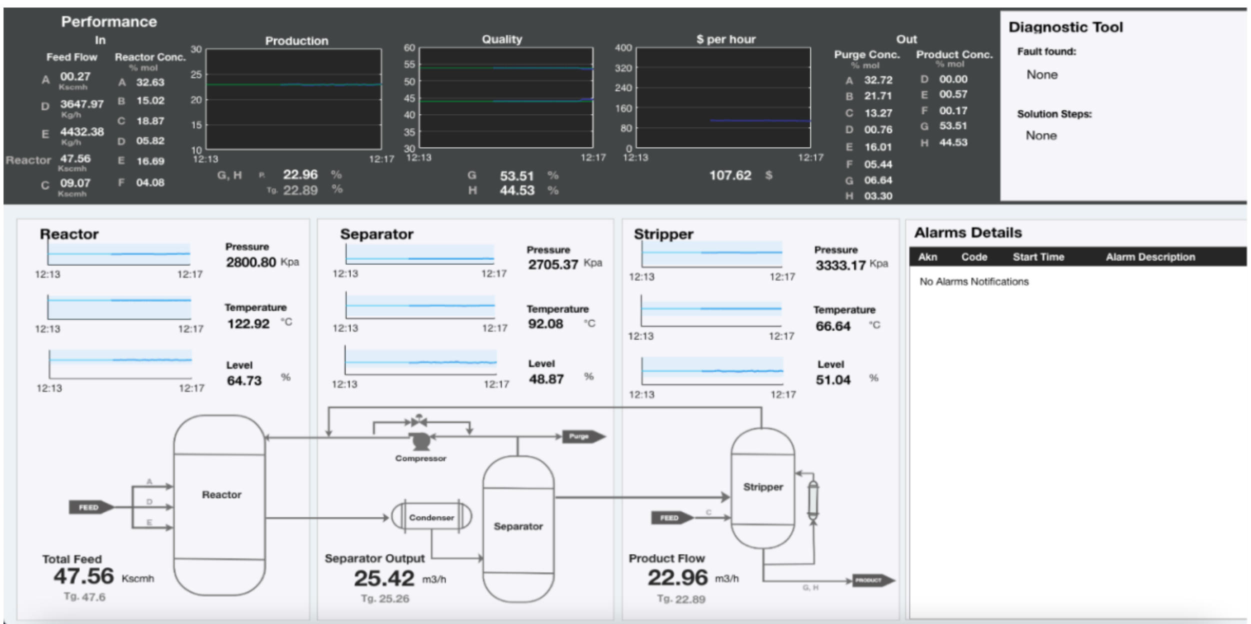

Figure 4.2 The environment system overview

These three blocks show the most critical equipment, i.e. the reactor, the separator and the stripper. For each of these equipment, we displayed its pressure, temperature and flow. At the bottom of the screen are three digital readouts indicating the incoming flow to the reactor, the outgoing flow from the separator and the overall flow of products for the stripper. To the right of these blocks is the alarm table. Under normal circumstances, there are no alarms displayed.

The second interface (**Figure 4.3**) represents detailed information per unit, displaying the variables related to the reactor, condenser, separator, compressor and stripper under different tabs (level 2). On the right side, the users can act on the process valves either in manual mode (openings adjustments) or in automatic mode (setpoint settings). If the control is in automatic mode, the operator can modify these setpoints. If the control is in manual mode, the operator can directly modify the valve opening.

At the top of the screen, there are tabs to navigate to other units of the system. The units on these tabs follow the process flow navigation from left to right. Some units are simpler than others, therefore, we combined them to save screen real estate; the condenser, separator and purge; and the stripper with the final product information.

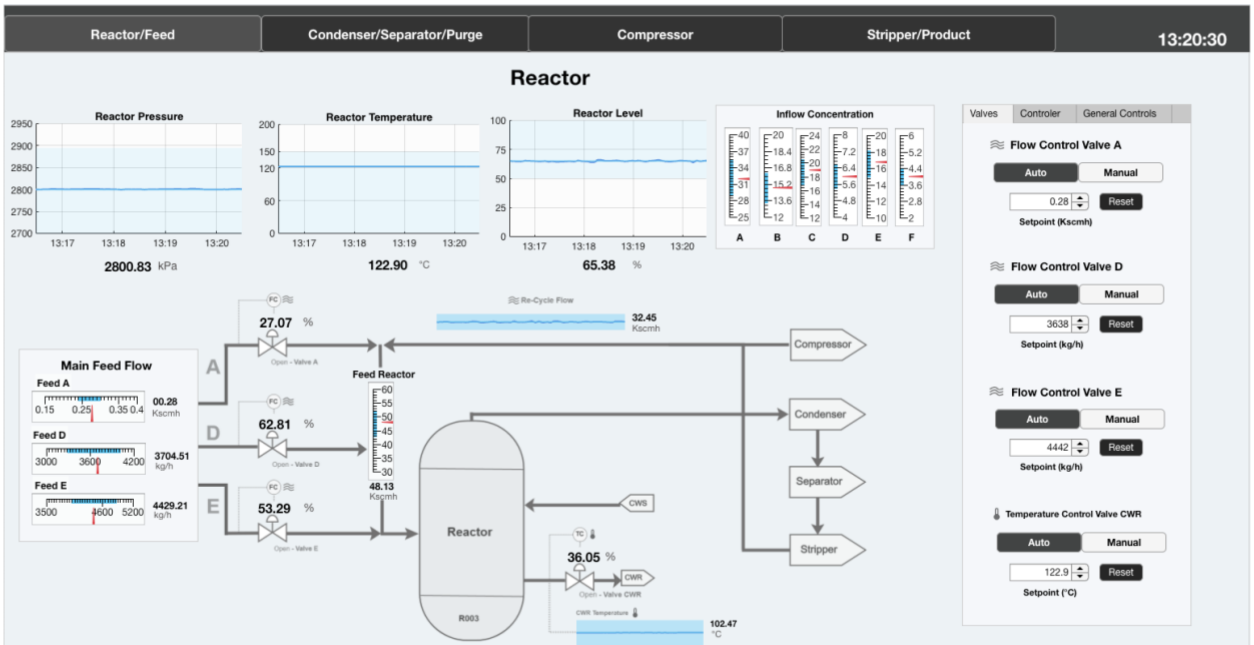

Figure 4.3 The reactor interface

The interface adds further information when the system is under a fault condition. **Figure 4.4** presents a fault condition scenario: a loss of feed of input A. The red dots indicate the variables and units have exceeded their normal threshold values and that an alarm has been triggered. They can be seen on the top KPIs section, but also on the overall TEP diagram. Furthermore, there is the alarm table on the bottom right side of the interface. There is a checkbox on each alarm line so that the operators can indicate that they have acknowledged the alarm. There is also the code of the alarm, its time of appearance, its description and finally the unit involved.

The prototyping environment follows the standards of a High Performance HMI: a two-level hierarchy was used and the data was grouped according to their corresponding sub-unit. Trend graphs and analog indicators were used to visualize if the value of a variable is within the normal range. The number of colors were limited by keeping the background gray, the operating limits in blue and the fault indications in red. By following these standards, the simulator environment closely resembles the interfaces used in the industry, and users have access to a functioning prototype that has an interface that represents those used in real-life-operations.

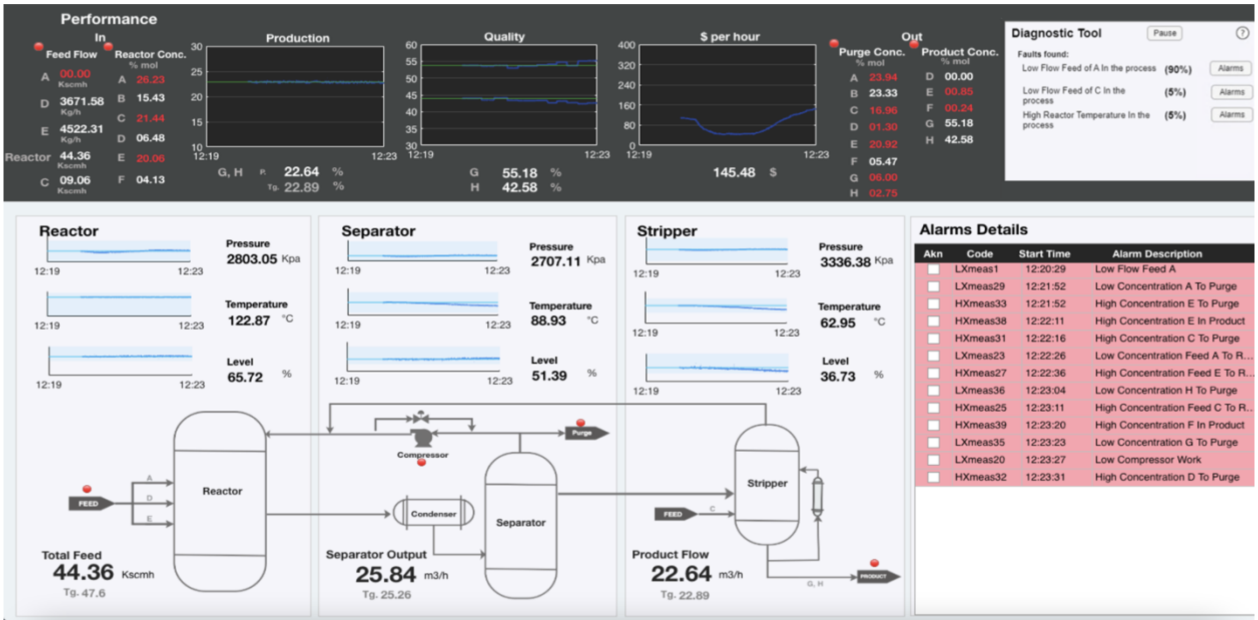

Figure 4.4 Abnormal condition: loss of flow A

4.5 Discussion

Although there are multiple process control simulator environments available in the market, not all are accessible and malleable. Our prototyping environment differs from others in the following ways: it has no financial costs, it is open-sourced and it is extremely modifiable. From our original files, users can change the codes and interfaces freely.

While developing this prototyping environment, there were a few limitations encountered. First, App Designer offered a limited library of graphical elements. Although the tool is very easy to use, the graphical elements provided by the program looked out-dated. The second disadvantage of this tool is that the more we added graphical elements to the interface, the heavier and slower the editing mode became. Finally, it is worth mentioning that the running speed to complete the simulation was less than 1 minute. The prototyping environment reflects the same speed as Simulink, and we therefore had to slow down the running speed on Simulink to be able to have a working prototype. Despite these limitations, users can change the alarm thresholds and behaviors, add or remove alarms, and configure the information provided by the diagnostic tool. Furthermore, users can remove the input controllers in case they want to reduce operator's control over the environment. We also developed this prototype to allow research with one operator, but modifications of the environment to study multiple operators simultaneously or team dynamics could be possible and worth investigating. We've created this prototyping environment to reflect our own goal, which is to study the impact of machine learning-based decision support systems, the diagnostic tool, to guide the operator during periods of alarm floods. But this prototype can be modified and adapted to countless other environments to study different aspects of alarm management in industrial settings.

4.6 Conclusion

This paper aimed to address the need for a prototyping environment to study human-machine in the process industry. We've proposed a prototyping environment that was built on the TEP simulator and HMI design guidelines and principles. With this public tool, we hope to encourage shared research on human-machine interaction and alarm management in relation to machine learning systems.

The next step for the research team is to test the prototyping environment with humans. Testing with humans will allow us to improve the prototype itself and to study the effects of the prototype's design on human cognition. Even though we followed HMI design principles, there are still many elements in the prototyping environment itself that can be improved, such as the way the variables are presented, the alarm presentation and the solutions conveyed by the diagnosis box. Also, testing the fault scenarios of the prototype with humans will allow us to perform fundamental research on alarm flood management and diagnosis, thereby investigating the interface elements that would help or hinder human diagnosis abilities. We have made this prototyping environment available to all in order to encourage shared learning and promote further work on improving the prototype.

The US Federal definition of research is "a systematic investigation, including development, testing, and. evaluation, designed to develop or contribute to generalizable knowledge" (45CFR46.102). We hope that our approach and design might generalize to other research in simulations or alarm floods in different domains. Don't hesitate to contact us for any collaborative work.

The PER4Mance (MATLAB, Simulink and App Designer files) are available to download via the following link: https://github.com/karine-ung/perf4mance

CHAPTER 5 ARTICLE 2: AUTOMATED DIAGNOSTIC TOOL SUPPORTS HUMAN PERFORMANCE DURING ALARM FLOODS: A CASE STUDY IN A CHEMICAL PLANT SIMULATOR

This article was submitted to the journal of IEEE Transactions on Human-Machine Systems on 12 November 2024.

Karine Ung (Polytechnique Montreal), Kairat Khismetzhan (The Hong Kong Polytechnic University), Moncef Chioua (Polytechnique Montreal), Philippe Doyon-Poulin (Polytechnique Montreal)

5.1 Abstract

During fault scenarios in complex systems, alarm management systems are used to assist the operator in solving the failure. Alarm floods situations are often difficult to manage for the operators due to the large quantity of alarms. Past work presented data-driven methods to help manage alarm floods. Yet, few research has investigated the interaction between operators' performance and alarm rationalization systems in alarm flood scenarios. This study aims to address this gap by introducing an automated diagnostic tool designed to help operators manage alarm floods. We developed an interface using a realistic chemical plant simulator based on the Tennessee Eastman Process (TEP) and added a diagnostic support tool for managing alarm floods. Twenty students enrolled in a chemical engineering program participated in the study and were presented with two alarm flood scenarios. The diagnostic tool was activated in one scenario and deactivated in the other. Results showed that using the diagnostic tool increased participants' situational awareness and assisted them in managing complex alarm flood situations. Eye tracking results showed that the diagnostic tool reduced the gaze on the alarms and increased the gaze toward areas of KPIs and diagnostic information. The results of this study illustrate the potential growth of the use of automated decision support systems in the process industry.

5.2 Introduction

According to the industrial standard ANSI/ISA-18.2 [38], an alarm is "an audible and/or visible means of indicating to the operator an equipment malfunction, process deviation, or abnormal condition requiring a response" and "an alarm system is the collection of hardware and software that detects an alarm state, communicates the indication of that state to operators, and records changes in the alarm state". Alarm systems have been an integral part of computerized monitoring systems, such as the distributed control systems (DCS) and supervisory control and data acquisition (SCADA) systems [192]. These systems control, monitor, and manage production in often large infrastructures, such as petrochemical operations, electric power generators, transportation systems and chemical facilities, among others [193].

Alarms are important because they provide warnings for operators during abnormal situations [194]. An abnormal situation is a disturbance or series of disturbances in an industrial process that causes plant operations to deviate from their normal operating state. During an abnormal situation in a large and interconnected infrastructure, a disturbance can cause the failure of multiple systems and trigger a cascade of associated alarms [36], [37]. These alarms are triggered in the control room and operators may not be able to properly investigate each of these alarms promptly [195]. This phenomenon is described as an alarm flood, defined as the duration where the rate of alarm annunciation is more than the response capability of an operator [39]. Alarm floods can lead to situations of loss of control, which in turn, may lead to accidents [49]. According to ANSI/ISA-18.2 (2009), an alarm flood is more than 10 alarms per 10 minutes.

During an alarm flood, each new system failure adds more alarms to the existing cascade of alarms, without any differentiation between the root causes [7]. Reviews of previous accidents involving an alarm flood revealed that it can affect hundreds or even thousands of alarms, with many unnecessary and redundant alarms being annunciated and displayed to the operators [14], [45]. This excess of information presents several human factors challenges, mainly due to the limited cognitive abilities of attention [196], [197]. The discrepancy between the amount of information presented and the amount of information to which individuals can effectively attend leads to increased mental workload, human error, and decreases in operational efficiency [42]. In such situations, one of the only responses available to the operators is to silence the alarms, sometimes without looking at them [41].

According to the Abnormal Situation Management (ASM) Consortium, petrochemical plants suffer on average one major accident every three years. The Engineering Equipment and Materials Users Association (EEMUA) stated that inadequate alarm systems “were a major contributor to incidents, which frequently involved the operator being overloaded with alarm floods” [3]. An example is the Milford Haven petrochemical plant accident at the Texaco refinery in Pembroke, South Wales, in July of 1994. An explosion resulted from 20 tons of hydrocarbons being released from the knock-out pot on the flare header. The two operators on duty were flooded with 275 alarms in the 10 min time window before the explosion. The numerous alarms and their lack of logic made it too difficult for the operators to manage the situation [43]. The Health and Safety Executive's investigation report identified the concern that alarms can overwhelm the operators, and instead of improving safety, can hinder the operator's diagnosis of the root cause. The contributing factors of the accident were the alarm flood, poor alarm management, and an alarm system that did not support fault diagnosis.

The poor design of alarm systems and misdiagnosis were reported to have contributed to accidents in the process industry [36]. Alarms being triggered in the order in which the disturbances occur do not support operators in finding the source of the problems. These accidents exemplify why alarm rationalization is essential to support the operators in finding the root cause of alarm floods. There is a need to provide rationale in the way alarms are shown to support the operators' diagnosis of the root cause during abnormal situations. A study found that 15 facilities with varying sizes had a high number of alarms installed, with a minimum of 500 and a maximum of 10,470 alerts [41]. During normal operations, the number of alarms ranged from 60-120 per hour. During abnormal conditions, operators experienced alarm loads of around 390-3750 per hour, and in one case over 300 alarms in just 5 min [198], [199] found alarms in various industries were significantly higher than EEMUA guidelines. These studies highlight the importance of proper alarm management and the need for better monitoring and control of alarm systems.

As a result, there has been increasing interest in the industry in addressing the issue of alarm floods and investigating methods to help operators diagnose faults. A promising avenue is the use of machine learning techniques that link the incoming flow of alarms to known faults and provide a diagnosis to the operators [200]. Further investigation is needed to study the difference between the use of "traditional" alarm management systems, which lead to the appearance of alarm floods, from "advanced and automated" systems, which offer support to the operator. Few research has investigated the interaction between operators and automated systems in alarm flood scenarios, and this study aims to address this gap.

The contributions of this paper are the following: we did a human-in-the-loop experiment with a diagnostic tool in a chemical plant simulator. We investigated how humans reacted to a diagnosis provided by the tool during an alarm flood. The rest of this paper is structured as follows. In this section, we provide an overview of related work in areas of alarm rationalization and human-automation interaction. In section III and IV, we describe the experimental methods and data analysis. In section V and VI, we present the main results and discuss the impacts of using the diagnostic tool on the operators' performance.

5.2.1 Alarm rationalization

Several guidelines have been developed to improve alarm management systems, such as the International Society of Automation and the ASM consortium [51], [182]. Various data-driven methods have been developed to enhance alarm systems. In the last few decades, algorithms and techniques have been created to reduce the number of alarms that are activated [50]. Among others, Foong et al. developed a fuzzy-logic based alarm prioritization (ALAP) system to prioritize alarms during alarm floods and reduce redundant or false alarms [201]. Higuchi et al. developed an alarm reduction method using data-mining to identify statistical similarities among alarms [202]. Cheng et al. used a modified Smith-Waterman algorithm to analyze alarm flood patterns and cluster similar ones [203].

Another methodology to mitigate alarm floods is that of alarm rationalization [51]. Alarm rationalization is a process that involves reviewing and validating alarms to ensure they are necessary for maintaining safety and normal operations [52]. It aims to reduce nuisance alarms and remove redundant ones, ensuring operators receive only those that require action [25]. This process increases efficiency and reduces time spent on identifying the true cause of alarm activations. Up to 80% of alarms during an event are redundant or nuisance alarms, which can divert attention and interfere with operator acknowledgment [204]. A subset of alarm rationalization worth mentioning is that of alarm classification. Alarm classification categorizes an incoming alarm flood on the basis that the new alarms are matched with a set of previously occurred alarm floods [62]. The ongoing alarm floods are matched to a known category and presented to the operator to help diagnose the fault causing the alarm flood [61]. Seyed Alinezhad et al. developed a semisupervised data-driven method for classifying ongoing alarm floods using historical data [63]. The method, based on the Gaussian mixture model, includes alarm clustering and labeling, and online early classification. The approach was validated using the Tennessee Eastman process (TEP) benchmark and an industrial alarm flood dataset. The results showed accurate early classification of alarm floods by considering historical alarm datasets [63]. Lucke et al. developed an alarm flood classification method that uses sequence mining and time series analysis to classify alarm floods based on past events [67]. The method involves two stages: detecting when the flood belongs to a new class and classifying when one class forms a basis to match incoming floods. The algorithm integrates a historical alarm classifier into fault detection and identification. A case study was conducted on an offshore oil-gas separation plant, revealing that the algorithm matched new alarm floods with historical floods corresponding to the same abnormal situation, enabling the classification and identification of the root cause of an alarm flood. They found an average accuracy classification of 92.2%.

Alarm flood classification targets recurrent types of alarm floods since it relies on training a classifier on past events. A limitation is that the algorithm is unable to provide a classification if an incoming alarm flood is not part of the historical set because the classifier wouldn't be able to match it. Therefore, the accuracy of all the methods does not reach 100% [67]. Even though the reliability rate isn't high enough to implement the algorithm in real-life process control environments, it is, to some extent, able to identify the root cause of alarm floods that are part of the historical set and propose the root cause to the operator. While these methods have significantly reduced alarm floods, they have not eliminated them [205].

The literature review explored data-driven methods to mitigate the impact of alarm floods [23]. A promising approach is alarm flood rationalization and classification based on historical events in the control room. However, few studies have involved humans to investigate the impacts of alarm flood management techniques on performance. This lack of performance-based research is likely due to the need for high- fidelity simulators of manufacturing processes to assess the impact of alarm design. This study aims to fill this gap by studying operators' performance in a high-fidelity chemical plant simulator.

5.2.2 Human-Automation interaction

Product refinement industries, such as mining, oil refining, chemical, and pharmaceutical manufacturing, use process control to monitor and react to abnormal processes for safety and efficiency [206]. Process monitoring requires demanding states of attention and skills from the operators, and automation can enhance these processes.

An automated system accomplishes a function previously carried out by a human operator and can help reduce operator workload in response to abnormal situations [207]. Research in the process control industry shows that operators perform better when permitted to work in-the-loop with automation [208]. Automation offers numerous benefits such as reduced workload, increased situational awareness, and improved performance [209]. However, inappropriate usage can negatively impact the operators' understanding of the system, task processes, and overall performance [210]. Negative impacts include excessive monitoring, boredom, over-reliance, low situational awareness, out-of-the-loop performance, and reduced trust in the automation [211]. Consequently, it is crucial to carefully consider the benefits and potential risks of automation in process control systems.

Previous work investigated the impacts of an automated alarm system in human-in-the-loop studies in process control simulators. A study examined the impact of alarm management and automation on process control operator workload (using NASA-TLX) and performance in a high-fidelity ethylene manufacturing simulator [179]. The experiment involved eleven console qualified operators in a simulator-based training exercise. The experimental design involved three levels of alarm management schemas (no alarm rationalization, with alarm rationalization, and smart alarm) and two levels of automation (no automation and with automation). Results showed that smart alarm management and automation can help operators reduce workload and material lost during abnormal operating conditions.

Jang et al. introduced a proactive alarm reduction method implemented in a nuclear power plant environment [180]. The researchers designed the alarm reduction method to investigate alarm processing techniques in coping with high volumes of alarms. They had eight nuclear power plant operators test the effectiveness of the alarm rationalization method. The results indicated that the operators’ situational awareness in the alarm reduction environment was higher than in the non-reduction environment. A study evaluated the effectiveness of a decision support scheme called Early Warning in a simulated setting of a chemical plant control room [50]. Early Warning predicted the time of occurrence of critical alarms before they were triggered. An experimental design was developed to assess the effectiveness of this decision support tool in enhancing operators' performance in specific tasks. Early Warning offered control room operators early warning of potential alarms within a specific time frame (e.g., in the next 90 sec.), enabling them to be proactive and take corrective actions before alarm thresholds are breached. Participants were asked to monitor the depropanizer unit and to diagnose the root cause of the fault. Results showed that Early Warning supports the operator’s diagnosis but does not enhance the accuracy of diagnosing the root cause.

Alarm management methods and automation are used in process control industries to prevent, manage, and mitigate abnormal processes. Under normal conditions, operators can usually passively supervise process units, focusing on efficiency through minor adjustments. However, when abnormal situations occur, operators need to proactively manage the situation by taking corrective actions. Automation is less error-prone and can produce repeatable actions, but it often fails to address unforeseeable abnormal situations. In contrast, humans can be flexible and produce creative solutions in response to unanticipated situations [26]. Despite automation and improved alarm management systems, humans remain crucial in controlling chemical plants, particularly during abnormal situations [173].

5.2.3 Wizard of Oz (WoZ)

Building automation technology in process control can be complex, time-consuming, and costly. To save time on resources and test the technology quickly, researchers can simulate the response of the technology by having a person “play” the role of the automation. This moderated research method is called The Wizard of Oz, WoZ [212]. In a WoZ study, participants interact with a system that appears to be autonomous but is actually partially or fully controlled by a human [213]. The WoZ method might be seen as a low-level deceit employed to manage participants’ expectations, but it has been used in several studies and is known to encourage participants’ natural behaviors to the new system [214]. The "wizard" acts as a proxy for the system, emulating its intelligence and interacting with the participant through an interface.

WoZ prototyping has been used in various contexts, including interface designs for automated cars [215], natural language dialogue systems [216], speech recognition systems [217], and even introducing children to machine learning concepts [218]. WoZ has also been used to test autonomous systems, as it allows the researchers to mimic the model's computations and receive feedback before the development process [219]. Recently, a study investigated the learning patterns for human and artificial intelligence (AI) teams using the WoZ method [220]. Teams of one human and an AI robot performed an Urban-Search-And- Rescue mission in a simulated environment. The robot was controlled by a WoZ researcher, allowing for the study of human-robot interaction without computational modeling of necessary robot competencies, such as environment sensing and natural language communication. This method allows researchers to test autonomous systems at a lower cost than developing a functioning system, and to gather early insights into the system's design, user perception, and behavior.

Despite more than 30 years of research in alarm management leading to the development of various data-driven methods and diagnostic tools, few studies have involved humans in investigating the implications for applicable industries. The purpose of this study is to conduct a high-fidelity human-in-the- loop experiment to quantify the effects of alarm rationalization in a chemical plant simulator. We will be using the WoZ method to simulate the effects of an automated diagnostic tool and its impact on the operator's performance.

5.3 Research objective

Extensive research has been done on developing and improving data-driven algorithms and methods to manage alarm floods. Alarm flood rationalization algorithms have often been tested with large samples of datasets, but not often challenged in a high-fidelity simulator with real operators. Few studies have investigated the interactions between humans and an automated fault diagnostic tool in a high-fidelity process control simulator. This paper presents a study to address this gap. This research aims to investigate how a fault diagnostic tool can support the operators during alarm floods in a control room. To this end, we investigated the effects of an automation-based fault diagnostic tool on the operator's performance during different alarm flood scenarios in a Tennessee Eastman Process (TEP) chemical plant simulator.

5.4 Methodologies

5.4.1 Participants

The subjects were 20 students in chemical engineering at Polytechnique Montreal University. 60% self-reported as female and 40% as male. 75% were between the ages of 20-29 years old, and 25% were 30 years old or over. An informed consent was obtained from all the participants.

5.4.2 Apparatus

5.4.2.1 Chemical plant simulator

We used the TEP as the simulator for representing a chemical process control. TEP is a well-established process control simulator that is downloadable onto MATLAB/Simulink [189]. TEP is a realistic simulator of a chemical process consisting of five main process units: a reactor, a separator, a stripper, a compressor, and a mixer. The process has a total of eight different chemical components identified as A through H. These components consist of four gaseous reactants: A, D, and E that are fed to the reactor, and C which is fed into the stripper. The reactor contains a small amount of inert gas B. The objective is to produce liquid products G and H, which exit the stripper base and are transferred to subsequent units and cells. There is also a liquid by-product F which is purged from the TEP. The operator can manipulate 12 input variables and monitor 41 output variables. The TEP simulator has 20 pre-defined fault scenarios [187].

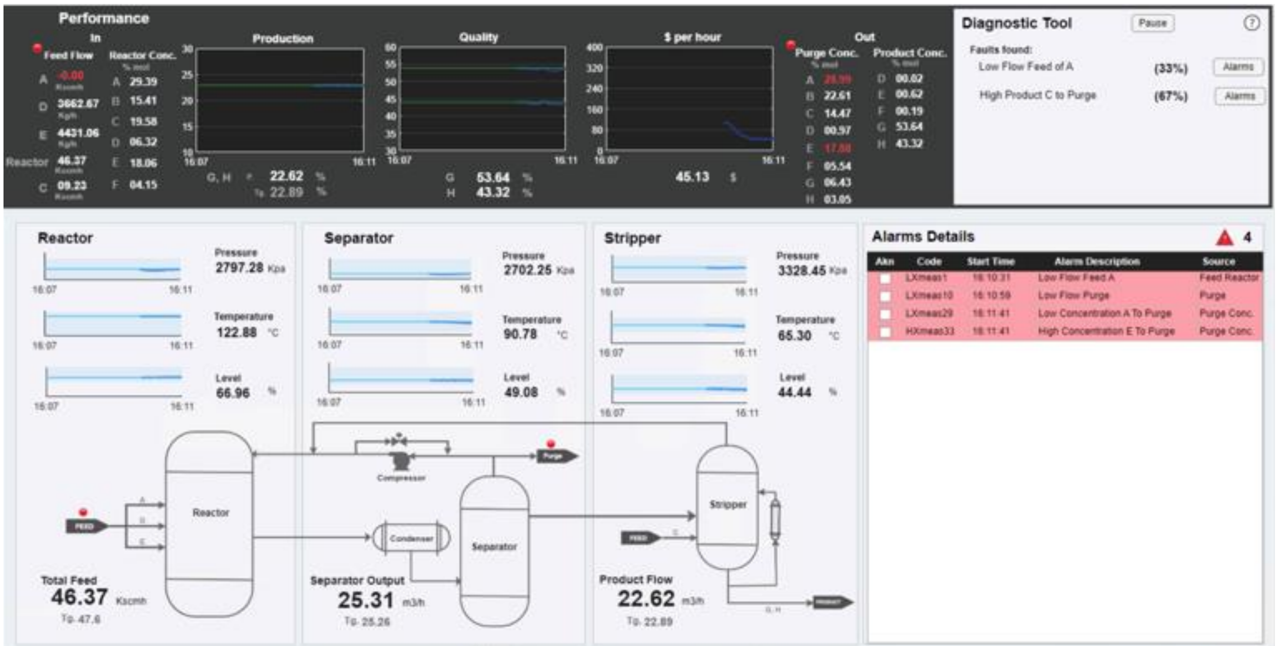

Figure 5.1 Overview interface of the chemical plant simulator.

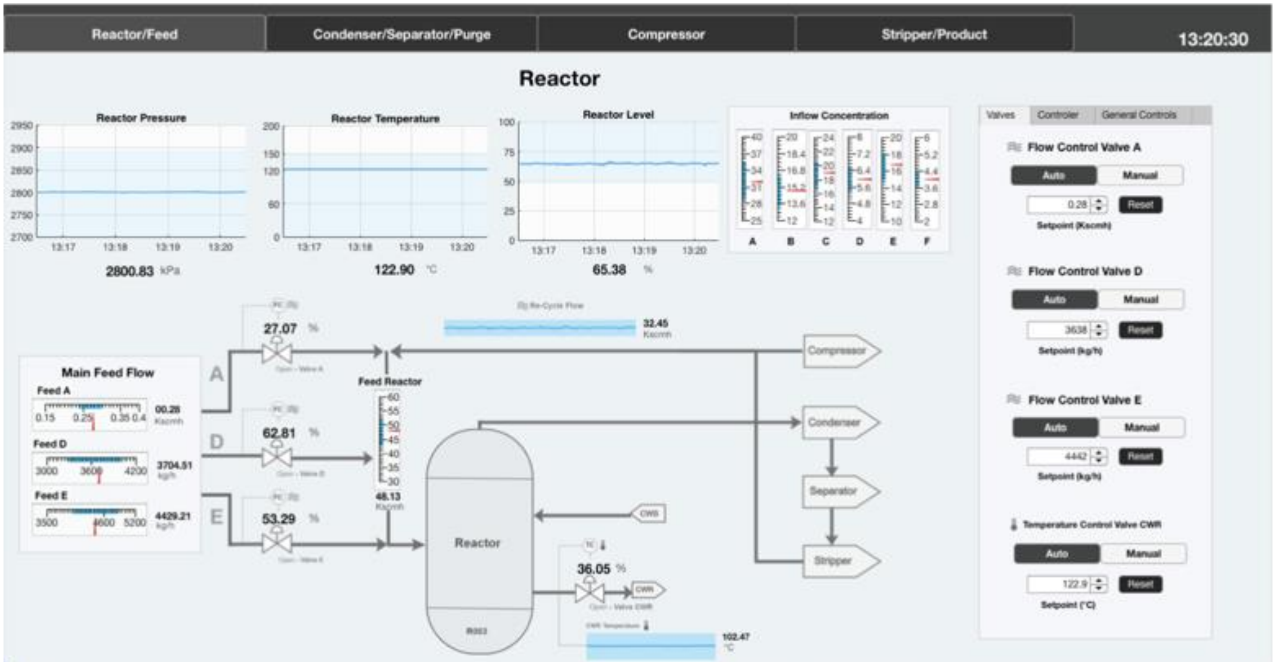

Figure 5.2 Detailed unit interface of the chemical plant simulator.

In a previous study, we developed PER4Mance, an open-source graphical user interface for the TEP simulator based on dashboards currently in use in the industry [221]. PER4Mance has two windows. The first window displays key performance indicators, the overall Piping and Instrumentation Diagram (P&ID), the alarm table, and the diagnostic tool (**Figure 5.1**). The second window shows time trends and the current value of process variables for each unit, organized as tabs, and the operator can control the unit's valves from a menu (**Figure 5.2**). The user can navigate between the units using tabs. Both windows were displayed at the same time on two computer screens. The alarms are triggered when the value of the variables exceeds its thresholds. Each variable has a low and a high threshold, within which are the normal operating values. When the value exceeds either threshold, an alarm appears in red on the overview interface along with a sound.

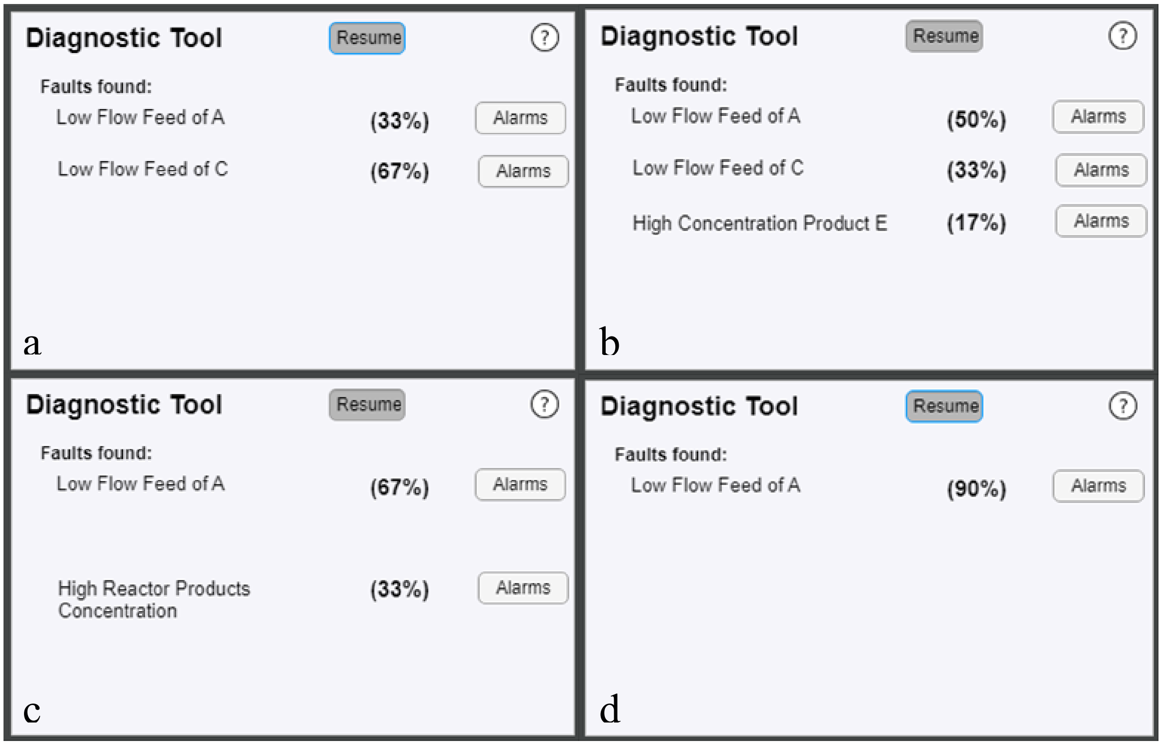

Figure 5.3 Diagnostic tool evolution for Fault 1: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

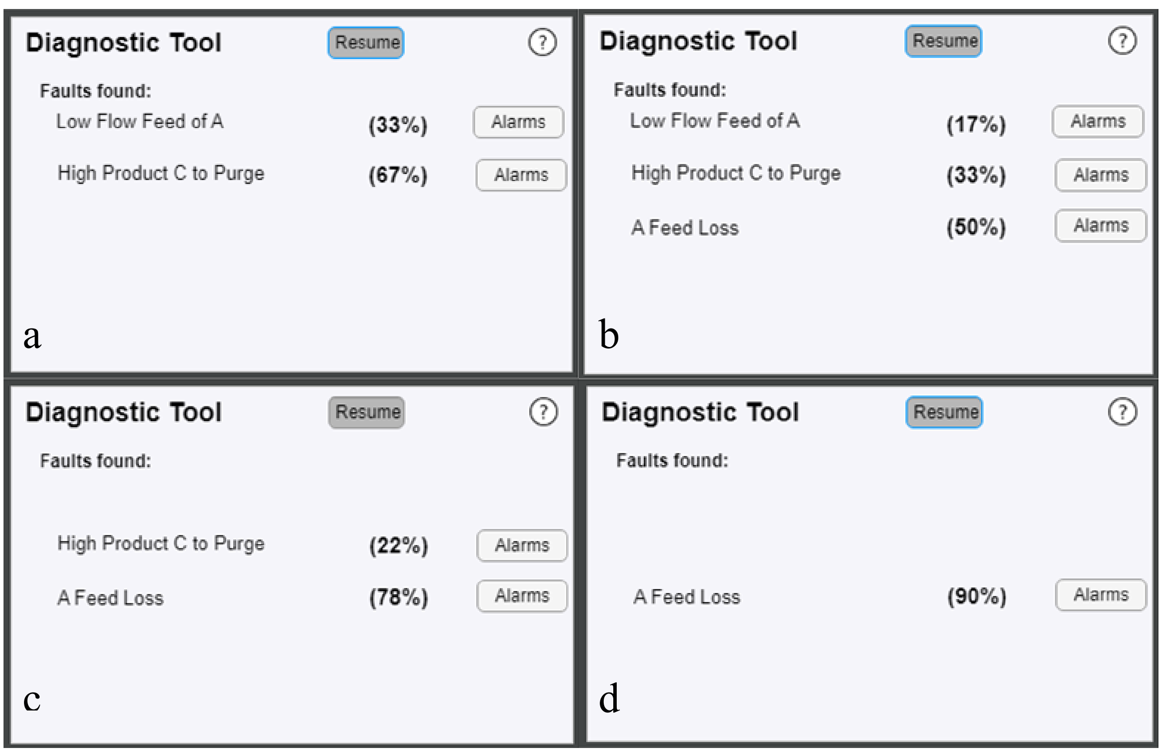

Figure 5.4 Diagnostic tool evolution for Fault 6: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

5.4.2.2 Diagnostic Tool

This study investigates the impact of a fault diagnostic tool on operator performance through a human-in-the-loop experiment. Training an algorithm to detect and identify the faults or alarm floods was going to be too time-consuming and costly. For this reason, we used the WoZ method to simulate the behavior of an automated fault diagnostic tool. Participants interacted with a diagnostic tool that they believed to be autonomous but was actually operated by a member of our research team. Participants thought that the fault diagnosis was being provided by the tool when it was activated by the researchers.

In the condition where the diagnostic tool was deactivated, the area did not provide any new information. It had the message “Fault found: Unknown” throughout the entire testing session. In the condition where the diagnostic tool was activated, the research team wanted the diagnostic tool to reflect a learning capability and to improve its accuracy of the likely fault as more information was available. In this case, the information provided to the tool were the alarms triggered. Hence, as there were more alarms triggered, more information was provided to help the tool classify the situation into one of different probable faults and the diagnostic message updates. Practically, a diagnostic message appeared after three alarms. The message was updated again after six, nine, and twelve alarms (**Figure 5.3** and **Figure 5.4**). After twelve alarms, the diagnostic was presented as the final message.

Furthermore, the research team wanted to reflect the limited confidence level of the algorithms. To do so, two to three possible diagnostics were provided at once, each with a confidence level expressed in percentage. When there were two or three possible diagnoses, the sum of the confidence levels was 100%. When there was only one diagnosis, the confidence level was at only 90%. After six alarms and again after nine alarms, the diagnostic message and its confidence levels were updated. After twelve alarms, the final diagnosis was provided. To simulate alarm grouping and to increase the tool’s transparency, clicking the “Alarms” button next to each fault highlighted the related alarms in the alarm table (**Figure 5.5**). Note that in the two failure scenarios participants experienced, the total number of alarms reached 15 and 27, respectively.

Figure 5.5 The “Alarms” button highlights in dark red the alarms related to the Low Flow Feed A fault selected in the alarm table.

5.4.3 Experimental variables

5.4.3.1 Independent variables

There were two independent variables: use of the fault diagnostic tool; and the type of fault used in the scenario.

Fault diagnostic tool

There were two experimental conditions: the activation or deactivation of the fault diagnostic tool. In condition 1, the fault diagnostic tool was deactivated. The only cues signaling the participants of a malfunction were the alarms and indications on the interfaces. In condition 2, the fault diagnostic tool was activated. The cues signaling a malfunction were the same as in condition 1 with the addition of messages for the faults found on the diagnostic tool. The Wizard of Oz method was used to simulate the behavior of the fault diagnostic tool. The research team pre-programmed the diagnostic tool’s message to be triggered on the operator’s screen during the alarm flood.

Type of faults: alarm flood scenarios

There were two different alarm flood scenarios. We reviewed all of TEP's predefined faults to identify those that led to an alarm flood and found two such faults [222]. These were “Fault 6: A feed loss” and “Fault 1: A/C feed ratio, B composition constant”.

Each participant completed a different fault for each condition to prevent any learning effect when executing the second scenario. Fault 6 was considered to be an “easy” fault due to the first alarm triggered locating the source of the fault (Low Flow Feed A). Fault 1 was considered to be a “difficult” fault because the alarms that were triggered did not clearly describe the source of the fault. To resolve Fault 6, participants had to acknowledge that there was no feed of product A, and to pause the plant until the refill was done. To resolve Fault 1 scenario, participants had to increase the flow of product A by opening its valve in manual mode.

5.4.3.2 Dependent variables

There were three dependent variables: mental workload; situational awareness; and fixation duration and count.

Mental workload questionnaire

Mental workload is a concept that has been invoked in human factors research and practice. Mental workload reflects the cost of mental resources necessary to achieve a particular level of performance during a task [223], [224]. Mental workload is viewed as the difference between the human's processing capacities that are required to perform the task and the capacity available at the given moment [58]. Sustained high mental workload causes mental fatigue, decreased performance, and can have detrimental health effects in the long run [225]. As a result, understanding subjects' mental workload under different alarm flood conditions could provide insights into the effects of the diagnostic tool.

The NASA task load index (NASA TLX) questionnaire is a tool for measuring and conducting a subjective mental workload assessment (see Appendix A). It is a well-established subjective method using a multidimensional rating scale that is the most widely used in human factors studies [226]. It assesses six dimensions related to the participant's capability and nature of the task: physical demand; mental demand; temporal demand; judgment of performance; effort required to perform the task; and level of frustration. Participants rated each dimension on a 20-point scale. Participants also completed 15 pairwise comparisons to determine the weight of each dimension. The total workload score is the weighted sum of each dimension rating, reported as percentage. In this study, the questionnaire was used to measure participants' workload after completing each scenario.

Situation Awareness Global Assessment Technique (SAGAT) questionnaire

Endsley's enduring definition of situational awareness (SA) is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [227]. Endsley's model of SA has three levels. Level 1: Perception of the elements in the environment is the first step in achieving SA. It is to perceive the status and dynamics of the relevant information in the environment. The information needs to be displayed in a way that allows successful and effective gathering of information from different sources. Level 2: Comprehension of the current situation includes being aware of the elements that are present and understanding the significance of these elements considering the operator's goals or system variables. Level 3: Projection of future status is the ability to anticipate the system dynamics in the near future. It represents the highest level of SA and is achieved through the perception of elements and the comprehension of the situation (level 1 and 2 SA). This level is important to allow operators to be proactive rather than reactive. Level 3 SA is concerned with what is going to happen, or what is most likely to happen, based on the current state and dynamics of elements [228].

In this study, we used the Situation Awareness Global Assessment Technique (SAGAT), a direct, objective assessment of participants' SA (see Appendix B). SAGAT is a widely used and validated metric that has been shown to be effective across a variety of domains to measure participants' SA [229]. The SAGAT questionnaire that was developed for this study had six questions. It was based on relevant probes from previous studies [230], [231] and covered all three levels of SA. There were four questions pertaining to SA level 1, one question on SA level 2, and one question on SA level 3.

The probes were administered at three predetermined moments: during normal operation, during the alarm flood, and after application of a corrective action. Participants were not previously aware of the timing of the probes. During each probe, the simulator was paused and displays were blanked. Participants answered the SAGAT questions on a printed questionnaire using a pen. After the participants completed the probing, the simulator was resumed from the exact moment it was stopped and the session continued. After the test, participants' answers were compared with the actual state of the simulator at the moment the probe occurred. A correct response was rated as 100% and an incorrect response as 0%. Level 1 SA is the averaged response score of all level 1 questions, averaged over all participants. The same was done for SA levels l 2 and 3, whereas the global SA was the averaged response score of all questions (levels 1, 2 and 3).

The freeze-probe approach used in the SAGAT can interrupt task flow and affect the measurement of situation awareness [232]. By requiring participants to recall system status during simulation pauses, the method may not fully reflect how SA is maintained during continuous operations. These pauses can also shift attention or change the participant's cognitive state. Nonetheless, SAGAT has been widely applied in fields such as air traffic control and medical training, where it provides a structured way to compare perceived and actual system states and has been effective in identifying gaps in operator awareness [233].

Fixation duration and count

Eye response metrics are among the most widely employed psychophysiological measures across various research domains [234]. Eye metrics include pupil diameter, blinks, gaze, and fixations. Fixations occur when eye-movements are nearly still in order to assemble the necessary information. Past research found that longer fixation duration is related to difficulty in interpreting the information presented or a greater involvement in its exploration [235]. It was found that more complex problems resulted in longer fixation duration [236]. High cognitive workload led to a failure to suppress irrelevant information, resulting in a longer mental processing time within a fixation [237]. This results in longer fixation durations and fewer fixation counts [238].

The eye-tracker used was the Pupil Invisible made by Pupil Labs. The eye tracker looks like a normal pair of glasses and measures 144 mm in width, 48mm in height, and 160 mm in length. The eye-tracker has a scene camera to record what the participant is looking at and an infra-red sensor to measure the participant's eye position. It came with an Android smartphone on which the PupilLab app is used to view real-time gaze and recordings. The recordings were uploaded onto the Pupil Lab Cloud, where we retrieved them for data analysis. The Captiv Neurolab software was used to analyze the fixation durations and counts.

5.4.4 Procedure

Two days prior to the experiment, participants were asked to watch a training video on the TEP simulator. It provided an overview of the simulator and gave examples of four fault scenarios, including the two fault scenarios that were used in the experiment. Two participants did not watch the training video before their arrival, so they watched the video before the start of the experiment. All the experimental sessions took place at a laboratory at Polytechnique Montreal University. At the participants' arrival, a briefing of the entire session was given. Subjects were seated at the computer simulator and were shown the two simulator interfaces. Then, they calibrated the eye-tracker. Participants were instructed to monitor the plant production and that if a fault occurred, they had to diagnose the fault and execute corrective actions to the best of their knowledge. At random, each participant was assigned one of the following scenarios:

1) Fault 1 with the diagnostic tool, followed by Fault 6 without the diagnostic tool.

2) Fault 1 without the diagnostic tool, followed by Fault 6 with the diagnostic tool.

3) Fault 6 with the diagnostic tool, followed by Fault 1 without the diagnostic tool.

4) Fault 6 without the diagnostic tool, followed by Fault 1 with the diagnostic tool.

In total, there were 5 participants per scenario. The order of faults and use of diagnostic tool presentation were balanced between participants. During each test, SAGAT probes were given at three predetermined moments. At the end of each test, the NASA-TLX questionnaire was administered. The eye- tracker recorded the entire session for each condition.

5.5 Data Analysis

5.5.1 NASA-TLX questionnaire

Each participant's ratings were inputted into the NASA-TLX application, and the weighted workload scores were calculated. The statistical tests ANOVA and Student's t-tests were performed to evaluate the impact of the type of fault and of condition on mental workload.

5.5.2 SAGAT questionnaire

We calculated participants' global SA, SA level 1, SA level 2, SA level 3, and SA for each of the three probes. We performed ANOVAs for each level and each probe, to analyze the impact of the diagnostic tool and type of fault on SA.

5.5.3 Eye-tracking fixation data

For each interface, we identified the areas of interest (AOI). The overview interface had four areas of interest (**Figure 5.6**): 1) key performance indicators, 2) plant overview, 3) alarm table, 4) fault diagnostic tool. The detailed unit interface had three areas of interest (**Figure 5.7**): 1) key performance indicators, 2) plant overview, 3) controls.

Figure 5.6 Overview interface AOIs.

Figure 5.7 Detailed unit interface AOIs.

In our study, the eye-tracker measured the total fixation duration and the total fixation count in each AOI. After gathering the raw fixation duration and count of each participant, we had to control for the differences in eye fixation between participants i.e., some participants had much higher fixation counts than others for the same duration of time. To do so, we calculated each participant's relative fixation duration (RFD) and relative fixation count (RFC). The RFD was calculated by dividing the fixation duration in each area by the total fixation duration on the entire interface (ex: fixation duration in AOI 4 / total fixation duration on the overview interface). The RFC was calculated by dividing the fixation count in each area by the total fixation count on the entire interface. Furthermore, RFD and RFC were calculated for each participant, for each AOI. Finally, ANOV As were performed for each AOI, to evaluate the impact of conditions and type of faults on RFD and RFC. Note that there were 15 missing recordings out of 40 due to technical problems with the eye-tracker. We report results from the 25 valid recordings from 16 participants.

5.6 Results

The following section will report statistically significant results only.

5.6.1 Mental workload

A two-way within subject ANOVA was performed to analyze the effects of the diagnostic tool and the faults on the workload. Results show that the type of fault had a significant impact on the workload ($F_{1,19} = 7.01$, $p < 0.05$), see **Figure 5.8**. The Tukey post- hoc test showed that workload during Fault 1 was significantly higher than during Fault 6 ($p < 0.05$). The diagnostic tool and the interaction between the fault and diagnostic tool showed no statistically significant effect on the workload.

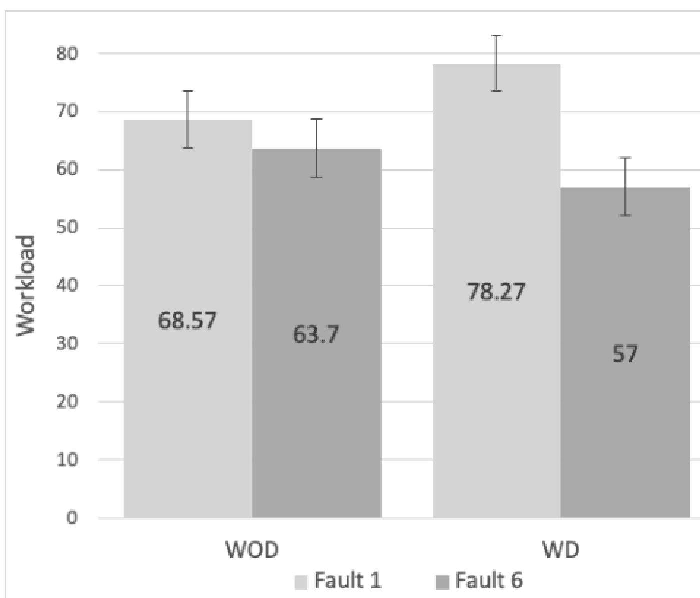

Figure 5.8 Mental workload between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.2 Situational awareness

Significant results were reported for the global SA, the SA level 3, and during probe 3 (after the corrective action). No significant results were found for SA levels 1 and 2, and for probes 1 (normal operations) and 2 (during the alarm flood).

5.6.2.1 Global SA

A two-way within subject ANOVA was performed to evaluate the impact of the diagnostic tool and the type of fault on participants' global SA. Results show that the use of diagnostic tool has a significant impact on the SA ($F_{1,19} = 6.32$, $p < 0.05$) see **Figure 5.9**. The Tukey post-hoc test showed that SA with the diagnostic tool was significantly higher than without it ($p < 0.05$). The type of fault and its interaction with the use of diagnostic tool showed no statistically significant effect on the global SA.

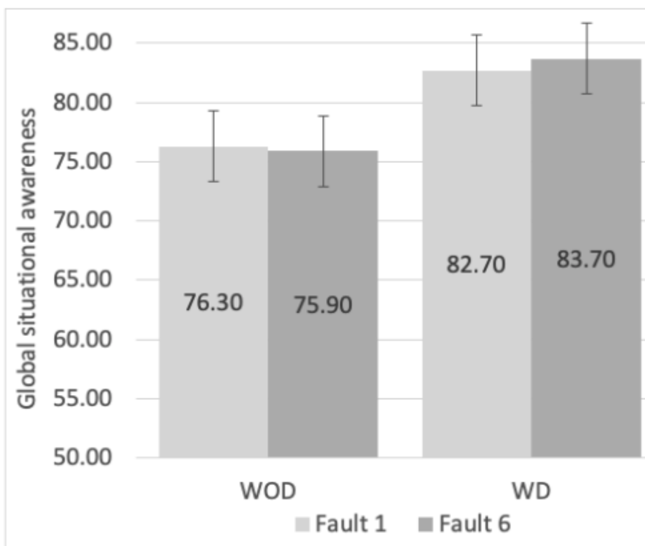

Figure 5.9 Global SA between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.2.2 SA level 3

Statistical analysis was performed to evaluate the effects of the diagnostic tool and faults on SA level 3. A two-way within subject ANOVA analysis showed that the use of diagnostic tool has a significant impact on the SA level 3 ($F_{1,19} = 4.00$, $p < 0.05$) see **Figure 5.10**. The Tukey post-hoc test showed that SA level 3 with the diagnostic tool was significantly higher than without it ($p < 0.05$).

The ANOVA also showed that the type of fault also had a significant impact on SA level 3 ($F_{1,19} = 4.00$, $p < 0.05$). Tukey post-hoc showed that SA level 3 was significantly higher during Fault 1 than Fault 6 ($p < 0.05$). The interaction between the use of diagnostic tool and the type of fault showed no statistically significant effect on SA level 3.

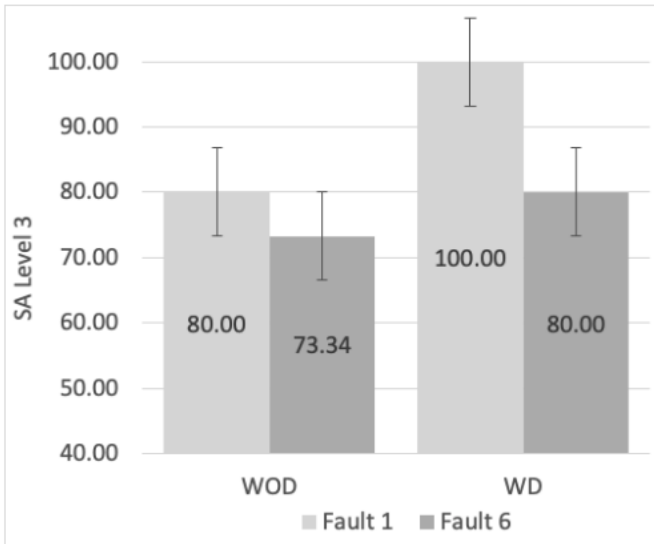

Figure 5.10 SA Level 3 between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

Student's t-tests were performed to evaluate the effect of the diagnostic tool on the SA level 3, in an easy scenario as compared to a difficult scenario. Results showed that the SA level 3 during Fault 1 (difficult scenario) was significantly higher with the diagnostic tool than without it ($p < 0.05$), see **Figure 5.11**. A t-test was also completed for Fault 6 (easy scenario) with and without the diagnostic tool, and the results showed no significant difference.

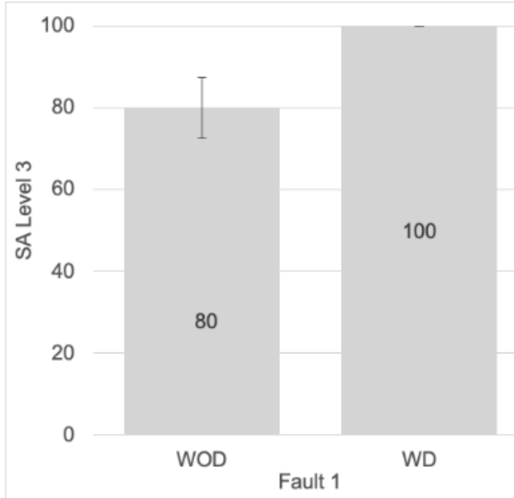

Figure 5.11 Difference in SA level 3 during Fault 1 with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.2.3 SA at probe 3

A two-way within subject ANOVA was performed to evaluate the effect of the diagnostic tool and the type of fault on the SA at probe 3, which was the moment after the resolution of the fault. Results showed no statistically significant effect. A paired t-test was performed to evaluate the effect of the diagnostic tool in an easy as compared to a difficult scenario. Results showed that the global SA at probe 3 during Fault 1 with the diagnostic tool was 73.33%, and 88.33% without it, see **Figure 5.12**. The difference was significant ($p < 0.05$). A t-test was also performed between conditions for Fault 6, but the results showed no statistically significant difference.

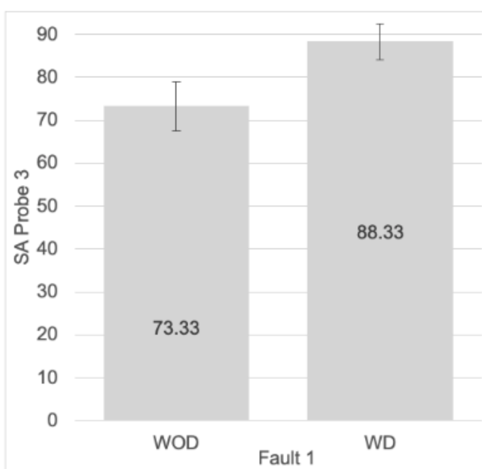

Figure 5.12 Difference in SA at probe 3 during Fault 1 between with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.3 Fixation duration and count

This paper reported the significant results that were found for the Overview interface on AOI 1, 3, and 4. Statistical tests were also completed for the Overview interface AOI 2, and all three areas of interest of the Detailed interface, and no statistically significant results were found.

5.6.3.1 Area of interest (AOI) 1: KPIs

An ANOVA analysis showed that the use of diagnostic tool had a significant impact on the RFD ($F_{1,15} = 5.30$, $p < 0.05$), see **Figure 5.13**. The interaction between the type of fault and the use of diagnostic tool also showed an effect on the RFD ($F_{1,15} = 5.03$, $p < 0.05$). The Tukey post hoc test showed that the RFD was significantly higher in AOI 1 when the diagnostic tool was activated ($p < 0.05$). The type of fault showed no significant effect on the RFD.

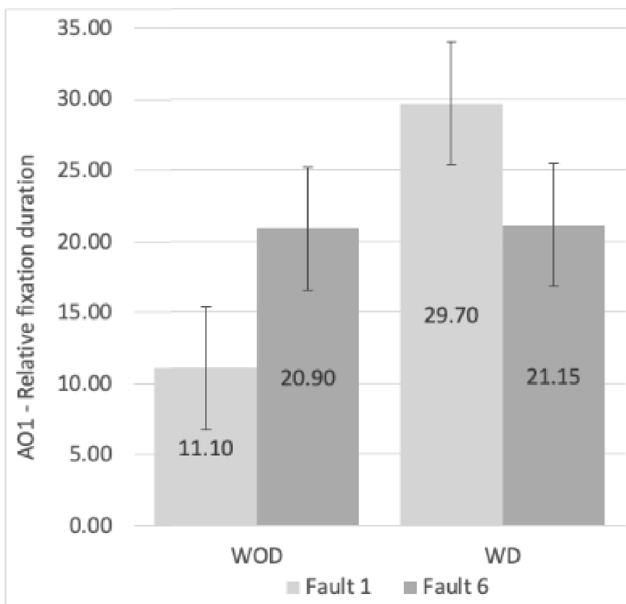

Figure 5.13 RFD in AOI1 between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

An ANOVA analysis showed that the use of the diagnostic tool had a significant impact on the RFC ($F_{1,15} = 6.21$, $p < 0.05$). The Tukey post hoc test showed that the RFC ($p < 0.05$) was significantly higher in AOI 1 when the diagnostic tool was activated (**Figure 5.14**). The type of fault showed no significant effect on the RFD.

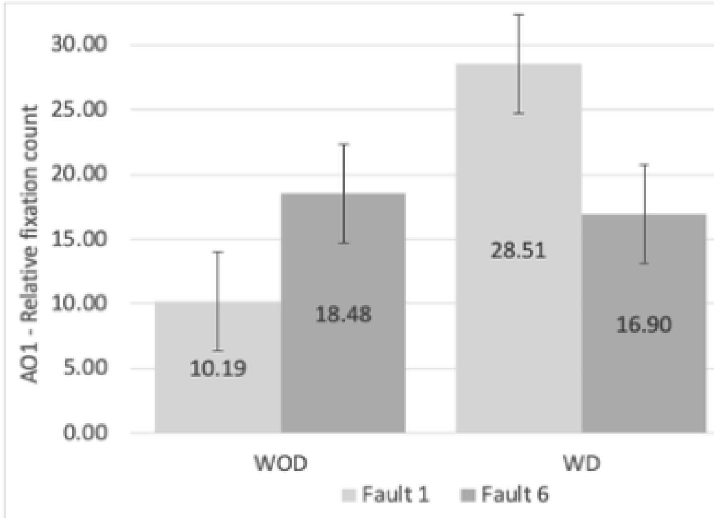

Figure 5.14 RFC in AOI1 between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.3.2 Area of interest (AOI) 3: Alarm table

A two-way ANOVA shows that the use of diagnostic tool affects the RFD on AOI 3 ($F_{1,15} = 4.59$, $p < 0.05$). Tukey post- hoc test shows that the RFD was significantly lower when the diagnostic tool was activated ($p < 0.05$), see **Figure 5.15**. The type of fault and its interaction with the use of diagnostic tool did not show any statistically significant results.

A two-way ANOVA was also performed to study the effect of the type of fault and the use of the diagnostic tool on the relative fixation count (RFC) in the AOI 3. Results showed no statistical significance.

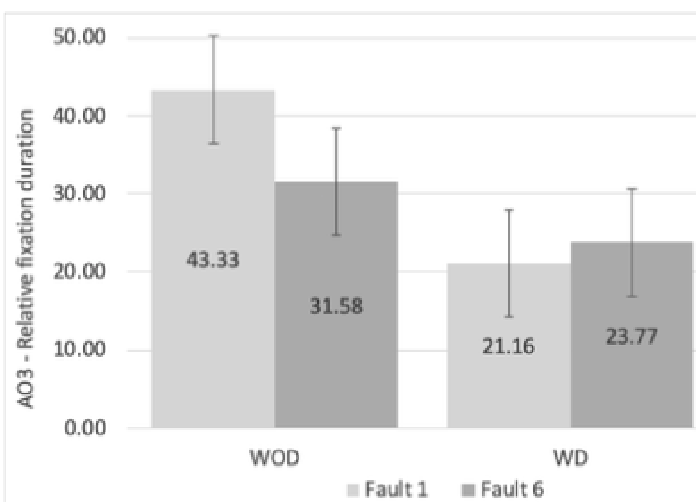

Figure 5.15 RFD in AOI3 between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.6.3.3 Area of interest (AOI) 4: Diagnostic tool

A two-way ANOVA test was performed to study the impact of the type of fault and diagnostic tool usage on the RFC on the AOI 4 (**Figure 5.16**). Results showed that the use of the diagnostic tool had an impact on the RFC ($F_{1,15} = 7.88$, $p < 0.05$). The post- hoc test showed that the RFC was significantly higher when the diagnostic tool was activated ($p < 0.05$). The type of fault and its interaction with the use of diagnostic tool did not show any statistically significant results. A two-way ANOV A was completed to evaluate the effect of the type of fault and the use of the diagnostic tool on the RFD in the AOI 4. Results showed no statistical significance.

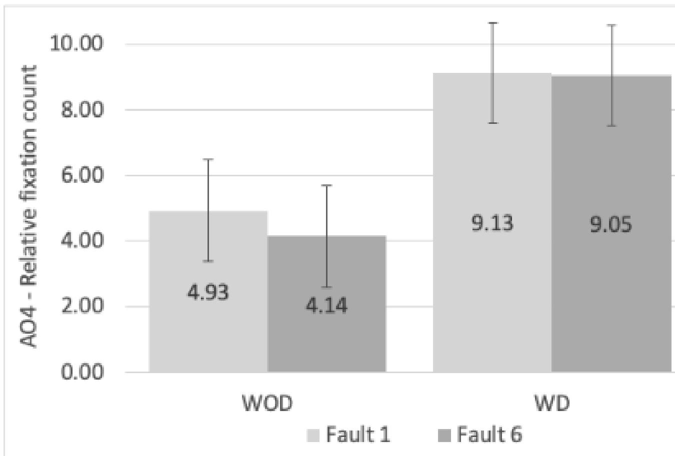

Figure 5.16 RFC in AOI4 between faults F1 and F6, with (WD) and without (WOD) the diagnostic tool. Error bars represent the standard error.

5.7 Discussion

The findings showed that the use of the diagnostic tool supported the operators during alarm floods. The global SA was high in all conditions (>80%), which suggests that PER4Mance was a good SA-oriented design for process control. PER4Mance was designed based on interfaces currently being used in the industry, to which we added a new window for the diagnostic support tool. Our results showed that having diagnostic support was beneficial and improved participants' SA without much of a change in user interface design. The use of the diagnostic tool improved global SA, SA level 3, and SA at probe 3. Having the diagnostic tool as a support increased operators' overall SA during the entire scenario, increased their ability to project future status, and increased their SA at the end of the scenario.

The use of the diagnostic tool showed benefits to operators' performance during difficult situations. Operators' workload was shown to be significantly higher during Fault 1 than during Fault 6. This confirmed the initial understanding that Fault 1 was the more difficult scenario and Fault 6 was the easier scenario. The use of the diagnostic tool increased SA level 3 during Fault 1, but not during Fault 6. In addition, the use of the diagnostic tool also increased SA at probe 3 during Fault 1 and showed no significant difference during Fault 6.

Although the findings suggested that the use of the diagnostic tool may not have helped operators during easy situations, they showed that the use of the diagnostic tool supported operators in situations that were complex with no obvious solution. Hence, the diagnostic tool showed no support to operators when the alarms clearly described the fault but showed significant benefit when the situation was complex, and the alarms were ambiguous.

The diagnostic tool led the operators' gaze to more solution- oriented areas of interest, rather than a diagnosis based on a multitude of alarms. As seen in the literature, alarms were annunciated without any differentiation between the root causes, which can overwhelm the operators. In addition, longer fixation duration was linked to high cognitive workload due to a failure to suppress irrelevant information [237]. When the diagnostic tool was activated, the fixation duration and count on AOI 1 (KPIs) increased, the fixation count on AOI 4 (diagnostic tool) increased, and the fixation duration on AOI 3 (alarm table) decreased. These findings suggested that the diagnostic tool reduced the attention from the alarm table and increased the exploration of the KPIs and the analysis of the information provided by the diagnostic tool. Taking the gaze away from the alarms and towards the KPIs and diagnostic tool is beneficial because it can reduce the risks of being overwhelmed by the alarms. This illustrated the potential of the diagnostic tool to help suppress the irrelevant information that required longer processing time such as the alarm table and rather focus on other effective information like the KPI's or the diagnostic tool.

Jang et al. (2013) introduced a proactive alarm reduction method used in a nuclear power plant environment and found that the operators' SA in the alarm reduction environment was greater than in the non-reduction environment. The findings in this paper, where participants' SA is higher in situations with the diagnostic tool activated, corroborate these results. Adhitya et al. (2014) developed a proactive system to inform chemical plant operators of an alarm before it happened, but the early warning did not improve operators' accuracy in identifying the root cause. In our study, we found that the diagnostic tool provided assistance to the operators in complex fault scenarios such as Fault 1 but did not show a significant impact in straightforward situations such as Fault 6. Therefore, our study partially confirms these results, where the system also did not support in easy scenarios yet supported in complex scenarios. This study supported previous research findings and added new findings related to the difficulty of the alarm flood scenario and the redirection of the eye gaze.

This study had three main limitations. First, the sample size was limited due to the availability of qualifying participants. A larger sample size would also have increased the internal and external validity of this study. Second, participants were not working operators from real chemical plants. The participating students in chemical engineering had limited exposure and experience in real-life chemical plants, and the simulator was a new setup for them. The third point of concern was the lack of validation from real plant operators in the simulated environment. Although the interfaces were developed based on best practices used in the industry, they hadn't been reviewed by industrial users. In addition, the behavior of the automated alarm rationalization system, i.e., the WoZ, was yet to be confirmed by a working operator.

Future directions include implementing the TEP simulator in real automation systems and improving designs based on system limitations and user feedback. To enhance the efficiency and realistic behavior of the simulated alarm rationalization system, more studies should be made to design better visualizations and interactive displays.

5.8 Conclusion

In this study, we conducted a human-in-the-loop experiment using a diagnostic support tool during two alarm flood scenarios. We found that the diagnostic support tool improved participants' situation awareness, especially in a complex scenario where the alarms triggered do not provide a clear root cause. Also, the diagnostic tool redirected participants' attention to the main operational KPIs and less on the list of alarms, encouraging a more solution-oriented approach when managing the chemical plant with the tool present. In a complex environment with human-automation interaction, issues related to trust in the machine can arise. There is a need for future studies to investigate which factors contribute to and how they impact the human's trust in an automated tool. Future work should also investigate the balance between trust in the automated tool, its transparency, and keeping the human's situational awareness when provided an erroneous diagnostic. This will be the next step of this project.

CHAPTER 6 ARTICLE 3 : THE EFFECTS OF AN IMPERFECT AI-BASED DIAGNOSTIC TOOL ON HUMAN SITUATIONAL AWARENESS, TRUST AND DECISION-MAKING DURING FAILURE MANAGEMENT

This article was submitted in the International Journal of Human-Computer Interaction on 10 February 2025.

Karine Ung (Polytechnique Montréal), Prakhar Shukla (Indian Institute of Technology Guwahati), Moncef Chioua (Polytechnique Montréal), Philippe Doyon-Poulin (Polytechnique Montréal)

6.1 Abstract

Process alarm management in complex industrial systems is challenging, especially during alarm floods, which can impair operator performance. AI shows potential in fault detection and diagnostics, supporting decision-making and situational awareness (SA). However, issues like inaccurate recommendations, insufficient transparency, and overreliance raise safety concerns. This study explored the effects of imperfect AI alarm analysis tools on operator performance. Twenty participants used an AI diagnostic tool in the Tennessee Eastman Process simulator, which provided either a correct or incorrect diagnosis. Performance measures included response times, diagnostic accuracy, SA, trust, reliance intentions, and workload. Findings showed incorrect AI diagnoses led to longer response times, lower SA, and reduced trust. Despite this, 85% of participants followed the AI's incorrect recommendations. Operators demonstrated reduced trust in the AI when its diagnostics were inaccurate but continued to rely on its recommendations. This disconnect between reported trust and actual reliance on AI highlights risks in alarm floods. Findings provide empirical evidence and recommendations to mitigate risks from imperfect AI, aiming to improve human performance and system safety in industrial environments.

6.2 Introduction

Artificial Intelligence (AI) refers to creating machines that exhibit intelligent behaviors, enabling them to perform tasks typically requiring human cognition [74]. AI systems are capable of analyzing data, recognizing patterns, and making decisions autonomously, contributing to various sectors from robotics to manufacturing industries [91]. The incorporation of AI systems in industrial process control has led to improvements in productivity and safety. According to a 2023 report by McKinsey, companies that adopt AI technologies see a 20-25% increase in operational efficiency [239]. Despite their efficiency, AI models remain challenging to interpret, as understanding how they reach conclusions is still an active area of research.

As AI became more integrated into industrial processes, it has allowed more complex tasks to be automated, and it transformed the nature of people's work [240]. However, the increasing reliance on these systems also exposes risks of AI failures, which can have severe consequences in safety-critical work settings [241]. AI can enhance situational awareness, lower out-of-the-loop issues, and improve overall performance by providing real-time insights [242]. However, during failures, operators may be left without the necessary context or control to address unexpected situations effectively [19]. This dual nature of AI highlights both its potential and the risks associated with its integration in critical industrial workplaces [243]. Our study will focus on exploring the impact of an unreliable AI-based diagnostic tool on human performance in a process control environment.

6.3 Literature Review

The literature review begins with an overview of autonomous systems, followed by a discussion on AI and automation failures in manufacturing, including challenges such as alarm floods. It then examines the implications of these issues for human performance, focusing on trust and reliance, out-of-the-loop challenges, and situational awareness. Next, the role of AI transparency is explored, culminating in an overview of AI-based fault detection and diagnosis.

6.3.1 Autonomous Systems

The integration of autonomous systems into industrial processes has advanced in recent years, reshaping the management of engineering, operations, and maintenance [244]. Gamer et al. envisioned a future in which industrial facilities function with minimal human intervention, enabled by technologies such as AI, machine learning, and robotics [245]. This vision is being realized as AI-driven systems demonstrate their capacity to optimize workflows in real time and adapt to rapidly changing conditions [246]. For instance, manufacturing and chemical industries exemplify this shift by harnessing automation to reduce downtime and enhance overall productivity [247]. These developments underscore the influence of autonomous systems on traditional operational models, setting new benchmarks for performance and reliability [248].

AI, Machine Learning (ML), and Deep Learning (DL) are terms often used interchangeably, yet they hold distinct, hierarchical relationships in the field of computer science. Machine Learning, a subset of AI, enables machines to learn and improve from data independently [81]. Deep Learning, a specialized branch within ML, uses multi-layer neural networks to recognize complex patterns in large datasets, driving advancements in areas like image and speech recognition and autonomous vehicles [249].

Despite the progress, achieving fully autonomous industrial systems remains challenging. Regulatory frameworks are still evolving, creating uncertainty for companies seeking to deploy these technologies at scale [245]. Technical hurdles, such as ensuring consistent performance in unpredictable settings, demand ongoing innovation [250]. The human element is another critical aspect requiring attention: while automation reduces repetitive tasks and enhances efficiency, it adds a supervisory role to human workers and alters their decision- making responsibilities [251]. This study aimed to explore the effects of an AI on the user, more specifically on human performance. We investigated the effects of AI failures on the operator’s trust, situational awareness and decision-making processes. The study took place in a chemical plant simulator with alarm flood scenarios. By studying these aspects, we investigated the risks posed by an unreliable AI, while fostering autonomous systems that support human resilience and adaptability in industrial environments.

6.3.2 AI/Automation Failures in Manufacturing and Alarm Floods

6.3.2.1 AI and automation

AI and automation have demonstrated effectiveness across various applications, yet it is not without flaws [252]. Failures in AI and failures in automation differ in nature and consequences. AI failures arise in data-driven systems due to biased training data or model/algorithmic flaws, resulting in incorrect or unintended outcomes [253]. For example, an AI failure in a chemical plant, where a predictive maintenance system wrongly predicts a pump failure, can lead to unnecessary shutdowns and production delays [254]. The accuracy of AI models largely depends on the quality and completeness of their training data, as AI may misclassify faults or fail to detect anomalies when encountering scenarios beyond its training experience [255].

In contrast, automation failures occur in rule-based systems due to hardware malfunctions, software bugs, or human errors [256]. For instance, a robotic arm could cease operation because of a misaligned sensor [257]. This paper focused on AI failures, specifically examining how an AI's misdiagnosis affected human performance in a chemical plant simulator. While automation systems will be referenced for context, the study emphasizes the challenges posed by AI-based systems failures in high-stakes environments.

6.3.2.2 AI/automation failures in manufacturing

Failure within automated systems, particularly in industrial plants, can lead to production delays, equipment damage, or safety risks. Studies highlighted that complex process control systems frequently exhibit unpredictable behaviors due to software or hardware issues, leading to serious failures [84], [258], [259]. For example, the Texaco Refinery Milford Haven accident in 1994 illustrated the potential dangers of automated systems when not adequately monitored. An excessive reliance on the system for managing equipment contributed to a series of failures that led to a massive explosion. The automated systems did not detect an unusual buildup of pressure, and the operators, overwhelmed by process alarms, failed to respond suitably [43]. Automation failures in manufacturing can result in dangerous outcomes, making it essential to design systems that support operators in effectively managing such situations [179].

6.3.2.3 Alarm floods

Such failures can lead to alarm floods, where cascading process alarms overwhelm operators, as seen in chemical plants where a single sensor failure triggers numerous alarms [260]. An "alarm flood" is when numerous alarms are triggered simultaneously, often resulting from cascading failures, typically more than ten alarms within a ten-minute period [261]. In many industrial settings, operators can receive hundreds or even thousands of alarms within minutes when critical system malfunctions, and they can hinder operators' ability to identify the root cause of the problem [37]. Alarm floods exacerbate system failures, as operators struggle to differentiate critical alarms from irrelevant or redundant ones, leading to delayed responses [262]. An example was the ExxonMobil Baton Rouge refinery incident in 2012, where an alarm flood overwhelmed operators, resulting in a chemical release and subsequent fire [263].

6.3.3 AI failures, implications for Human Performance

AI systems are particularly vulnerable to failures in complex, dynamic environments where conditions change rapidly [164]. When AI makes mistakes, prompt human intervention is essential to address potential failures, especially in industrial settings where undetected AI issues could have severe safety consequences [168]. As operators increasingly depend on AI, they can become disengaged from the system, impairing their ability to respond effectively when unexpected situations arise that require human take-over [167]. This underscored the necessity of establishing an appropriate level of trust in AI systems to facilitate effective human-AI collaboration and ensure operators can assume control when required [264]. To provide a deeper understanding of these challenges, we will now explore the concepts of trust and reliance, out-of-the-loop phenomena, and situational awareness.

6.3.3.1 Trust and reliance

Trust is central to effective collaboration between humans and AI systems [171], [265], [266]. Mayer et al. conceptualized trust as "a willingness to be vulnerable to the actions of another, based on the belief that the other will act in a manner beneficial to the trustor" [267]. Lee and See refined this definition by characterizing trust as an attitudinal construct and reliance as its behavioral counterpart.

Reliance refers to the act of depending on another entity to fulfill specific tasks or responsibilities [268], [269]. Trust lays the groundwork for reliance, as it shapes the individual's confidence in the system's capability and dependability [266]. Without trust, reliance is unlikely to materialize because individuals hesitate to depend on systems they find unreliable [171]. Successful interactions where systems perform as anticipated tend to bolster trust and encourage continued reliance [270].

The appropriate calibration of trust is essential for the optimal utilization of AI systems [271]. Calibration ensures that reliance corresponds appropriately to the system's actual performance capabilities, which is essential for promoting both operational safety and efficiency. Properly calibrated trust enables users to confidently delegate routine responsibilities to AI systems while maintaining active oversight in more complex or high-stakes situations [272], [273]. When there is insufficient trust, users perceive these AI systems as unreliable, they can be reluctant to rely on them, reject or underutilize systems [274]. This usually results in unnecessary manual interventions or overrides that disrupt workflows and reduce operational efficiency [275].

On the other hand, overtrust—a scenario in which trust exceeds a system's actual capabilities—can result in over-reliance and detrimental consequences [276]. Overtrust causes users to disengage from essential oversight roles, undermining the overall reliability of operations [241], [270]. For example, in aviation, excessive reliance on autopilot systems has caused pilots to overlook critical flight data, thereby increasing the risk of adverse events during crucial phases such as takeoff or landing [277]. Excessive trust can delay the detection and correction of errors, compromising system integrity and effectiveness [278].

Studies showed that individuals exhibiting higher levels of trust in AI agents are more likely to rely on their guidance [266], [274]. We aimed to explore the implications of overtrust and over-reliance, particularly focusing on alarm flood scenarios where participants risk failing to reject erroneous AI diagnosis.

6.3.3.2 Out-of-the-loop (OOTL)

The OOTL problem arises when operators disengage from monitoring tasks due to an over- reliance on AI, hindering their ability to respond effectively during critical failures [152], [279]. For instance, in industrial manufacturing, operators rely on automated quality control systems to detect defects. When these systems fail, operator's response delays can impact safety, production, and product quality [280], [281]. Studies show that OOTL operators often struggle to re-engage during system malfunctions, leading to delayed responses and increased risks of accidents [209], [282], [283].

The concept of OOTL was useful for our study as we investigated the effects of an unreliable AI when participants re-entered the loop after a period of monitoring, and tried to regain control of the fault situation.

6.3.3.3 Situational Awareness (SA)

Situational awareness is a fundamental component of effective decision-making and performance. Maintaining SA is essential to ensuring optimal human performance in complex and dynamic environments. Endsley's model divides situational awareness into three hierarchical levels: 1) Perception, 2) Comprehension, and 3) Projection. First, perception involves detecting relevant environmental information. Second, comprehension is about interpreting and understanding the significance of these elements. Third, projection is the ability to anticipate future events based on the current understanding [166]. AI enhances SA by transforming complex data into actionable insights and detecting anomalies [174], [284], [285]. In industrial settings, AI monitors variables like temperature and pressure, identifying patterns that signal potential issues [286], [287].

Studies show that when AI insights are clear and relevant, they improve SA, reduce OOTL effects, and enable faster decision-making [152], [280], [288]. AI can analyze patterns to prioritize issues and assist with complex choices [258], [289]. When AI provides an accurate assessment and relevant projections, it can significantly enhance SA, reduce OOTL issues, and improve human performance by enabling quick, informed responses [62]. However, when AI provides an inaccurate assessment or fails to justify its recommendations, it can undermine trust and create significant challenges for operators in maintaining SA. This issue becomes particularly critical in emergencies, where decisions require rapid verification and interpretation of AI-driven guidance, potentially leading to confusion, mistrust, and OOTL problems [241], [290]. For example, during an alarm flood, an AI system could prioritize a specific alarm without explaining that it detected an underlying equipment failure as the root cause. This lack of clarity can significantly impair operators' SA, leaving them uncertain about the broader context and the reliability of the AI's judgment [276]. Unclear AI logic delays actions, compromising SA, safety and performance [291], [292], [293].

Finally, AI systems can also be useful for executing decisions and action, especially for routine and repetitive tasks [294]. For instance, in industrial settings, AI can autonomously adjust machinery to maintain optimal performance or shut down equipment in case of detected malfunctions, reducing the need for constant human oversight [295]. However, during AI failures or emergencies, operators must quickly regain control of critical systems, but sudden manual intervention combined with low SA and OOTL issues can lead to confusion, delays, and compromised safety and performance [251], [262], [282].

Therefore, effective AI design should prioritize maintaining SA, calibrating trust, and minimizing OOTL issues in complex environments [296]. Enhancing AI transparency plays a critical role in keeping operators actively engaged, ensuring readiness for emergency takeovers, and supporting system effectiveness by fostering trust, reducing OOTL risks, and preserving SA in high-stakes situations [19].

6.3.3.4 AI Transparency

Transparent AI systems provide operators with clear, understandable explanations for recommendations, allowing for informed decision-making [297]. Enhancing transparency in AI systems is critical for improving SA and mitigating OOTL issues [298], particularly in manufacturing environments. For instance, AI can present real-time sensor data, such as temperature, vibration, and pressure readings, alongside specific justifications for maintenance actions [258]. In the automotive industry, transparency reveals how an AI-based system predicts maintenance needs in vehicles by disclosing the data and criteria used in autonomous driving decisions [291]. If the AI recommends shutting down a machine, it can identify which safety thresholds have been exceeded and explain the associated risks. These suggestions can be supported by summaries of past incidents, offering operators a comprehensive understanding of the system's reasoning [276].

Further, transparency can be enhanced by incorporating clear language, confidence levels, and alternative options considered by the AI system [299]. Displaying confidence levels allows operators to evaluate the reliability of AI-generated recommendations, particularly in uncertain or high-stakes scenarios [271]. Such features promote trust and empower operators to critically assess and decide whether to act on AI outputs [300]. By keeping operators engaged in the decision-making process, transparency mitigates OOTL issues and ensures a collaborative dynamic between humans and AI systems [241]. This dynamic is particularly vital in emergencies, where a rapid yet well-informed evaluation of AI recommendations is necessary [293]. Consequently, transparent AI systems support improved SA, proactive responses to emerging risks, and enhanced decision-making quality.

6.3.4 AI-Based Fault Detection and Diagnosis (FDD)

AI can be used for managing faults and mitigating alarm floods in manufacturing environments. By analyzing historical data, suppressing non-critical alarms, and prioritizing essential ones, AI systems enable operators to focus on critical issues while filtering out less significant alarms [301]. For example, in industrial plants, AI can group related alarms stemming from a single fault, allowing operators to address the root cause more effectively [302]. Additionally, AI-based systems automate routine monitoring tasks, enabling human operators to focus on complex challenges [303].

Building on these capabilities, Fault Detection and Diagnosis (FDD) can be used to support the safety, efficiency, and reliability of industrial operations [30]. It focuses on identifying system faults and analyzing their root causes using advanced techniques like artificial intelligence and machine learning [255]. AI-based FDD systems detect anomalies and diagnose issues in real-time, supporting faster corrective actions [304]. FDD can be useful in high-risk industrial environments, where timely fault resolution is essential to prevent accidents. Traditional FDD methods relied on rule-based systems, predefined thresholds, and expert judgment, which limited their adaptability and accuracy [51]. These systems were primarily reactive, identifying faults only after their occurrence and required human expertise to interpret alarms [41]. Statistical process control techniques, commonly used in manufacturing, were constrained by their inability to detect subtle anomalies in dynamic environments [305].

AI-based FDD systems detect faults by leveraging machine learning algorithms to process large amounts of sensor data, identify patterns, and detect faults in real time. These systems outperform traditional methods in both speed and accuracy [175]. For example, machine learning models have been shown to predict equipment failures early, helping manufacturers prevent downtime and improve productivity [84]. Chang et al. implemented an AI system for detecting and diagnosing operational issues in solar projects, achieving a fault detection precision of 98.6% [176]. Another study demonstrated that an AI-based system achieved a 98% agreement between predicted and actual anomalies in chemical plants over three months, improving fault detection accuracy [177]. Similarly, a deep learning model applied to the Tennessee Eastman process in chemical production achieved a fault detection accuracy of 95.6% [178].

The application of FDD systems in real-world scenarios has also been explored. For instance, one experiment evaluated AI-based alarm systems in a high-fidelity ethylene manufacturing simulator [179]. The study assessed the impact of alarm management strategies on operator workload and performance using the NASA-TLX scale. Participants included eleven console-qualified operators who engaged in training simulations incorporating three levels of alarm management (no rationalization, rationalized alarms, and smart alarms) and two levels of automation (manual and AI-based). Results indicated that smart alarm management combined with AI-based automation reduced operator workload and material losses during abnormal conditions.

Another study by Jang et al. introduced a proactive alarm reduction methodology designed for high-volume alarm environments in nuclear power plants [180]. This approach prioritized alarms to reduce cognitive overload and improve situational awareness among operators. Eight nuclear power plant operators tested the method, and results showed that situational awareness improved in environments with proactive alarm reduction compared to those without.

Additionally, the effectiveness of decision-support tools in alarm analysis has been studied. For example, Cappelli et al. investigated the Early Warning system, which predicts critical alarms within a specific time frame (e.g., 90 seconds before activation) in chemical plant control rooms [286]. Operators monitored a depropanizer unit and diagnosed faults using this system. While Early Warning enhanced operators' ability to respond proactively to potential alarms, it did not significantly improve the accuracy of diagnosing root causes.

Our previous study investigated the impact of a reliable automated FDD tool on human performance in a chemical plant simulator during an alarm flood, compared to the same scenario without any support tool [306]. The findings demonstrated that the diagnostic tool was particularly helpful when the fault was complex, with alarms that did not clearly indicate the root cause. In such scenarios, the tool significantly reduced cognitive workload, improved situational awareness, and aided operators in identifying root causes and applying corrective actions. However, for simpler alarm flood scenarios where the fault triggers a clearly identifiable root cause, the diagnostic tool showed no significant benefits. We anticipate these effects may differ when the diagnostic tool is unreliable.

Collectively, these studies document the role of AI-based FDD systems in fault detection, alarm analysis, and operator performance. This background will be particularly useful for our research, as we will be studying the effects of an AI-based FDD tool in a chemical plant simulator during alarm flood scenarios.

This study aims to investigate scenarios of imperfect process alarm analysis, focusing on situations where an AI-based diagnostic tool may misdiagnose faults. The literature shows that previous research has suggested that AI-based fault diagnostic systems are never 100% accurate and are therefore prone to failures and misdiagnoses. Process alarm analysis algorithms have often been improved using large datasets, but they have rarely been tested in high-fidelity simulators with real operators. Moreover, few studies have examined the interactions between humans and imperfect AI-based fault diagnostic tools in such realistic settings. While some human-in-the-loop experiments in simulators have been conducted, only a subset of these studies report significant findings. This research aims to bridge these gaps by evaluating the impact of unreliable AI-based diagnostic tools on operator performance and decision-making during alarm floods in a high-fidelity process control simulator.

6.4 Research Objectives

This study addresses key gaps in human-AI interaction research, focusing on imperfect alarm analysis tools and operator performance. Using a human-in-the-loop experiment in a chemical plant simulator, we examined the impact of AI failures on participants' situational awareness, trust, and decision-making during alarm flood scenarios. The experiment also evaluated operators' ability to detect AI misdiagnosis during emergencies and the subsequent effects on decision- making, particularly in diagnosing malfunctions and determining the need for intervention. By assessing the effects of imperfect AI decision-support tools, this study seeks to improve operator performance and safety in complex, high-stakes environments.

The structure of this article is as follows: the next section outlines the experimental methodology, providing details on the apparatus and diagnostic tool. This is followed by a description of the data analysis methods and a presentation of the results. The article concludes with a discussion of the key findings, contributions, and limitations of the research.

6.5 Methodology

6.5.1 Participants

The participants were 20 students enrolled in the chemical engineering program at Polytechnique Montreal. 70% self-reported as female and 30% as male. 90% were aged 20-29, and 10% were aged 30-35 years old. This experiment was reviewed and approved by Polytechnique Montreal's Ethics Committee (CER-2122-48-D). Informed consent was obtained, and each participant was given 20$ for their participation at the end of the experiment.

6.5.2 Apparatus

6.5.2.1 Chemical plant simulator

In a previous study, we developed PER4Mance, an open-source graphical user interface designed to control the Tennessee Eastman Process (TEP) chemical plant simulator [221]. The TEP is a well-known benchmark used in process control and includes a comprehensive representation of a chemical production system. TEP simulator serves as a powerful tool for understanding complex chemical processes, training operators, and refining control strategies in process industries.

It consists of five main units: a reactor, which facilitates the chemical reaction between gaseous reactants; a separator, responsible for dividing the gaseous and liquid phases; a stripper, which removes unwanted components and concentrates the desired liquid products; a compressor that increases the pressure of the gaseous outputs; and a mixer, which blends different components including the liquid by-product F. The TEP process has eight components, which include both gaseous reactants and inert gas. The primary objective is to produce liquid products G and H, which exit through the base of the stripper and are subsequently transferred to other units for further processing. In addition to these main products, a liquid by-product, F, is removed from the process to manage waste and ensure efficiency [307].

Operators using PER4Mance can manipulate 12 input variables to optimize the process conditions and control the chemical reactions. They can monitor 41 output variables, providing critical feedback on the performance of the system and allowing for real-time adjustments. The simulator is equipped with 20 pre-defined fault scenarios that simulate various operational challenges and failures, enhancing training and preparedness [185], [308]. To ensure safety and proper operation, the simulator includes a robust alarm system. When monitored variables exceed either the low or high thresholds, an alarm is triggered, which appears in red on the overview interface, accompanied by a sound alarm. This immediate feedback helps operators respond quickly to potential issues, reducing the risk of process deviations and enhancing overall safety in the chemical production environment [307].

PER4Mance featured two interfaces: The Plant Overview (**Figure 6.1**) interface presented the key performance indicators (KPIs) at the plant-level, the overall Piping and Instrumentation Diagram (P&ID), an alarm table, and the diagnostic tool; The Detailed Unit (**Figure 6.2**) interface offered in-depth information about each of the five units within the system, including time trends, and controls for the unit's valves. Both windows were displayed simultaneously on two computer screens, and alarms were triggered when variables exceeded their thresholds.

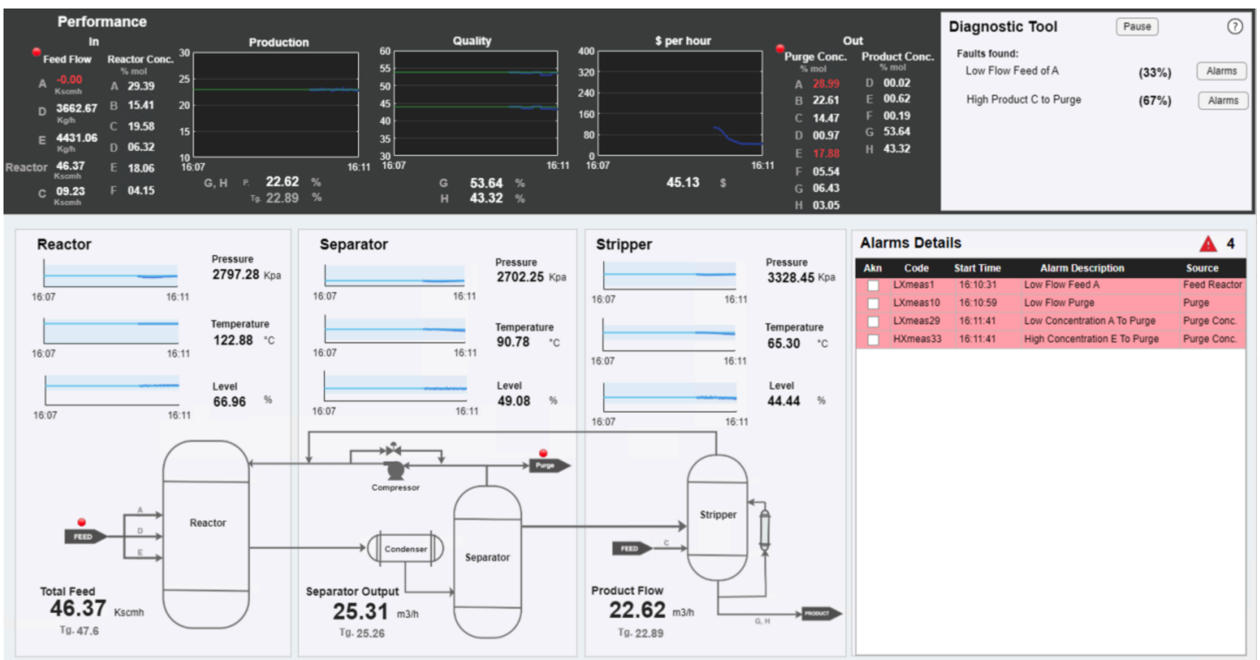

Figure 6.1 Overview interface of the chemical plant simulator.

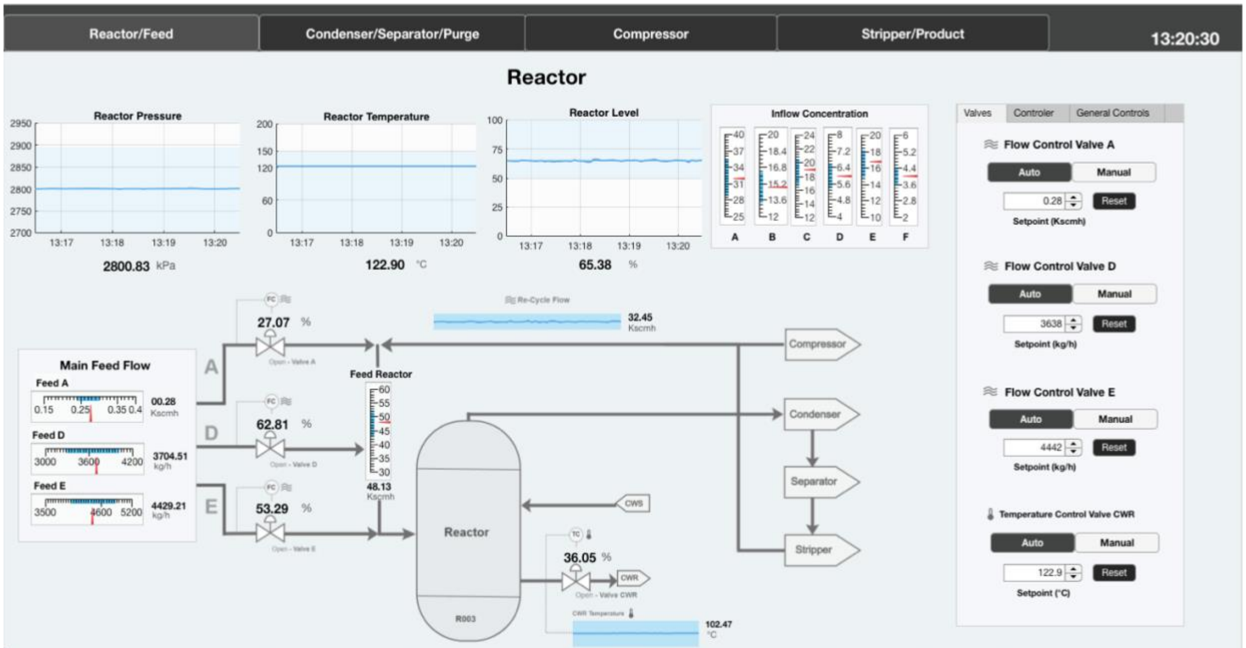

Figure 6.2 Detailed unit interface of the chemical plant simulator.

6.5.2.2 Wizard-of-Oz (WoZ) method to Simulate AI

The Wizard-of-Oz (WoZ) experimental method is a cost-effective approach used to simulate AI systems by having a human "wizard" manually operate aspects of the system while participants believe they are interacting with an autonomous AI [216]. This method has been successfully used in complex environments to evaluate how operators interact with AI-based systems. A paper presented an open-source WoZ interface designed for human-robot interaction experiments, relevant to manufacturing environments where robots and humans collaborate [309]. In a simulated urban search-and-task using WoZ to control the AI team partner, the study evaluated the human-robot team's situational awareness and performance [220]. A WoZ vehicle was developed to explore human interactions with AI-driven cars, using simulated AI systems to study how people engage with and trust AI processes [310]. Finally, a tool for conducting WoZ studies on machine learning (ML) systems was developed to simulate ML errors during user experience assessment. The study investigated the importance of considering and preventing ML errors in design [311]. In all these cases, the WoZ method helped identify challenges in human-AI interaction, such as over-reliance on the simulated AI, and provided valuable feedback for developing real AI systems. In our study, we will use the WoZ method to simulate an imperfect AI-based fault diagnosis system making recommendations to the participants for fault diagnosis.

6.5.2.3 Diagnostic Tool

The study investigated the influence of a fault diagnostic tool on operator performance through a human-in-the-loop experiment. Utilizing the Wizard of Oz (WoZ) method, researchers simulated an AI-controlled system. Participants interacted with a diagnostic tool they believed to be powered by artificial intelligence, when in fact, it was manually operated by a team member. To emulate AI behavior, the research team simulated the diagnostic tool's learning capabilities and transparency by integrating additional information. As more alarms were triggered, the tool progressively improved its ability to identify and classify the situation into increasingly likely faults, thereby updating its diagnostic message accordingly. An initial diagnostic message was generated after three alarms triggered, which was then refined following subsequent activations at six, nine, and twelve alarms. The final diagnostic message was delivered after twelve alarms.

Two alarm flood scenarios, Fault 1 and Fault 6, were selected for use in the experiment. For Fault 1, the correct diagnoses are illustrated in **Figure 6.3**, and the incorrect diagnoses in **Figure 6.4**. Fault 1 “A/C feed ratio, B composition constant", involved a step change in the feed ratio of components A and C while maintaining a constant composition of component B. This disruption altered the balance in the reactor, leading to deviations in product quality and process stability. The interconnected nature of the process made this fault challenging to detect, as its effects propagated through variables such as concentrations, temperatures, and pressures in both upstream and downstream units. To resolve Fault 1, participants had to increase product A flow by opening its valve manually.

For Fault 6, the correct diagnoses are illustrated in **Figure 6.5**, and the incorrect diagnoses in **Figure 6.6**. Fault 6 “A feed loss", represented a complete loss in the flow rate of component A, a critical reactant. This disruption directly impacted the chemical reaction dynamics, resulting in reduced product yield, altered reactor conditions, and instability throughout the system. Detecting and diagnosing this fault was complex due to its cascading effects on multiple process variables. To resolve Fault 6, participants had to acknowledge the loss of feed A verbally, or pause the plant manually.

To illustrate the inherent uncertainty in AI algorithmic confidence, the research team presented two to three potential diagnoses simultaneously, each accompanied by a confidence level expressed as a percentage. For scenarios with two or three diagnoses, the total confidence levels summed to 100%, while in cases with a single diagnosis, the confidence level was set at 90%. The diagnostic message and its associated confidence levels were updated at six, nine, and twelve alarms. The diagnosis tool also had a button to highlight the alarms used to make the diagnosis. During the failure scenarios, the total number of alarms reached 15 during Fault 1 and 27 during Fault 6.

This approach highlighted key practices in human-machine interaction, focusing on transparency and iterative learning to build operator trust in automated systems. Transparency ensures that operators understand how tools function and make decisions, enhancing their confidence in the system's reliability [262]. Additionally, iterative learning enables systems to continuously improve based on new data, which is crucial in dynamic environments like chemical facilities, where rapid changes demand precise and educated decision-making [311]. By delivering timely updates through real-time alarm data, the diagnostic tool mimics AI behaviors.

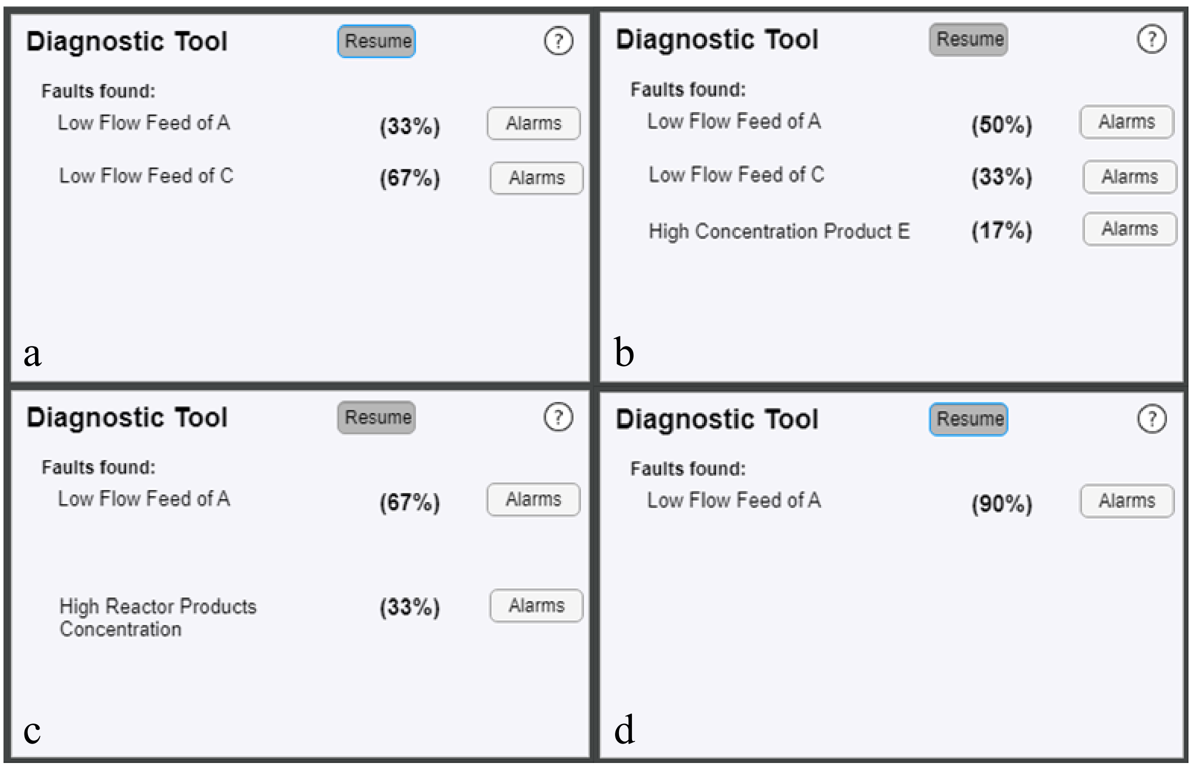

Figure 6.3 AI's evolution for Fault 1 with a correct diagnosis after: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

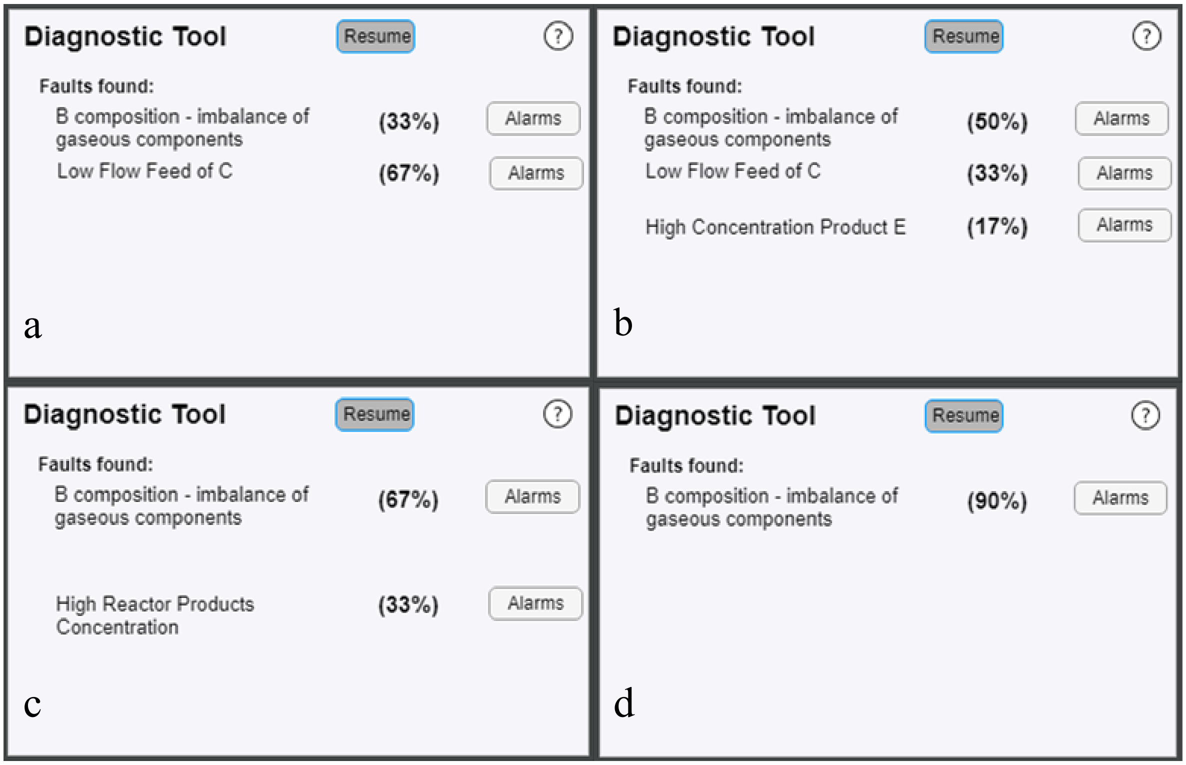

Figure 6.4 AI's evolution for Fault 1 with an incorrect diagnosis: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

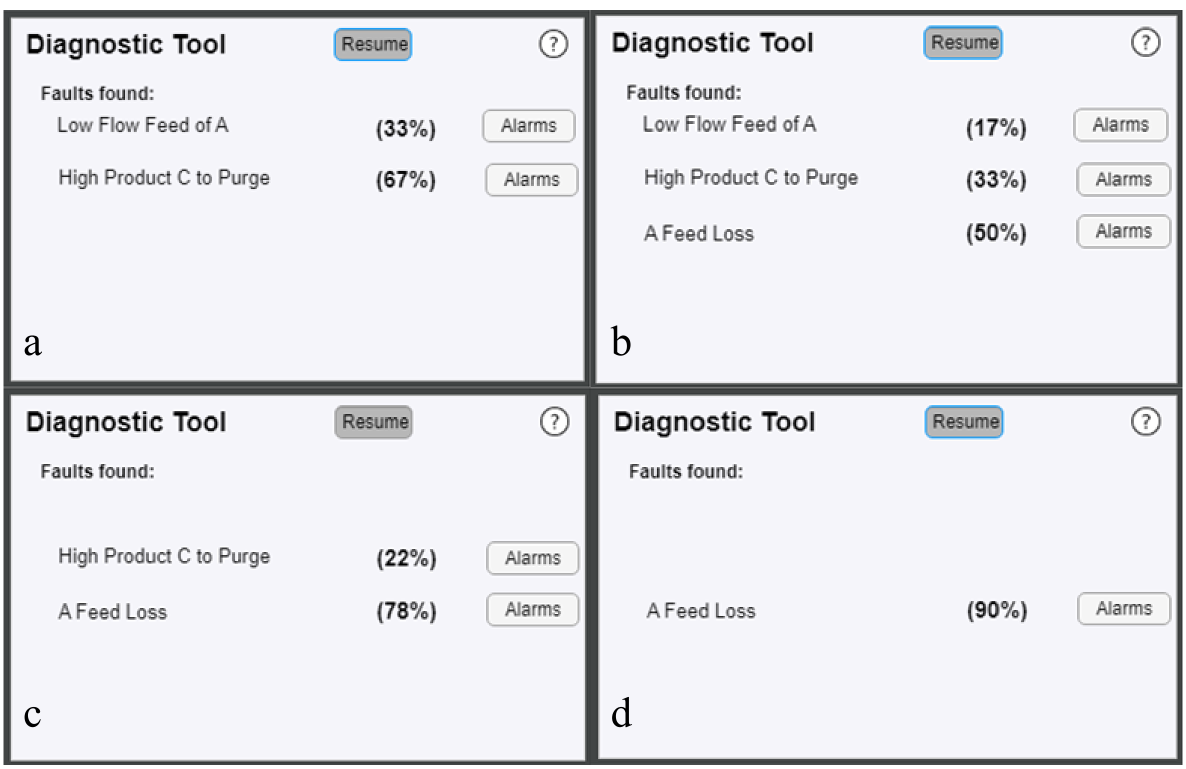

Figure 6.5 AI's evolution for Fault 6 with a correct diagnosis after: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

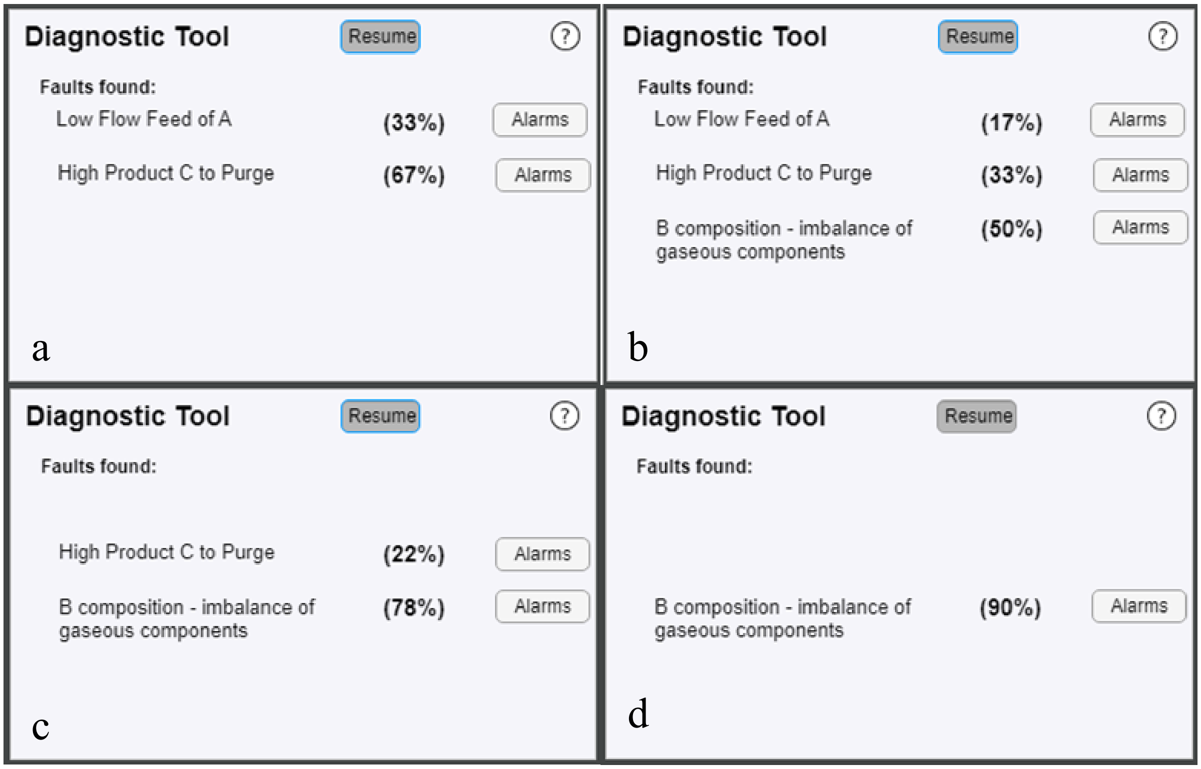

Figure 6.6 AI's evolution for Fault 6 with an incorrect diagnosis: a) 3 alarms, b) 6 alarms, c) 9 alarms, and d) 12 alarms.

6.5.3 Experimental variables

6.5.3.1 Independent variables

The study had two independent variables: the type of fault used in the scenario (Fault 1 and Fault 6) and the type of diagnosis provided by the fault diagnostic tool (correct or incorrect).

Type of fault

Fault scenarios in the TEP were predefined disturbances or malfunctions introduced into the process to test monitoring, fault detection, and diagnosis techniques. Two alarm flood scenarios were chosen as scenarios to use in the experiment: Fault 1 and Fault 6. Participants were required to complete different faults for each different condition to prevent any learning effect during the execution of the second scenario.

Fault diagnostic tool

The study tested the functionality of a fault diagnostic tool under two experimental conditions. In the first condition, the tool provided a correct diagnosis. In the second condition, the tool provided an incorrect diagnosis. In both conditions, the out-of-bound variables, alarms triggered and highlighted were the same. The only difference was the diagnostic message provided by the tool. If a participant wanted to identify the right diagnosis in the incorrect diagnosis situation, it was possible by analyzing the alarms and variables.

The diagnostic tool provided the participants with the root cause of the alarm flood but did not specify the corrective action to apply. For instance, the AI tool identified the issue as "low feed A," prompting participants to determine the appropriate response. In this case, they needed to recognize that increasing the opening of valve A would restore the feed flow. This approach was deliberately designed to prevent participants from simply applying corrective actions suggested by the tool without engaging in analysis. Participants had to be enrolled in a chemical engineering program to be qualified to participate in our study, ensuring they possessed the necessary expertise to perform these basic deductions and analyze process dynamics effectively.

6.5.3.2 Dependent variables

The study analyzed four dependent variables: response time, assessment accuracy, situational awareness and trust.

Response time

The response time consisted of the duration between the first alarm and the execution of the corrective action.

Assessment accuracy

For each test, the participant’s corrective action was recorded. Their grading was “passed” if they applied the right corrective action, and “failed” if they applied a wrong corrective action.

Situational awareness

This study employed the Situation Awareness Global Assessment Technique (SAGAT), which provides a direct and objective measure of participants' SA. This method of evaluating SA has been supported in the literature, emphasizing its effectiveness in understanding how operators maintain awareness in dynamic and complex environments [233]. The SAGAT questionnaire consisted of six questions that addressed all three SA levels and was administered at three predetermined intervals:

- Probe 1 during normal operations,
- Probe 2 during an alarm flood,
- Probe 3 following corrective actions.

Participants were unaware of when the probes would occur. During each probe, the simulator was paused, screens were blanked, and participants responded to the SAGAT questions using a printed questionnaire. After completing the questions, the simulation resumed, allowing participants to continue their tasks. The accuracy of their responses was then compared with the actual state of the simulator at the time of the probe, with correct responses scoring 100% and incorrect ones scoring 0%. Level 1 SA scores were averaged for all questions at that level, and the same was done for levels 2 and 3, while global SA was calculated as the average score across all questions.

Trust

We used a 10-item questionnaire developed to capture trust and reliance (see Appendix C). This form was adapted from Lyons & Guznov where trust is the intention to be vulnerable to another entity with little control or observability [312]. The questionnaire consisted of 4-item measuring the trust in AI versus interpersonal trust, and 6-item to measure reliance intentions. Participants answered using a 5-point likert scale (from 1-strongly disagree to 5-strongly agree).

6.5.4 Procedure

Two days before the experiment, participants watched a training video providing an overview of the TEP interfaces and examples of four fault scenarios. All participants provided a signed informed consent prior to arriving at the laboratory. The experimental sessions took place at Polytechnique Montreal, with participants briefed on the simulator and shown the interfaces. They were instructed to monitor plant production and diagnose faults and execute corrective actions. They were warned that the AI could make mistakes and provide a wrong diagnosis. Participants were assigned to one of the following scenarios:

1) Fault 1 correct diagnosis, followed by Fault 6 incorrect diagnosis.
2) Fault 1 incorrect diagnosis, followed by Fault 6 correct diagnosis.
3) Fault 6 correct diagnosis, followed by Fault 1 incorrect diagnosis.
4) Fault 6 incorrect diagnosis, followed by Fault 1 correct diagnosis.

The order of faults and use of the diagnostic tool presentation was balanced between participants. Five participants completed each scenario. SAGAT probes were given at three predetermined moments during each test, and the trust questionnaire was administered at the end of each test.

6.6 Data Analysis

Response time

The duration of Fault 1 and Fault 6 was calculated from the first to the last alarm. The faults had different duration, where Fault 1 lasted 487 seconds and Fault 6 was 912 seconds. The recorded response times were analyzed using the Student's t-test to compare the effect of the diagnostic tool. A t-test was completed for each fault because their duration was different.

Assessment accuracy

The compiled grading was analyzed using Fisher's exact test, and the odds ratio was calculated to quantify the association between the accuracy and the effect of the diagnostic tool.

SAGAT questionnaire

Participants' global SA, SA level 1, SA level 2, SA level 3, and SA for each probe were calculated, and two-way within-subject ANOVAs were performed to analyze the impact of the diagnostic tool and fault type on SA.

Trust

Following Kyons (2019), we analyzed three compound measures from the trust questionnaire: global trust score (average of questions 1 to 10), trust in AI versus interpersonal trust (average of questions 1 to 4) and reliance intentions (average of questions 5 to 10). Two-way within-subject ANOVAs were used to analyze the effects of the diagnostic tool and type of fault on participants' trust.

6.7 Results

6.7.1 Response time

A paired t-test was performed to evaluate the effect of the diagnostic tool on participants' response time during Fault 1 and Fault 6, see **Figure 6.7**. During Fault 1, results showed that the mean response time with the correct diagnostic tool was 64.3s and 141.4s with the incorrect diagnosis. The difference was statistically significant ($t_{(19)} = -4.88$, $p < 0.001$).

During Fault 6, the results showed that the response time with the correct diagnosis was 67.9s, and 173.8s with the incorrect diagnosis. The difference was significant ($t_{(19)} = -9.02$, $p < 0.0001$). For both faults, the response time was significantly slower when the diagnosis was incorrect.

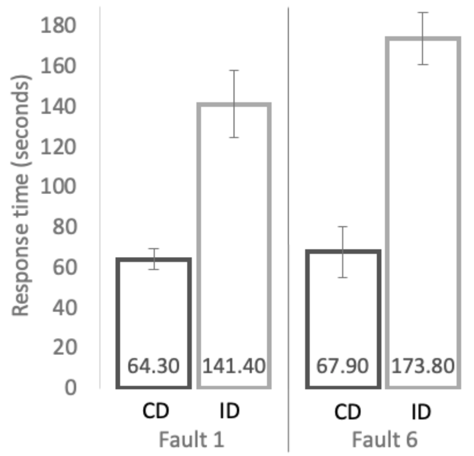

Figure 6.7 Difference in response time with correct (CD) and incorrect (ID) diagnosis during Fault 1 and Fault 6. Error bars represent the standard error.

6.7.2 Assessment accuracy

A t-test was completed to assess the impact of the assessment accuracy, results showed no significant difference.

Fisher's exact test and odds ratio were calculated to evaluate the effect of the diagnostic tool on participants' accuracy of their corrective action. Results showed a statistically significant positive association between the Pass grade and the correct diagnosis ($p < 0.05$, odds ratio = 6.93), see **Figure 6.8**. Participants were more likely to have a Pass grade when the diagnosis was correct, and a Fail grade when the diagnosis was incorrect. In the latter case, it means that participants were more likely to follow the inaccurate diagnosis presented by the tool.

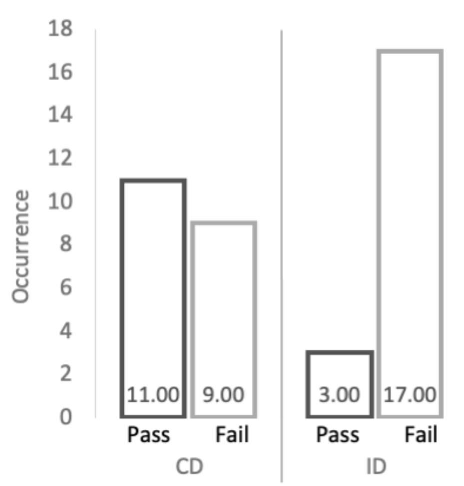

Figure 6.8 Occurrences of Pass and Fail grading with correct (CD) and incorrect (ID) diagnosis.

6.7.3 Situational awareness

6.7.3.1 Global SA

An ANOVA was performed to evaluate the impact of the diagnostic tool and the type of fault on participants' global SA. Results show that the tool's diagnosis had a significant impact on the global SA ($F_{1,19} = 28.44$, $p < 0.0001$) see **Figure 6.9**. The Tukey post-hoc test showed that the global SA with the incorrect diagnosis was significantly lower than with a correct diagnosis ($p < 0.001$). The type of fault and its interaction with the diagnostic tool showed no statistically significant effect on the global SA.

6.7.3.2 SA Level 1

Statistical analysis was performed to evaluate the effects of the diagnostic tool and type of fault on SA level 1. A two-way within-subject ANOVA analysis showed that the diagnostic tool had a significant impact on the SA level 1 ($F_{1,19} = 14.81$, $p < 0.0001$) see **Figure 6.9**. The Tukey post-hoc test showed that SA level 1 with the incorrect diagnosis was significantly lower than with a correct one ($p < 0.001$). The type of fault and its interaction with the diagnostic tool showed no statistically significant effect on the SA level 1.

6.7.3.3 SA Level 2

The ANOVA showed that the type of diagnosis had a significant impact on SA level 2 ($F_{1,19} = 13.93$, $p < 0.001$). Tukey post-hoc showed that SA level 2 was significantly lower when the diagnosis was incorrect than correct ($p < 0.001$). The interaction between the use of diagnostic tool and the type of fault showed no statistically significant effect on SA level 2.

6.7.3.4 SA Level 3

An ANOVA analysis showed that the tool's diagnosis had a significant impact on the SA level 3 ($F_{1,19} = 7.89$, $p < 0.01$). The Tukey post hoc test showed that the SA level 3 was significantly lower when the diagnosis was incorrect ($p < 0.01$). The type of fault and the interaction between the two independent variables showed no significant effect.

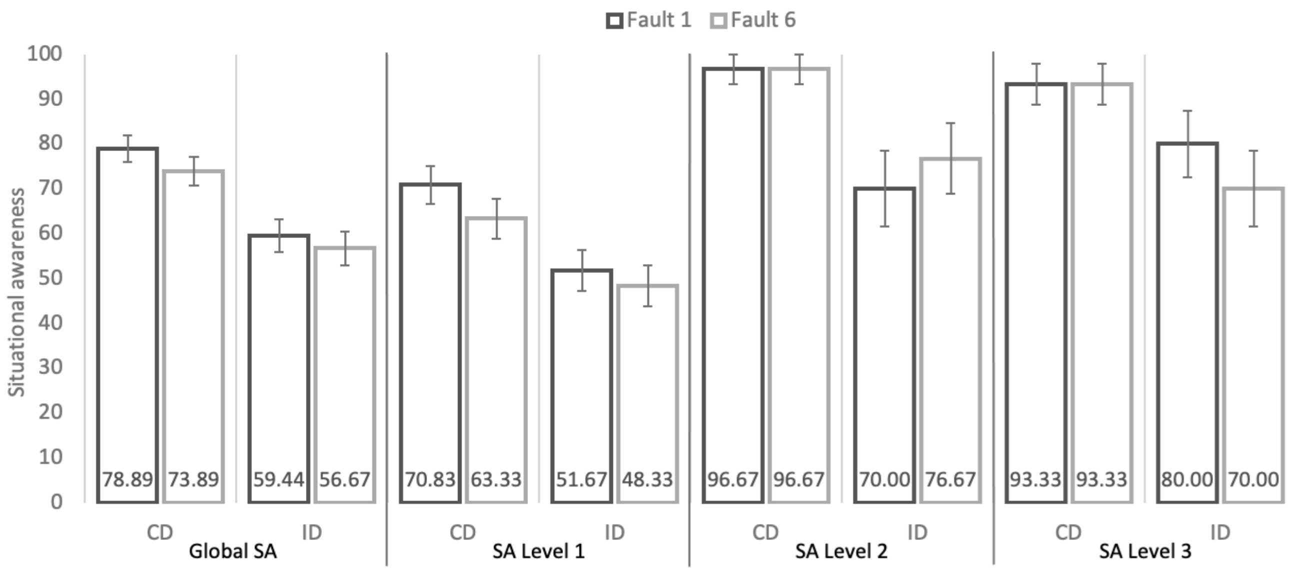

Figure 6.9 Difference in global SA, SA at levels 1, 2 and 3 during Fault 1 and Fault 6 between correct (CD) and incorrect (ID) diagnosis. Error bars represent the standard error.

6.7.3.5 SA at probe 1

An ANOVA was performed to evaluate the effect of the diagnostic tool, the type of fault, and their interaction on the SA at probe 1, which was the moment before the start of the fault. Results showed no statistically significant effect ($p > 0.05$).

6.7.3.6 SA at probe 2

An ANOVA was performed to study the impact of the type of fault and type of diagnosis on SA at probe 2, which was during the alarm flood (i.e., when 10 alarms were triggered). Results showed that the diagnostic tool had an impact on SA at probe 2 ($F_{1,19} = 7.61$, $p < 0.01$), see **Figure 6.10**. The post-hoc test showed that the SA at probe 2 was significantly lower when the diagnostic tool was incorrect ($p < 0.01$). The type of fault and its interaction with the diagnostic tool did not show any statistically significant results.

6.7.3.7 SA at probe 3

An ANOVA was performed to evaluate the effect of the diagnostic tool and the type of fault on the SA at probe 3, which was the moment after the resolution of the fault. Results showed that the diagnostic tool had a significant effect on SA at probe 3 ($F_{1,19} = 31.87$, $p < 0.0001$), see **Figure**

6.10. The post-hoc showed that the SA at probe 3 was significantly lower when the diagnosis was incorrect ($p < 0.001$). The type of fault and the interaction showed no significant difference.

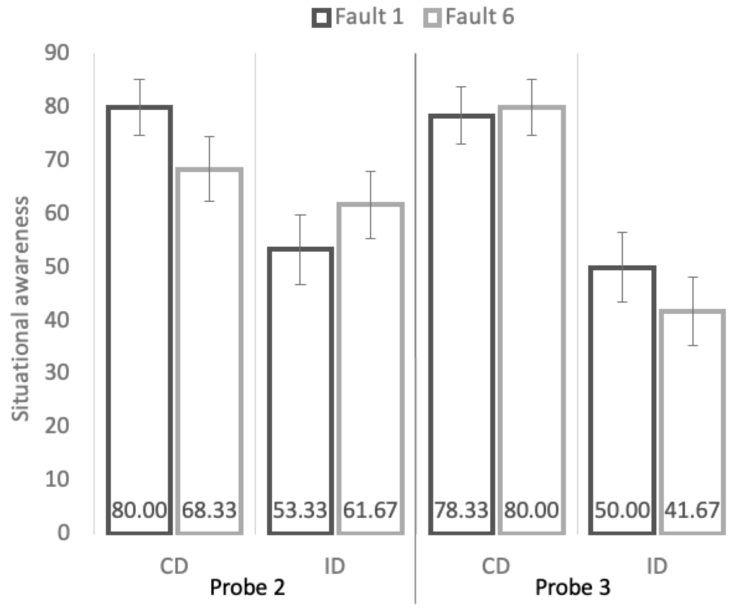

Figure 6.10 Difference in SA at probes 2 and 3 during Fault 1 and Fault 6 between correct (CD) and incorrect (ID) diagnosis. Error bars represent the standard error.

6.7.4 Trust

6.7.4.1 Global trust

An ANOVA analysis was performed to evaluate the effects of type of diagnosis and fault on the participants' global trust score. Results showed that the tool's diagnosis had a significant impact on the global trust score ($F_{1,19} = 5.52$, $p < 0.05$), see **Figure 6.11**. The Tukey post hoc test showed that the global trust was significantly lower when the diagnosis was incorrect ($p < 0.05$). The type of fault and the interaction between the independent variables showed no significant effect.

6.7.4.2 Trust in AI versus interpersonal trust

An ANOVA was performed to study the impact of the type of fault and type of diagnosis on the trust in the AI versus interpersonal trust. Results showed that the tool's diagnosis had a significant effect ($F_{1,19} = 4.47$, $p < 0.05$), see **Figure 6.11**. The post-hoc test showed that the trust in AI was significantly lower when the diagnostic tool was incorrect ($p < 0.05$). The type of fault and its interaction with the diagnostic tool did not show any statistically significant results.

6.7.4.3 Reliance intentions

Statistical analysis was performed to evaluate the effects of the diagnostic tool and type of fault on reliance intentions. An ANOVA analysis showed that the diagnostic tool had a significant impact on the reliance intentions ($F_{1,19} = 4.28$, $p < 0.05$) see **Figure 6.11**. The Tukey post-hoc test showed that reliance intentions with the incorrect diagnosis was significantly lower than with a correct one ($p < 0.05$). The type of fault and its interaction with the diagnostic tool showed no statistically significant effect.

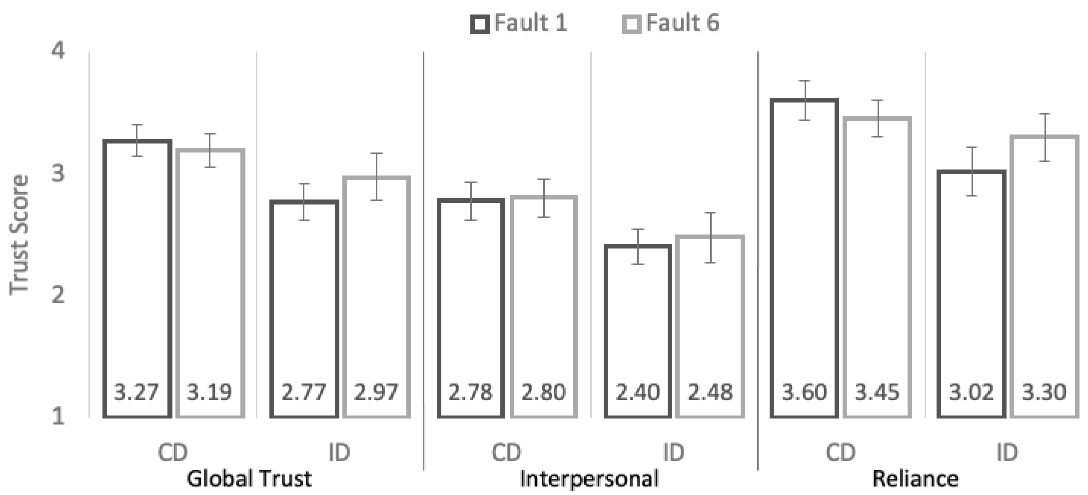

Figure 6.11 Difference in global trust, interpersonal trust and reliance during Fault 1 and Fault 6 between correct (CD) and incorrect (ID) diagnosis. Error bars represent the standard error.

6.7.5 Workload

An ANOVA was performed to evaluate the effect of the diagnostic tool, the type of fault, and their interaction on the workload, see **Figure 6.12**. Results showed no statistically significant effect ($p > 0.05$).

The mean workload with a correct diagnosis was 63.78 and the one with an incorrect diagnosis was 68.20. The literature defines that these NASA-TLX numbers in a process control environment represent high levels of workload [226].

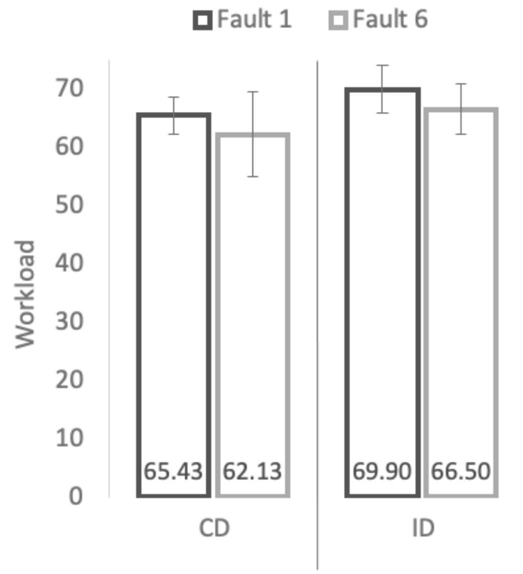

Figure 6.12 Difference in workload during Fault 1 and Fault 6 between correct (CD) and incorrect (ID) diagnosis. Error bars represent the standard error.

6.8 Discussion

6.8.1 Trust and reliance

Our results showed that response times for faults 1 and 6 were significantly slower when the AI diagnosis was incorrect, indicating participants hesitated to rely on the AI's suggestion when it was under an incorrect condition. These results supported previous research by Pearson et al. and Hoffman et al. which showed that trust in AI diminishes when reliability is in question, leading to slower decision-making and increased hesitation [274], [278].

In terms of accuracy, results from the global trust questionnaire provided further clarity on these dynamics. Global trust scores were significantly lower when the diagnosis was incorrect, supporting earlier findings by J. D. Lee & See that trust in AI erodes following system failures [171]. The compound *trust in AI versus interpersonal trust* revealed a significant decline in trust toward AI when it provided incorrect diagnosis. This aligned with previous findings by Klingbeil et al. suggesting that successful interactions where systems perform as anticipated tend to develop trust [270]. Since our participants were using the simulator for the first time and had no prior successful interactions with the AI system, they had not yet developed trust in its capabilities.

Furthermore, *reliance intentions* were significantly reduced when the diagnosis was incorrect, corroborating prior studies by Pearson et al. that found that operators were less likely to trust the

AI they perceived as unreliable [274]. However, this contrasts with participants' actual behavior, as shown in the assessment accuracy results where they mainly followed the AI's incorrect suggestion. Observations showed that participants were more likely to achieve a Pass grade when the diagnosis was correct but tended to Fail when it was incorrect. Accuracy assessment results during an AI misdiagnosis showed that 85% of participants followed the AI's incorrect suggestion. This result revealed a vulnerability to over-rely, as participants used the incorrect diagnoses. This observation aligned with findings from Buçinca et al., who noted that people frequently over-rely on the AI's suggestion even when that suggestion is wrong, as well as Leveson, who indicated that operators often follow procedures without considering context [241], [259]. This may be attributed to the fact that participants were students acting as novice operators who lacked the practical expertise to critically evaluate AI outputs.

The results from the trust and reliance questionnaire showed that participants measured low reliance intentions in the AI, but they still relied on it for diagnostic recommendations. This can be explained by an overtrust on the AI's capabilities specifically, which aligned with findings by Miró- Nicolau et al., who noted that trust that exceeds a system's actual capabilities can result in detrimental consequences [276].

In summary, while operators demonstrated hesitation and reduced trust when AI diagnoses were incorrect, they still exhibited overreliance when the AI failures were presented and their workload was high. The findings highlighted a critical disconnect between reported trust and actual reliance, particularly under high cognitive workload, as seen in alarm flood scenarios.

6.8.2 Situational Awareness and Out-of-the-Loop

Our study revealed a significant decline in SA across all levels when the diagnostic tool provided incorrect assessments. At SA Level 1 (perception), which involves the gathering and integration of system information, participants demonstrated reduced SA under conditions of incorrect diagnoses. This supported Endsley's assertion that faulty or misleading information disrupts the foundational stage of SA formation, which is critical for building a coherent understanding of system status [294]. The degradation of perception in our study also exacerbated OOTL effects. Studies by Lorenz et al. and Gouraud et al. have shown that operators often disengage cognitively during routine monitoring, making it difficult to re-engage during malfunctions [209], [282]. Our findings mirrored this phenomenon, as participants in monitoring roles showed reduced SA and delayed responses.

For SA level 2 (comprehension), which requires operators to interpret and contextualize system information, SA similarly declined when incorrect diagnostic outputs were presented. Our results suggested that reliance on an imperfect AI can lower comprehension of the situation. Additionally, results highlighted increased OOTL issues, exacerbated delays and reduced accuracy when manual intervention was required. Our NASA-TLX results demonstrated that the experiment required high cognitive effort, which increased the likelihood of failed corrective action and delayed response times. These findings corroborated earlier research, including Dixon and Wickens who identified the challenges OOTL operators encountered in re-establishing situational awareness [262]. Additionally, our high workload results were consistent with Lewis et al., who emphasized the cognitive strain involved in regaining control following periods of disengagement from automated systems [283].

At SA level 3 (projection), which involves anticipating system behavior to enable proactive decision-making, participants exhibited significantly lowered SA during incorrect diagnoses by the AI tool. This finding further validates Endsley's assertion that accurate perception and comprehension are prerequisites for predictive abilities [294].

During the second SA probe – measured during the alarm flood – participants demonstrated lower SA when diagnostic information was incorrect. At probe 3, which followed corrective actions, SA remained significantly impaired in the scenario of an incorrect diagnosis. These results further supported previous research highlighting the influence of overreliance and OOTL issues on SA degradation [19], [251].

Our findings demonstrated that incorrect AI diagnostic outputs disrupted SA across all levels and exacerbated the challenges of re-engaging during fault situations. These disruptions aligned with prior research showing that operators in OOTL states struggle to reconnect with system operations, resulting in delayed responses and compromised decision-making and performance [262]. Our study validated previous research by confirming that imperfect AI impairs SA and intensifies OOTL states and mistrust, leading to delayed recovery and reduced operator performance [298].

6.8.3 Limitations

The limitations of this study must be carefully considered when interpreting the findings. Most notably, the use of students as participants instead of professional operators raised questions about the generalizability of the results to real-world contexts. Novice operators lacked the expertise and situational familiarity required to critically evaluate AI outputs, often defaulting to the guidance provided by the AI, even when it was incorrect. These findings aligned with previous research suggesting that inexperienced operators rely heavily on rule-based decision-making due to a lack of contextual understanding, which limits their ability to adapt to changing operational conditions [207]. Furthermore, the results supported Hollnagel's observations that unexpected AI outputs can lead operators to rigidly adhere to predefined rules, even when the context demands a more flexible approach [167].

Additionally, the use of the Wizard-of-Oz (WoZ) method in this study introduced a limitation, as it relied on a simulated AI system controlled by a human operator instead of a fully autonomous system. While this approach enabled a controlled exploration of human-AI interactions in our specific context, it did not accurately replicate the adaptive behaviors, decision-making processes, or error dynamics of a real operational AI. Participants interacted with pre-determined outputs, which may not reflect how they would respond to a functional AI system. This limitation restricted the study's ability to capture the full range of operator-AI interactions present in real-world scenarios.

Conducting the study in a controlled laboratory environment introduced another limitation. The simulation lacked the dynamic, high-pressure conditions and operational complexities of real industrial control rooms. Although the controlled setting facilitated consistency in experimental conditions, it did not reflect the cognitive and multitasking demands operators face in real-world environments. Furthermore, the absence of organizational and team dynamics, which are integral to decision-making in actual control rooms, limited the ecological validity of the findings.

6.9 Conclusion

The study demonstrated that imperfect AI diagnostic outputs undermined situational awareness, intensified out-of-the-loop effects, delayed response times, and diminished trust in the AI-based systems, collectively degrading operator performance. Participants exhibited hesitation and lower trust when encountering AI failures, yet they were prone to over-rely on the system, even when it provided an incorrect diagnosis. This revealed a disconnect between operators' self-reported trust in AI and their actual reliance on it. These challenges were particularly evident during alarm floods, where heightened cognitive demands led operators to follow AI recommendations without adequately evaluating their appropriateness. Such reliance resulted in delays and poor decisions, especially in high-pressure scenarios. Novice operators were particularly affected, struggling to manage erroneous outputs, regain situational awareness, and maintain control in complex, dynamic environments.

While accurate AI outputs supported improved SA and decision-making, imperfect AI systems introduced risks that adversely affected operator performance. The persistence of OOTL effects and reductions in trust and SA observed in the study emphasized the importance of addressing these challenges. Future research should focus on the development and evaluation of strategies to optimize human-AI interaction in high-stakes environments. Enhancing AI transparency is critical to fostering trust, while advanced operator training can strengthen critical thinking skills regarding an imperfect AI. Additionally, designing adaptive interfaces that reduce cognitive overload and support the maintenance of SA is essential. These measures are necessary to ensure operators remain engaged, effectively manage AI failures, and take control when required. The PER4mance open-source simulator provides a valuable platform for conducting such studies, enabling the exploration and validation of these investigations.

CHAPTER 7 GENERAL DISCUSSION

This chapter presents a synthesis of the research objectives accomplished in the thesis, along a review of the main findings from the three studies conducted. It then presents the limitations applicable to this work and future research stemming from this thesis.

7.1 Research Objectives

Modern process industries rely on fault detection and diagnosis systems to monitor operations, identify anomalies, and support decision-making. AI-based automated diagnostic tools improved fault detection and reduced downtime, but they remained prone to misdiagnoses, false alarms, and undetected failures. In safety-critical environments, these errors led to incorrect operator decisions, increased cognitive workload, and reduced situational awareness.

A key challenge in AI-based automated diagnostic tools is their integration with alarm management systems and their effect on human performance, particularly during alarm flood episodes. While alarm analysis algorithms had been developed using large datasets, their effectiveness in real-world scenarios with human operators remained insufficiently validated. Limited research had explored how operators interacted with imperfect AI-based automated diagnostic tools in realistic operational settings. Most studies on human-AI collaboration in fault detection and diagnosis focused on algorithmic accuracy rather than human factors such as trust, reliance, and SA. A major concern was out-of-the-loop effects, where excessive reliance on AI reduced an operator's ability to respond effectively when automation failed.

This thesis addressed these gaps by examining how AI-based automated diagnostic tools influenced operator performance and decision-making during alarm flood episodes using a high-fidelity process control simulator. The research had three main objectives. The first was to develop a high-fidelity simulator designed to study fault detection, diagnosis, and alarm management in realistic industrial settings. The second was to determine whether an AI-based automated diagnostic tools improved operator performance during alarm flood episodes. The third was to examine how an AI-based diagnostic tool affected human decision-making in failure management, particularly in scenarios where the system provided incorrect or misleading diagnoses.
The thesis tested three hypotheses:

- **H_1:** The implementation of a diagnostic tool enhances operator performance during episodes of alarm flood.
- **H_2:** The impact of a diagnostic tool is significantly greater when managing difficult alarm flood episodes compared to easier ones.
- **H_3:** Operators are capable of identifying incorrect AI diagnoses during episodes of alarm flood.

The following discussion examines whether the findings supported or rejected these hypotheses.

7.2 Synthesis of main findings

7.2.1 H_1: The implementation of a diagnostic tool enhances operator performance during episodes of alarm flood

Chapter 4 focused on the development of PER4Mance, an open-source prototyping environment for alarm management research. The simulator was built using HMI design principles and followed industry standards and guidelines. It provided a testing environment where researchers could examine operator interactions with alarm systems and evaluate the efficiency of decision support tools. The simulator allowed modifications to alarm thresholds, the addition or removal of alarms, and the integration of automated diagnostic tools. These features enabled researchers to study how operators responded to different fault scenarios and assess the impact of interface design on their performance. By replicating alarm flood episodes, the simulator facilitated controlled experiments on alarm management strategies and human-machine interactions.

Although the simulator was developed based on best HMI practices and using the well-established TEP simulation, it had not yet been tested with human participants at this stage. The chapter concluded that PER4Mance provided a structured environment for examining alarm management strategies, but its effectiveness in improving operator performance remained to be evaluated. Chapter 4 concluded with stating that further experimental research was required to determine how well the interface supported operator decision-making and whether it facilitated performance improvements in practical applications.

Chapter 5 presented the first documented use of PER4Mance to evaluate operator performance in alarm flood episodes. This study introduced an automated diagnostic tool within the simulator and examined its effects on situational awareness, workload, and eye fixations. The interface integrated real-time fault diagnostic information while preserving standard process control displays. Through a controlled experiment, this research assessed the impact of the automated diagnostic tool on operator performance. Each participant completed two conditions: a control condition, where the automated diagnostic tool was deactivated, and an experimental condition, where the automated diagnostic tool was activated and provided fault diagnostic recommendations to them.

When the automated diagnostic tool was activated, fixation duration and count on AOI 1 (KPIs) and AOI 4 (diagnostic tool) increased, while fixation duration on AOI 3 (alarm table) decreased. Eye-tracking data showed that participants using the tool spent less time scanning alarms and focused more on KPIs and diagnostic recommendations, suggesting that the tool helped them prioritize relevant system information. In addition, the SAGAT assessment revealed that SA with the automated diagnostic tool was significantly higher than without it. Participants without the automated diagnostic tool relied on manually interpreting multiple alarms, which increased eye fixations on the alarm table and lowered situational awareness.

The results indicated that the automated diagnostic tool improved participants performance by increasing situational awareness and directing attention toward KPIs and diagnostic recommendations, suggesting that it helped participants prioritize relevant system information more effectively. These findings support Hypothesis 1: The implementation of a diagnostic tool enhances operator performance during episodes of alarm flood.

7.2.2 H_2: The impact of a diagnostic tool is significantly greater when managing difficult alarm flood episodes compared to easier ones

In the Chapter 5 experiment, participants experienced two different fault scenarios to avoid any learning effect during the second session. Each participant completed one condition with Fault 1, the most difficult scenario, and one with Fault 6, the easier scenario. Fault 6 was easier because the first alarms that appeared clearly indicated the source of the issue, which was the absence of feed A. Fault 1 was more complex because the alarms were triggered by secondary variables affected by the fault, making it harder to isolate the root cause. This design allowed for a comparison of operator performance with and without the use of the automated diagnostic tool while accounting for differences in fault complexity.

The workload analysis showed that participants experienced greater cognitive demand during Fault 1 than Fault 6, confirming that Fault 1 was the more difficult scenario. The results also indicated that participants had higher global SA scores when using the automated diagnostic tool. A significant increase was observed in SA level 3 (projection), which involved anticipating system behavior. The tool improved SA level 3 during Fault 1, but no significant effect was found in Fault 6. Additionally, SA at probe 3, measured at the end of the experimental session, was higher when the tool was used in Fault 1, while no significant difference was observed in Fault 6.

The automated diagnostic tool helped participants project future system status, enabling them to anticipate system dynamics in the near future. Additionally, the automated diagnostic tool improved awareness at the end of the experiment, indicating that participants retained a better understanding of system conditions throughout the task. This study indicate that the automated diagnostic tool supported their ability to interpret process conditions more effectively, particularly in Fault 1, where alarms alone did not provide sufficient diagnostic information. The study also found that the effectiveness of the automated diagnostic tool depended on the complexity of the fault scenario. In Fault 6, where alarms clearly indicated the fault, the tool had little impact on performance. However, in Fault 1, where alarms provided symptomatic alarms, the tool improved SA and workload.

The results indicated that the implementation of an automated diagnostic tool improved participant performance, especially in scenarios where alarms alone did not provide clear diagnostic information (Fault 1) and had less effect in simpler scenarios (Fault 6). The findings supported Hypothesis 2, which states that the impact of the diagnostic tool is significantly greater when managing difficult alarm flood episodes compared to easier ones.

7.2.3 H_3: Operators are capable of identifying incorrect AI diagnoses during episodes of alarm flood

Chapter 6 examined the effects of imperfect AI-based fault detection and diagnosis tools on operator performance, focusing on how incorrect AI recommendations influenced decision-making during alarm flood episodes. Literature has shown that AI-based diagnostic tools can assist operators in identifying faults and responding to system disturbances. However, these tools can also provide incorrect diagnoses, which may lead to delayed responses, reduced situational awareness, and misplaced trust in automation. The study investigated whether participants would recognize inaccurate AI recommendations and adjust their decisions or continue to rely on them despite errors.

The objective was to assess the impact of incorrect AI diagnoses on operator performance. It aimed to determine whether participants would identify AI misdiagnoses or accept inaccurate recommendations without verification. The research also examined how reliance on AI tools affected situational awareness at different levels, particularly in high-cognitive-load scenarios during episodes of alarm flood.

The study used the PER4Mance simulation to present participants with two alarm flood scenarios. An AI-based diagnostic tool provided either correct or incorrect fault identifications. Participant performance was evaluated based on response times, diagnostic accuracy, situational awareness scores, workload, trust and reliance ratings. This experimental design allowed for a direct comparison of participant behavior when interacting with accurate versus inaccurate AI-based diagnoses.

The mean workload score was 63.78 when the AI provided a correct diagnosis and 68.20 when the diagnosis was incorrect. Although a lower workload was expected when the AI provided a correct recommendation, the difference was not statistically significant. Nonetheless, these results indicate that the workload during the experiments were high for process control environments [226]. This could also reflect the high operator workload that may occur during real-life alarm flood episodes.

Response times were significantly longer when the AI provided an incorrect diagnosis. Global SA was lower with incorrect diagnoses, with significant reductions observed across all SA levels. SA Level 1, SA Level 2, and SA Level 3 were all lower when the diagnosis was incorrect. SA scores at probe 2 and probe 3 also decreased under incorrect AI recommendations.

Participants who received incorrect AI diagnoses demonstrated lower SA scores and longer response times, indicating that inaccurate AI recommendations negatively affected situational awareness and delayed corrective actions.

The accuracy assessment showed a statistically significant positive association between receiving a Pass grade and the AI providing a correct diagnosis. Participants were more likely to receive a

Pass grade when the AI diagnosis was accurate and a Fail grade when it was incorrect. This indicated that participants tended to follow the AI's recommendations regardless of their accuracy.

Global trust in the AI was significantly lower when the AI-based diagnostic tool provided incorrect information. Trust in AI and reliance intentions ratings also decreased under incorrect diagnoses. However, despite the lower reliance ratings, 85% of participants still followed the incorrect AI diagnosis, suggesting a continued dependence on automation even when reporting reduced trust in the tool.

These results showed that providing accurate AI recommendation positively impacted situational awareness, response times and accuracy, trust, and reliance. This reconfirms Hypothesis 1, that a diagnostic tool enhances operator performance during episodes of alarm flood. On the other hand, incorrect AI recommendations resulted in delayed and inaccurate responses, reduced situational awareness, and lower trust, yet participants continued to follow inaccurate AI suggestions. The study rejects Hypothesis 3 that operators are capable of identifying incorrect AI diagnoses during episodes of alarm flood.

7.3 Limitations

Developing a fully functional FDD system was beyond the scope of this study. The research team used a Wizard-of-Oz method to simulate AI-based diagnostic behavior by pre-programming the PER4Mance simulator to display diagnostic outputs at specific moments during the experiments. This ensured consistent AI responses across participants and scenarios. However, this approach introduced limitations. The system did not process data autonomously or respond to operator interactions. Outputs were static and did not adapt to changing conditions, which differs from how an autonomous AI system would function in practice. The simulation also excluded features typical of real AI tools, such as learning from new data or adjusting behavior over time. These constraints may have affected how participants interpreted and relied on the tool, particularly when it provided incorrect or missing recommendations. Future studies could incorporate interactive AI systems that respond in real time to better reflect operational settings and capture more representative human-AI interactions.

Another key limitation of the studies was the use of student participants instead of professional operators. While chemical engineering students provided relevant insights, their lack of industry experience may have influenced their performance. Additionally, the limited number of available participants and the protocol constraints imposed by COVID-19 at the time further restricted the study. The small sample size may have reduced the statistical power of the findings, limiting the ability to detect significant effects. These factors prevented the examination of team dynamics, which are an important aspect of real-world industrial operations.

The use of a laboratory-based simulation introduced several constraints. While the PER4Mance simulator provided a controlled environment for studying human-AI interaction, it did not fully replicate the complexities of real-world industrial settings. Team dynamics, time constraints, and multitasking are integral to operational environments and can influence how operators engage with AI-based systems. The absence of these factors may have affected participants' behavior, limiting the study's ecological validity, which refers to the extent to which research findings can be applied to real-world conditions. Findings from a controlled setting may not fully translate to industrial environments, where additional pressures and collaborative decision-making processes shape operator interactions with AI tools.

7.4 Theoretical implications

7.4.1 Automation failures do not always cause negative consequences

Referring back to the Jamieson and Skraaning taxonomy [138] and the findings from Chapter 5, the AI-based automated diagnostic tool used in the study exhibited an elementary automation failure, meaning it did not fully meet its intended function. The study found that the impact of the AI-based automated diagnostic tool was significantly greater in difficult alarm flood episodes compared to easier ones.

Automation failures do not always produce negative consequences, particularly when automation plays a limited role in task execution and decision-making. In operational settings with low cognitive demands, operators rely on their expertise and established procedures, reducing the impact of automation failures. In these cases, operators can detect and address issues manually, minimizing disruptions and maintaining workflow continuity.

For example, if an automation system designed for routine monitoring fails, an operator can still manually assess the system's status and make necessary adjustments. Similarly, if an automated function intended to provide supplemental information malfunctions, operators may recognize the issue and rely on existing procedures to manage the situation effectively.

In such cases, automation acts as a secondary aid rather than a primary decision-making tool. Its failure does not significantly affect workflow because operators have alternative methods to detect and resolve issues. Therefore, the impact of an automation failure is largely dependent on the level of operator reliance and the complexity of the task—when cognitive demands are low and manual processes remain effective, automation failures are less likely to cause operational disruptions.

Thus, the impact of automation failure on performance is influenced by user experience and operational context. This aligns with the observations from previous research [139], [140], that emphasized that the effects of automation failure vary depending on the operational environment and the level of user knowledge. The thesis suggests that in routine scenarios, operators can rely on their skills, making elementary automation failures less significant.

7.4.2 Ghost Failures

The automation-induced human performance challenges taxonomy defines systemic automation failure as a failure triggered by situational factors, leading to a system-wide breakdown of integrated functions. Unlike isolated technical malfunctions affecting a single function, these failures arise when automation fails to coordinate effectively across interconnected processes and disrupting operations.

Chapter 6 examined a systemic automation failure that was more difficult for participants to detect than the elementary automation failure in Chapter 5. In Chapter 5, the failure was more apparent, allowing operators to recognize the issue and adjust accordingly. In contrast, the failure in Chapter 6 was embedded within the broader interaction between automation and human decision-making, making it harder to identify. Participants often did not detect the AI failure and followed incorrect diagnostic recommendations provided by the system, suggesting that automation failures are not always apparent, particularly when the system presents information that appears reliable.

Process faults and alarm flood episodes can obscure systemic automation failures by diverting the operator's attention. In high-demand situations, operators prioritize resolving immediate process disruptions, reducing their capacity to evaluate automation performance. The cognitive and operational workload in these conditions makes it less likely that incorrect or unreliable automation outputs will be detected. When managing real-time system faults, operators may not have the cognitive resources to recognize automation failures, leading to delays or missed detections. The literature review identified masked failures, where automation malfunctions are difficult to detect due to system complexity or a lack of direct feedback. In this case, the process fault itself conceals the automation failure, as operators focus on stabilizing the system rather than verifying automation performance. This differs from other masked failures in the literature, where automation defects remain hidden due to design limitations or infrequent system interactions. Here, the masking occurs because the operator's attention is occupied by process-related issues, preventing them from assessing whether automation is functioning correctly. As a result, the automation failure persists undetected.

These findings align with automation bias, where users rely on automation without verifying its outputs. Systemic automation failures reinforce this bias, as operators may not receive clear cues that the system is providing incorrect recommendations. In some cases, process faults or alarm flood episodes camouflage automation failures, limiting the operator’s ability to detect errors in system outputs. Without direct feedback or conflicting information, operators are more likely to accept automation recommendations without independent assessment.

This relates to John Sweller’s cognitive load theory developed in 1988, which explains that under high workload conditions, cognitive resources are strained, and can negatively impact decision-making, learning, and task performance. When managing multiple tasks or responding to system faults, operators may prioritize efficiency over verification, increasing reliance on automated recommendations. In such cases, automation bias and cognitive overload interact, reducing the likelihood that automation failures will be identified. These findings emphasize the importance of automation designs that facilitate operator engagement and verification, especially in high-demand environments where cognitive capacity is strained and AI failures may go unnoticed.

7.4.3 Alarm management, or automation management?

This thesis indicates that AI-based automated diagnostic tools can influence alarm management by refining how alarms are processed, ranked by priority, and communicated to operators. Conventional alarm systems often generate a high volume of alerts, contributing to alarm fatigue, where operators become desensitized to alarms or struggle to differentiate between essential and non-essential alarms. AI-based automated diagnostic tools can support alarm management by filtering out non-relevant alarms, detecting patterns, and providing contextual information to improve decision-making. Furthermore, not all alarms require immediate attention from operators. Some alarms result from self-correcting conditions, minor variations, or low-priority events that do not require intervention. AI systems can evaluate alarm relevance in real time, differentiating between those that necessitate immediate action and those that can be recorded for later review.

During alarm flood episodes, automation can help operators manage high volumes of alarms by highlighting the most relevant information and isolating root causes. However, in stable conditions with fewer alarms, operators can rely on their expertise without additional automation support. Both of these situations raise questions about the usefulness of modern alarm systems. If automation can effectively prioritize critical alarms during high-demand situations and operators can manage simple conditions, it may be worth reconsidering the necessity of traditional alarm systems. Instead of relying on alarms as the primary means of alerting operators, alternative approaches, such as AI-driven diagnostics or predictive monitoring, could provide more effective and context-aware decision support. This challenges the conventional role of alarms and suggests that eliminating or significantly redesigning them.

Completely removing alarms could be problematic, as operators must remain aware of system conditions. Rather than eliminating alarms, AI can restructure how they are presented by grouping related notifications, delaying non-urgent signals, or summarizing less critical events. Instead of displaying separate alarms for similar issues, the system could consolidate them into a single, concise message. If a condition is likely to stabilize without intervention, the system could temporarily delay the alarm, minimizing distractions. Rather than generating an alert for every minor fluctuation, AI could produce periodic summaries that highlight trends and potential concerns, allowing operators to focus on more critical tasks. Additionally, alarm lists or tables could be placed outside the operator’s primary line of sight, ensuring that non-urgent information is accessible without interfering with immediate decision-making.

AI can refine alarm presentation, but operator oversight remains necessary to maintain control. Automated systems should be designed to allow operators to access delayed or hidden alarms when required and to override AI-based prioritization when needed. Maintaining a balance between automation and human intervention is essential to ensuring that alarm management supports decision-making and operational reliability. As AI takes on a greater role in filtering and prioritizing alarms, it raises the question of whether the focus for humans is shifting from alarm management to automation management, where operators increasingly monitor and manage AI-driven decision processes rather than the alarms themselves.

7.5 Future Research Directions

7.5.1 Adaptive automation

This thesis demonstrates that automation’s effect on human performance and decision-making is shaped by multiple factors, such as task complexity, operator expertise, system conditions, and the limitations of human information processing. The usefulness of automation depends on how well it adapts to these elements, ensuring it aids decision-making without diminishing operator involvement or awareness.

The operational context and environment play a key role in determining the level of automation support needed [313]. In demanding situations, such as alarm flood episodes, where multiple faults occur at once and the cause of failure is unclear, operators must analyze large amounts of information under time constraints. Automation can help by filtering alarms, prioritizing critical data, and providing structured decision support, reducing cognitive workload. Conversely, in simpler fault conditions where system status remains stable, excessive automation may be unnecessary. In these cases, operators can depend on their expertise without additional support, allowing them to stay engaged in monitoring and control.

An adaptive automation system should assess operator workload, task complexity, and system conditions to determine the appropriate level of assistance. Additionally, automation should adjust based on the operator’s experience level. Less experienced users may benefit from structured guidance, while experienced operators may require only targeted support to enhance efficiency rather than direct their decision-making.

Future research should focus on automation that can dynamically adjust its level of support to match changing conditions, ensuring that it enhances performance without diminishing operator engagement. Verification mechanisms should also be incorporated to encourage operators to reassess automation outputs, maintaining a balance between automation assistance and human oversight. By aligning automation with workload, task complexity, operator expertise, and system demands, adaptive automation can improve operational efficiency while supporting effective decision-making.

7.5.2 Training the human

This thesis has shown that Human-AI interactions are prone to automation bias and overreliance on AI**,** often caused by trust miscalibration—where operators either place too much or too little trust in automation [314] [315] [316]. Well-designed training programs can help mitigate these effects by improving operators' ability to assess AI reliability, recognize potential automation failures, and develop appropriate verification strategies.

Training in simulated environments with varied AI accuracy and failure scenarios can help operators calibrate trust and develop assessment skills [317]. Exposure to both correct and incorrect AI recommendations allow trainees to practice decision-making under different conditions and verify automation outputs. Scenario-based exercises requiring operators to analyze AI recommendations, identify errors, and justify decisions improve their ability to assess AI reliability.

Training programs should focus on AI awareness and understanding. Instruction on AI capabilities, limitations, how recommendations are generated, potential errors, and factors influencing accuracy can help operators assess AI outputs. This knowledge allows them to determine when further verification is needed, reducing over-reliance on automation.

Training should also include methods for cross-checking AI recommendations against independent system data [278]. Operators should apply manual verification techniques, refer to secondary data sources, and analyze historical trends before taking action. These approaches help ensure that decisions are based on validated information rather than unverified automation outputs. Training should include performance feedback to help operators understand when they trust AI too much or too little. Real-time assessments and feedback can help AI systems decide when to provide more explanations or ask for operator verification, preventing over-reliance in high-demand situations.

Future research should explore how different training methods affect operator trust calibration over time. Studies could investigate how long-term exposure to AI recommendations influences verification behaviors and decision-making patterns. Additionally, research should examine how adaptive training programs—where training intensity and content adjust based on an operator's experience and performance—impact the ability to recognize AI errors. Further research is needed to assess how training interventions affect operators' ability to develop automation judgment, mitigating automation bias and overreliance. These findings can inform training strategies that enhance AI-assisted decision-making while ensuring human oversight and control.

CHAPTER 8 CONCLUSION AND RECOMMANDATIONS

This thesis contributes to the understanding of AI integration in high-risk environments by examining its impact on human performance and decision-making. As automation becomes more common in many industries, including aviation, healthcare, manufacturing, and industrial process control, it is necessary to design systems that support rather than replace human expertise. The findings highlight how automation bias, trust miscalibration, and overreliance on AI affect fault detection, diagnosis, and response. By identifying strategies to improve Human-AI interaction, such as adaptive automation and training interventions, this research provides a foundation for developing AI systems that enhance operator performance while maintaining human oversight.

As systems become more interconnected, the number of alarms being programmed into technologies continues to grow. New aircraft include more alarms, modern trains integrate additional monitoring systems, and novel medical devices generate a higher volume of alarms. While these systems aim to improve safety and efficiency, they also contribute to alarm flood episodes. As technology advances, alarm management will become an increasing challenge, requiring solutions that help operators prioritize information, reduce cognitive workload, and improve response times.

This research provides insights that can guide policymakers, system designers, and industry leaders in shaping AI design, regulation, and implementation. Addressing alarm management and AI-assisted decision-making is necessary to ensure that automation remains a tool that supports human decision-making rather than introducing new challenges in complex operational environments.

By building on these insights, this research contributes to the development of AI-driven solutions that enhance human performance and decision-making. With thoughtful design and training, automation can be a valuable tool that improves efficiency, supports human expertise, and ensures safer and more reliable operations. As industries continue to evolve, the integration of AI and automation presents an opportunity to create systems that are more responsive, adaptive, and aligned with human capabilities.

REFERENCES

[1] S. Lai, F. Yang, T. Chen, and L. Cao, “Accelerated multiple alarm flood sequence alignment for abnormality pattern mining,” *J. Process Control*, vol. 82, pp. 44–57, 2019, doi: 10.1016/j.jprocont.2019.06.004.

[2] F. Yang, T. Chen, P. Duan, and S. L. Shah, *Capturing Connectivity and Causality in Complex Industrial Processes*, 1st ed. 2014. in SpringerBriefs in Applied Sciences and Technology. Cham: Springer International Publishing : Imprint: Springer, 2014. doi: 10.1007/978-3-319-05380-6.

[3] Engineering Equipment and Materials Users’ Association, Ed., *Alarm systems: a guide to design, management and procurement*, Edition 3. in Publication, no. 191. London: The Engineering Equipment and Materials Users’ Association, 2013.

[4] NASA Safety Center, “The case for safety: The North Sea Piper Alpha disaster,” 2013. [Online]. Available: https://sma.nasa.gov/docs/default-source/safety-messages/safetymessage-2013-05-06-piperalpha.pdf

[5] U.S. Chemical Safety and Hazard Investigation Board, “Refinery Explosion and Fire.” 2007. [Online]. Available: https://www.csb.gov/bp-america-texas-city-refinery-explosion/

[6] exida.com, “When Good Alarms Go Bad: Learning from Incidents.” exida.com L.L.C., Jan. 2015. [Online]. Available: https://www.exida.com/articles/When%20Good%20Alarms%20Go%20Bad.pdf

[7] NTSB, “Organizational Factors in Metro-North Railroad Accidents.” Nov. 19, 2014. [Online]. Available: https://www.ntsb.gov/safety/safety-studies/Documents/SIR1404.pdf#search=Search%2E%2E%2EJune%202009%20Washington%20Metro%20train%20collision

[8] N. G. Nageswari Amma and N. G. Bhuvaneswari Amma, “An Intelligent Alarm System to Detect and Control Railroad Crossings Using Wireless Sensor Networks: RailAlarm,” in *Advances in Computer and Electrical Engineering*, P. N. Mahalle, D. G. Takale, S. Sakhare, and G. B. Regulwar, Eds., IGI Global, 2024, pp. 189–212. doi: 10.4018/979-8-3693-3940-4.ch009.

[9] H. S. Savitha, V. S. Kavivendhan, K. Gagan Kumar, R. Hariprasad, and G. Keerthana, “Railway Accident Prevention System,” in *International Conference on Recent Trends in Computing & Communication Technologies (ICRCCT’2K24)*, International Journal of Advanced Trends in Engineering and Management, Nov. 2024. doi: 10.59544/UYCO7266/ICRCCT24P111.

[10] S. McDougall and J. Edworthy, “Soundscaping: Sound, meaning and vision in healthcare alarm systems,” presented at the Proceedings of the 32nd International BCS Human Computer Interaction Conference, 2018. doi: 10.14236/ewic/HCI2018.228.

[11] L. Rypicz, I. Witczak, M. Supinova, H. P. Salehi, and O. Jarabicova, “Alarm fatigue and sleep quality in healthcare workers,” *Eur. J. Public Health*, vol. 34, no. Supplement_3, p. ckae144.890, Nov. 2024, doi: 10.1093/eurpub/ckae144.890.

[12] H. R. Anderson *et al.*, “Stats on the desats: alarm fatigue and the implications for patient safety,” *BMJ Open Qual.*, vol. 12, no. 3, p. e002262, Jul. 2023, doi: 10.1136/bmjoq-2023-002262.

[13] Z. Pruitt *et al.*, “Informing Healthcare Alarm Design and Use: A Human Factors Cross-Industry Perspective,” *Patient Saf.*, pp. 6–14, Mar. 2023, doi: 10.33940/med/2023.3.1.

[14] J. P. Bliss, “Investigation of Alarm-Related Accidents and Incidents in Aviation,” *Int. J. Aviat. Psychol.*, vol. 13, no. 3, pp. 249–268, Jul. 2003, doi: 10.1207/S15327108IJAP1303_04.
[15] K. J. Ruskin, C. Corvin, S. Rice, G. Richards, S. R. Winter, and A. Clebone Ruskin, “Alarms, alerts, and warnings in air traffic control: An analysis of reports from the Aviation Safety Reporting System,” *Transp. Res. Interdiscip. Perspect.*, vol. 12, p. 100502, Dec. 2021, doi: 10.1016/j.trip.2021.100502.
[16] J. E. Veitengruber, “Design Criteria for Aircraft Warning, Caution, and Advisory Alerting Systems,” *J. Aircr.*, vol. 15, no. 9, pp. 574–581, Sep. 1978, doi: 10.2514/3.58409.
[17] Y. Chen and T. Chen, “Application of principal component pursuit to process fault detection and diagnosis,” *2013 Am. Control Conf.*, pp. 3535–3540, 2013, doi: 10.1109/ACC.2013.6580378.
[18] J. Wang, W. Hu, and T. Chen, *Intelligent Industrial Alarm Systems: Advanced Analysis and Design Methods*, 1st ed. 2024. Singapore: Springer Nature Singapore, 2024. doi: 10.1007/978-981-97-6516-4.
[19] M. R. Endsley, “Understanding Automation Failure,” *J. Cogn. Eng. Decis. Mak.*, p. 15553434231222059, Jan. 2024, doi: 10.1177/15553434231222059.
[20] B. R. Hollifield and E. Habibi, *The alarm management handbook: a comprehensive guide : practical and proven methods to optimize the performance of alarm management systems*. 2010.
[21] B. W. Bequette, *Process control: modeling, design, and simulation*. in Prentice-Hall international series in the physical and chemical engineering sciences. Upper Saddle River, N.J: Prentice Hall PTR, 2003.
[22] G. Stephanopoulos, *Chemical process control: an introduction to theory and practice*. in Prentice-Hall international series in the physical and chemical engineering sciences. Englewood Cliffs, N.J: Prentice-Hall, 1984.
[23] E. Russell, L. H. Chiang, and R. D. Braatz, *Data-driven methods for fault detection and diagnosis in chemical processes*. in Advances in industrial control. London : New York: Springer, 2000.
[24] S. R. Kondaveeti, I. Izadi, S. L. Shah, and T. Black, “Graphical Representation of Industrial Alarm Data,” *IFAC Proc. Vol.*, vol. 43, no. 13, pp. 181–186, 2010, doi: https://doi.org/10.3182/20100831-4-FR-2021.00033.
[25] I. Izadi, S. L. Shah, D. S. Shook, and T. Chen, “An Introduction to Alarm Analysis and Design,” *IFAC Proc. Vol.*, vol. 42, no. 8, pp. 645–650, 2009, doi: 10.3182/20090630-4-ES-2003.00107.
[26] J. Nachtwei, “The many faces of human operators in process control: a framework of analogies,” *Theor. Issues Ergon. Sci.*, vol. 12, no. 4, pp. 297–317, Jul. 2011, doi: 10.1080/14639221003728609.
[27] P. Marsden and M. Green, “Optimising procedures in manufacturing systems,” *Int. J. Ind. Ergon.*, vol. 17, no. 1, pp. 43–51, Jan. 1996, doi: 10.1016/0169-8141(94)00102-2.
[28] M. H. Rahaman, H. S. Alinezhad, and T. Chen, “Identification of Most Critical Alarms for Alarm Flood Reduction,” *IFAC-Pap.*, vol. 58, no. 14, pp. 835–840, 2024, doi: 10.1016/j.ifacol.2024.08.441.
[29] R. Isermann, “Process fault detection based on modeling and estimation methods—A survey,” *Automatica*, vol. 20, no. 4, pp. 387–404, Jul. 1984, doi: 10.1016/0005-1098(84)90098-0.

[30] R. Isermann and P. Ballé, “Trends in the application of model-based fault detection and diagnosis of technical processes,” *Control Eng. Pract.*, vol. 5, no. 5, pp. 709–719, May 1997, doi: 10.1016/S0967-0661(97)00053-1.

[31] K. Parsa, M. Hassall, and M. Naderpour, “Enhancing Alarm Prioritization in the Alarm Management Lifecycle,” *IEEE Access*, vol. 10, pp. 99–111, 2022, doi: 10.1109/ACCESS.2021.3137865.

[32] Center for Chemical Process Safety, “Introduction to Operating Procedures,” *American Institute of Chemical Engineers*, 120 Wall Street, FL 23 New York, NY 10005-4020, 2021. [Online]. Available: https://www.aiche.org/ccps/introduction-operating-procedures

[33] T. Kourti, “Process analysis and abnormal situation detection: from theory to practice,” *IEEE Control Syst.*, vol. 22, no. 5, pp. 10–25, Oct. 2002, doi: 10.1109/MCS.2002.1035214.

[34] F. E. Mustafa *et al.*, “A review on effective alarm management systems for industrial process control: Barriers and opportunities,” *Int. J. Crit. Infrastruct. Prot.*, vol. 41, p. 100599, Jul. 2023, doi: 10.1016/j.ijcip.2023.100599.

[35] S. Dasani, S. L. Shah, T. Chen, J. Funnell, and R. W. Pollard, “Monitoring Safety of Process Operations Using Industrial Workflows,” *IFAC-Pap.*, vol. 48, no. 8, pp. 451–456, 2015, doi: 10.1016/j.ifacol.2015.09.009.

[36] P. Goel, A. Datta, and M. S. Mannan, “Industrial alarm systems: Challenges and opportunities,” *J. Loss Prev. Process Ind.*, vol. 50, pp. 23–36, Nov. 2017, doi: 10.1016/j.jlp.2017.09.001.

[37] D. Beebe, S. Ferrer, and D. Logerot, “Alarm floods and plant incidents,” *ProSys*, 2007.

[38] ANSI/API, “Process Safety Performance Indicators for the Refining and Petrochemical Industries, First Edition.” American Petroleum Institute, 2010.

[39] A. Kabir, I. Iman, and C. Tongwen, “Similarity Analysis of Industrial Alarm Flood Data,” *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 2, pp. 452–457, Apr. 2013, doi: 10.1109/TASE.2012.2230627.

[40] B. R. Hollifield and E. Habibi, *Alarm management: seven effective methods for optimum performance*. Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2007. Accessed: Jul. 26, 2021. [Online]. Available: http://app.knovel.com/hotlink/toc/id:kpAMSEMOP1/alarm-management-seven

[41] M. L. Bransby and J. Jenkinson, *The management of alarm systems*. in Contract research report / Health & Safety Executive, no. 166. Sudbury: HSE Books, 1998.

[42] N. A. Stanton, P. Salmon, D. Jenkins, and G. Walker, *Human Factors in the Design and Evaluation of Central Control Room Operations*, 0 ed. CRC Press, 2009. doi: 10.1201/9781439809921.

[43] Great Britain and Health and Safety Executive, *The explosion and fires at the Texaco refinery, Milford Haven, 24 July 1994*. Sudbury: HSE Books, 1997.

[44] Longford Royal Commission, *The Esso Longford Gas Plant Accident Report of the Longford Royal Commission*. Government Printer for the State of Victoria, 1998. [Online]. Available: https://www.parliament.vic.gov.au/papers/govpub/VPARL1998-99No61.pdf

[45] The Joint Commission, “Medical device alarm safety in hospitals,” *Sentin. Event Alert*, no. 50, Apr. 2013, [Online]. Available: https://www.pwrnewmedia.com/2013/joint_commission/medical_alarm_safety/downloads/SEA_50_alarms.pdf

[46] M. S. Afzal, T. Chen, A. Bandehkhoda, and I. Izadi, “Performance assessment of time-deadbands,” in *2017 American Control Conference (ACC)*, Seattle, WA, USA: IEEE, May 2017, pp. 4815–4820. doi: 10.23919/ACC.2017.7963700.

[47] R. Jiang, “Optimization of alarm threshold and sequential inspection scheme,” *Reliab. Eng. Syst. Saf.*, vol. 95, no. 3, pp. 208–215, Mar. 2010, doi: 10.1016/j.ress.2009.09.012.

[48] S. Deb and D. Claudio, “Alarm fatigue and its influence on staff performance,” *IIE Trans. Healthc. Syst. Eng.*, vol. 5, no. 3, pp. 183–196, Jul. 2015, doi: 10.1080/19488300.2015.1062065.

[49] J. P. Bliss, M. J. Freeland, and J. C. Millard, “Alarm Related Incidents in Aviation: A Survey of the Aviation Safety Reporting System Database,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 43, no. 1, pp. 6–10, Sep. 1999, doi: 10.1177/154193129904300102.

[50] A. Adhitya, S. F. Cheng, Z. Lee, and R. Srinivasan, “Quantifying the effectiveness of an alarm management system through human factors studies,” *Comput. Chem. Eng.*, vol. 67, pp. 1–12, Aug. 2014, doi: 10.1016/j.compchemeng.2014.03.013.

[51] J. Errington, D. V. Reising, C. Burns, and ASM Joint R & D Consortium, *Effective alarm management practices*. Phoenix: ASMConsortium, 2009.

[52] K. Parsa, M. Hassall, and M. Naderpour, “Process Alarm Modeling Using Graph Theory: Alarm Design Review and Rationalization,” *IEEE Syst. J.*, vol. 15, no. 2, pp. 2257–2268, Jun. 2021, doi: 10.1109/JSYST.2020.3019041.

[53] Y. T. P. Nunes, R. Souza de Abreu, G. Leitão, and L. A. Guedes, “An Optimization-based Method to Define Dead-band Values of Industrial Alarms,” 2022, doi: 10.20906/CBA2022/3453.

[54] Z. Geng and Q. Zhu, “Multi-swarm PSO and its application in operational optimization of ethylene cracking furnace,” in *2008 7th World Congress on Intelligent Control and Automation*, Chongqing, China: IEEE, 2008, pp. 103–106. doi: 10.1109/WCICA.2008.4592907.

[55] N. Sanchez-Pi, A. C. B. Garcia, and L. A. P. Leme, “Intelligent agents for alarm management in petroleum ambient,” *J. Intell. Fuzzy Syst.*, vol. 28, no. 1, pp. 43–53, 2015, doi: 10.3233/IFS-141198.

[56] Y. Laumonier, J.-M. Faure, J.-J. Lesage, and H. Sabot, “Towards alarm flood reduction,” in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Limassol: IEEE, Sep. 2017, pp. 1–6. doi: 10.1109/ETFA.2017.8247625.

[57] F. O. Onyekwe, O. Odujobi, F. E. Adikwu, and T. Y. Elete, “Revolutionizing process alarm management in refinery operations: Strategies for reducing operational risks and improving system reliability,” *Magna Sci. Adv. Res. Rev.*, vol. 9, no. 2, pp. 187–149, 2023, doi: https://doi.org/10.30574/msarr.2023.9.2.0156.

[58] D. Dunn and N. P. Sands, “Alarm Management Fundamentals – What is the Impact of Electrical Equipment Automation on Industrial Control Systems,” *IEEE IAS Pet. Chem. Ind. Tech. Conf. PCIC*, pp. 1–10, 2024, doi: 10.1109/pcic47799.2024.10832189.

[59] A. Abulaban, S. Imtiaz, and S. Ahmed, “AlarmSoft: An Advanced Cloud-based Alarm Management Application,” *2022 IEEE Int. Symp. Adv. Control Ind. Process. AdCONIP*, doi: 10.1109/AdCONIP55568.2022.9894248.

[60] J. Folmer and B. Vogel-Heuser, “Computing dependent industrial alarms for alarm flood reduction,” *Int. Multi-Conf. Syst. Signals Devices*, pp. 1–6, doi: https://doi.org/10.1109/SSD.2012.6198008.

[61] C. Tian, P. Song, C. Zhao, and J. Ding, “Structure Feature Extraction for Hierarchical Alarm Flood Classification and Alarm Prediction,” *IEEE Trans. Autom. Sci. Eng.*, pp. 1–11, 2024, doi: 10.1109/TASE.2023.3290256.

[62] G. Dorgo, A. Palazoglua, and J. Abonyi, “Decision trees for informative process alarm definition and alarm-based fault classification,” *Process Saf. Environ. Prot.*, vol. 149, pp. 312–324, 2021, doi: https://doi.org/10.1016/j.psep.2020.10.024.

[63] H. Seyed Alinezhad, J. Shang, and T. Chen, “Early Classification of Industrial Alarm Floods Based on Semisupervised Learning,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 3, pp. 1845–1853, Mar. 2022, doi: 10.1109/TII.2021.3081417.

[64] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Advances in alarm data analysis with a practical application to online alarm flood classification,” *J. Process Control*, vol. 79, pp. 56–71, Jul. 2019, doi: 10.1016/j.jprocont.2019.04.010.

[65] N. Javanbakht, A. Neshastegaran, and I. Izadi, “Alarm-Based Root Cause Analysis in Industrial Processes Using Deep Learning,” Mar. 21, 2022, *arXiv*: arXiv:2203.11321. doi: 10.48550/arXiv.2203.11321.

[66] M. Fullen, P. Schüller, and O. Niggemann, Eds., *Semi-supervised Case-based Reasoning Approach to Alarm Flood Analysis*. in Technologien für die intelligente Automation, Technologies for Intelligent Automation, no. 11. Berlin, Heidelberg: Springer Vieweg, 2020. doi: 10.1007/978-3-662-59084-3.

[67] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “On improving fault detection and diagnosis using alarm-range normalisation,” *IFAC-Pap.*, vol. 51, no. 24, pp. 1227–1232, 2018, doi: 10.1016/j.ifacol.2018.09.695.

[68] E. L. Russell, *Data-driven methods for fault detection and diagnosis in chemical processes.* Place of publication not identified: Springer, 2012.

[69] K. Severson, P. Chaiwatanodom, and R. D. Braatz, “Perspectives on process monitoring of industrial systems,” *Annu. Rev. Control*, vol. 42, pp. 190–200, 2016, doi: 10.1016/j.arcontrol.2016.09.001.

[70] I. Akyar, “Standard Operating Procedures (What Are They Good For ?),” in *Latest Research into Quality Control*, I. Akyar, Ed., InTech, 2012. doi: 10.5772/50439.

[71] T. Qing, Z. Liu, Y. Tang, H. Hong, L. Zhang, and S. Chen, “Effects of Automation for Emergency Operating Procedures on Human Performance in a Nuclear Power Plant,” *Health Phys.*, vol. 121, no. 3, pp. 261–270, 2021, doi: 10.1097/HP.0000000000001445.

[72] A. Shirshahi and M. Aliyari-Shoorehdeli, “Diagnosing root causes of faults based on alarm flood classification using transfer entropy and multi-sensor fusion approaches,” *Process Saf. Environ. Prot.*, vol. 181, pp. 469–479, Jan. 2024, doi: 10.1016/j.psep.2023.11.036.

[73] G. Manca, F. C. Kunze, and A. Fay, “A Novel Process Plant Alarm Dataset and Methodology for Alarm Data Generation,” in *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, Padova, Italy: IEEE, Sep. 2024, pp. 1–8. doi: 10.1109/ETFA61755.2024.10710862.

[74] W. Ertel, *Introduction to Artificial Intelligence*. in Undergraduate Topics in Computer Science. Wiesbaden: Springer Fachmedien Wiesbaden, 2025. doi: 10.1007/978-3-658-43102-0.

[75] I. Ahmed, G. Jeon, and F. Piccialli, “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022, doi: 10.1109/TII.2022.3146552.

[76] T. Yang, X. Yi, S. Lu, K. H. Johansson, and T. Chai, “Intelligent Manufacturing for the Process Industry Driven by Industrial Artificial Intelligence,” *Engineering*, vol. 7, no. 9, pp. 1224–1230, Sep. 2021, doi: 10.1016/j.eng.2021.04.023.

[77] K. Bauer, M. Von Zahn, and O. Hinz, "Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing," *Inf. Syst. Res.*, vol. 34, no. 4, pp. 1582–1602, Dec. 2023, doi: 10.1287/isre.2023.1199.

[78] O. B. Akinnagbe, "Human-AI Collaboration: Enhancing Productivity and Decision-Making," *Int. J. Educ. Manag. Technol.*, vol. 2, no. 3, pp. 387–417, Nov. 2024, doi: 10.58578/ijemt.v2i3.4209.

[79] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook," *IEEE Access*, vol. 8, pp. 220121–220139, 2020, doi: 10.1109/ACCESS.2020.3042874.

[80] Mr. P. Durgaprasad, "Effect of AI on HI," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 7, pp. 976–978, Jul. 2024, doi: 10.22214/ijraset.2024.63682.

[81] S. Rafatirad, H. Homayoun, Z. Chen, and S. M. Pudukotai Dinakarrao, "What Is Applied Machine Learning?," in *Machine Learning for Computer Scientists and Data Analysts*, Cham: Springer International Publishing, 2022, pp. 3–33. doi: 10.1007/978-3-030-96756-7_1.

[82] J. Song, "A Review of The Application of Natural Language Processing in Human-Computer Interaction," *Appl. Comput. Eng.*, vol. 106, no. 1, pp. 111–117, Nov. 2024, doi: 10.54254/2755-2721/106/20241328.

[83] D. Shamoo, "Computer Vision Promising Innovations," *World J. Adv. Res. Rev.*, vol. 23, no. 3, pp. 610–619, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2725.

[84] M. Shahin, F. F. Chen, A. Hosseinzadeh, and N. Zand, "Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: an early failure detection diagnostic service," *Int. J. Adv. Manuf. Technol.*, vol. 128, no. 9–10, pp. 3857–3883, Oct. 2023, doi: 10.1007/s00170-023-12020-w.

[85] V. S. Gurav, A. Gugnani, Y. R. Meena, V. Marathe, S. A. A. Vijay, and S. Nanda, "The Impact of Industrial Automation on the Manufacturing Industry in the Era of Industry 4.0," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India: IEEE, Jun. 2024, pp. 1–6. doi: 10.1109/ICCCNT61001.2024.10725011.

[86] D. Patil, "Human-Artificial Intelligence Collaboration In The Modern Workplace: Maximizing Productivity And Transforming Job Roles," 2025. doi: 10.2139/ssrn.5057414.

[87] D. A. Spencer, "AI, automation and the lightening of work," *AI Soc.*, May 2024, doi: 10.1007/s00146-024-01959-3.

[88] F. Sultana, U. Talpur, M. S. Iqbal, A. Ahmed, and K. Hussain, "The Macroeconomic Implications of Automation and AI on Labor Markets and Employment," *Crit. Rev. Soc. Sci. Stud.*, vol. 2, no. 2, pp. 497–507, Nov. 2024, doi: 10.59075/f9hdkk61.

[89] L. Qudus, "Leveraging Artificial Intelligence to Enhance Process Control and Improve Efficiency in Manufacturing Industries," *Int. J. Comput. Appl. Technol. Res.*, Jan. 2025, doi: 10.7753/IJCATR1402.1002.

[90] B. P. Agrawal *et al.*, "AI-Driven Robotics for Real-Time Manufacturing Processes:," in *Advances in Computational Intelligence and Robotics*, D. Pandey, B. Muniandi, B. K. Pandey, and A. S. George, Eds., IGI Global, 2024, pp. 199–212. doi: 10.4018/979-8-3693-7367-5.ch014.

[91] National Association of Manufacturers, "The future of AI in manufacturing," 2023. [Online]. Available: https://www.nam.org

[92] T. Y. Elete, Emmanuella Onyinye Nwulu, Ovie Vincent Erhueh, Oluwaseyi Ayotunde Akano, and Adeoye Taofik Aderamo, "Exploring advanced techniques in process

automation and control: A generic framework for oil and gas industry applications," *Eng. Sci. Technol. J.*, vol. 5, no. 11, pp. 3127–3159, Nov. 2024, doi: 10.51594/estj.v5i11.1704.

[93] J. Xu, S. Zhou, Y. Tang, D. Huang, and Q. Zhu, "Alarm Ranking Model for Intelligent Management of Metro Systems Based on Statistical Machine Learning Methods," in *2020 Global Reliability and Prognostics and Health Management (PHM-Shanghai)*, Shanghai, China: IEEE, Oct. 2020, pp. 1–8. doi: 10.1109/PHM-Shanghai49105.2020.9280930.

[94] M. K. Chakravarthi, Y. V. P. Kumar, and G. P. Reddy, "Potential Technological Advancements in the Future of Process Control and Automation," in *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, Vilnius, Lithuania: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/eStream61684.2024.10542581.

[95] T. Lincy and S. S. Blessy, "Artificial Intelligence (AI) Driven Industrial Automation," in *Industry Automation: The Technologies, Platforms and Use Cases*, 1st ed., New York: River Publishers, 2024, pp. 85–100. doi: 10.1201/9781003516668-4.

[96] Y. Wenduo, G. Zongjin, S. Pengxiang, S. Maolin, P. Huiyang, and S. Ziyi, "Automation Design and Application of Electrical Control Systems in Industrial Production," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China: IEEE, Jul. 2024, pp. 187–190. doi: 10.1109/ICPICS62053.2024.10796180.

[97] K. R. Limkar and F. A. Tamboli, "Impact of Automation," *Int. J. Sci. Res. Mod. Sci. Technol.*, vol. 3, no. 8, pp. 13–17, Aug. 2024, doi: 10.59828/ijsrmst.v3i8.243.

[98] Z. I. Samigulina, A. K. Kurmasheva, and M. K. Kazbek, "Development of a process automation system for heating, ventilation and air conditioning for the food industry on the basis of honeywell equipment," *Her. Kazakh-Br. Tech. Univ.*, vol. 21, no. 1, pp. 28–41, Mar. 2024, doi: 10.55452/1998-6688-2024-21-1-28-41.

[99] A. A. Alfarhan and H. G. Alhazeem, "The Impact of Technology on Industrial Process Automation," *Int. J. Eng. Res. Appl.*, vol. 14, no. 12, pp. 20–27, Dec. 2024, doi: 10.9790/9622-14122027.

[100] S. Deng, "The Application of PLC in Industry Fields," *Highlights Sci. Eng. Technol.*, vol. 114, pp. 168–172, Oct. 2024, doi: 10.54097/28d22249.

[101] A. Saxena, K. A. Jabbar, and L. H. A. Fezaa, "Enhancing Industrial Automation: A Comprehensive Study on Programmable Logic Controllers (PLCs) and their Impact on Manufacturing Efficiency," in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan: IEEE, Nov. 2023, pp. 1182–1187. doi: 10.1109/ICTACS59847.2023.10390129.

[102] A. Bozzi, J.-F. Jimenez, C. Hernandez-Rodriguez, E.-M. Gonzalez-Neira, and D. Trentesaux, "Platoon-Based Distributed Control for Automated Material Handling Systems," in *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*, Rome, Italy: IEEE, Jul. 2023, pp. 2257–2262. doi: 10.1109/CoDIT58514.2023.10284111.

[103] I. B. Kadhim, "Automation in industry using Programmable Logic Controllers (PLCs) and SCADA," *Int. J. Circuit Comput. Netw.*, vol. 6, no. 1, pp. 05–10, Jan. 2025, doi: 10.33545/27075923.2025.v6.i1a.83.

[104] M. Nalini, "Industry Automation: The Contributions of Artificial Intelligence (AI)," in *Industry Automation: The Technologies, Platforms and Use Cases*, 1st ed., New York: River Publishers, 2024, pp. 25–55. doi: 10.1201/9781003516668-2.

[105] N. Kaur and A. Sharma, "Robotics and Automation in Manufacturing Processes," in *Intelligent Manufacturing*, 1st ed., Boca Raton: CRC Press, 2025, pp. 97–109. doi: 10.1201/9781032655758-7.

[106] G. Dorgo and J. Abonyi, "Sequence Mining Based Alarm Suppression," *IEEE Access*, vol. 6, pp. 15365–15379, 2018, doi: 10.1109/ACCESS.2018.2797247.

[107] Z. Ye, "Research on application of automatic control system in chemical safety production," in *Ninth International Symposium on Sensors, Mechatronics, and Automation System (ISSMAS 2023)*, L. Pan and Z. Zhou, Eds., Nanjing, China: SPIE, Mar. 2024, p. 101. doi: 10.1117/12.3014842.

[108] T. Ha *et al.*, "AI-driven robotic chemist for autonomous synthesis of organic molecules," *Sci. Adv.*, vol. 9, no. 44, p. eadj0461, Nov. 2023, doi: 10.1126/sciadv.adj0461.

[109] S. Ramamoorthy, "Role of AI, Automation & Robotics in Pharmaceutical Industry," *J. - Gener. Res. 50*, Dec. 2024, doi: 10.70792/jngr5.0.v1i1.45.

[110] S. Pawar, S. Baravani, S. Panhalkar, M. Kowadkar, and S. Bilgoji, "Introduction to AI in Automation-Transforming Industries through Intelligence," *Int. J. Multidiscip. Res.*, vol. 6, no. 3, p. 20440, May 2024, doi: 10.36948/ijfmr.2024.v06i03.20440.

[111] P. Madhavan and D. A. Wiegmann, "Effects of Information Source, Pedigree, and Reliability on Operator Interaction With Decision Support Systems," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 49, no. 5, pp. 773–785, Oct. 2007, doi: 10.1518/001872007X230154.

[112] E. J. De Visser *et al.*, "Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams," *Int. J. Soc. Robot.*, vol. 12, no. 2, pp. 459–478, May 2020, doi: 10.1007/s12369-019-00596-x.

[113] J. B. Lyons, T. Vo, K. T. Wynne, S. Mahoney, C. S. Nam, and D. Gallimore, "Trusting Autonomous Security Robots: The Role of Reliability and Stated Social Intent," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 63, no. 4, pp. 603–618, Jun. 2021, doi: 10.1177/0018720820901629.

[114] R. Wiczorek and D. Manzey, "Supporting Attention Allocation in Multitask Environments: Effects of Likelihood Alarm Systems on Trust, Behavior, and Performance," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 56, no. 7, pp. 1209–1221, Nov. 2014, doi: 10.1177/0018720814528534.

[115] S. McGregor, "Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 17, pp. 15458–15463, May 2021, doi: 10.1609/aaai.v35i17.17817.

[116] H. C. Joshi and S. Kumar, "Artificial Intelligence Failures in Autonomous Vehicles: Causes, Implications, and Prevention," *Computer*, vol. 57, no. 11, pp. 18–30, Nov. 2024, doi: 10.1109/MC.2024.3449435.

[117] A. Gautam and S. Thapaliya, "A Chronology of AI Failures in Safety and Cybersecurity," *NPRC J. Multidiscip. Res.*, vol. 1, no. 6, pp. 1–12, Nov. 2024, doi: 10.3126/nprcjmr.v1i6.71734.

[118] D. Virmani, M. K. L., G. P. Selvi, S. Javed, R. A. M., and M. R. Palav, "AI-Driven Predictive Maintenance in Asset Management:," in *Advances in Computational Intelligence and Robotics*, D. Pandey, B. Muniandi, B. K. Pandey, and A. S. George, Eds., IGI Global, 2024, pp. 397–408. doi: 10.4018/979-8-3373-1032-9.ch025.

[119] D. Patil, "Artificial Intelligence-Driven Predictive Maintenance In Manufacturing: Enhancing Operational Efficiency, Minimizing Downtime, And Optimizing Resource Utilization," 2025. doi: 10.2139/ssrn.5057406.

[120] S. Pokorni, “Reliability of artificial intelligence,” in *11th International Scientific Conference on Defensive Technologies - OTEX 2024 - zbornik radova*, Military Technical Institute, Belgrade, 2024, pp. 643–646. doi: 10.5937/OTEH24118P.

[121] E. E. Umoh, “Reliability of AI Algorithms in Safety Applications,” *Int. J. Eng. Adv. Technol. Stud.*, vol. 12, no. 2, pp. 74–85, Feb. 2024, doi: 10.37745/ijeats.13/vol12n27485.

[122] R. Cardoso Rial, “AI in analytical chemistry: Advancements, challenges, and future directions,” *Talanta*, vol. 274, p. 125949, Jul. 2024, doi: 10.1016/j.talanta.2024.125949.

[123] N. Bagheri and G. A. Jamieson, “The impact of context-related reliability on automation failure detection and scanning behaviour,” in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, The Hague, Netherlands: IEEE, 2004, pp. 212–217. doi: 10.1109/ICSMC.2004.1398299.

[124] E. Ferrara, “The Butterfly Effect in artificial intelligence systems: Implications for AI bias and fairness,” *Mach. Learn. Appl.*, vol. 15, p. 100525, Mar. 2024, doi: 10.1016/j.mlwa.2024.100525.

[125] A. S. Tejani, Y. S. Ng, Y. Xi, and J. C. Rayan, “Understanding and Mitigating Bias in Imaging Artificial Intelligence,” *RadioGraphics*, vol. 44, no. 5, p. e230067, May 2024, doi: 10.1148/rg.230067.

[126] Q. Huang, Z.-W. Gao, and Y. Liu, “Sensor Fault Reconstruction Using Robustly Adaptive Unknown-Input Observers,” *Sensors*, vol. 24, no. 10, p. 3224, May 2024, doi: 10.3390/s24103224.

[127] T. Yoshihiro, “A Two-step Automatic Calibration Method for Sensor Accuracy Management,” in *2023 19th International Conference on Intelligent Environments (IE)*, Uniciti, Mauritius: IEEE, Jun. 2023, pp. 1–4. doi: 10.1109/IE57519.2023.10179097.

[128] L. Bainbridge, “Ironies of automation,” *Automatica*, vol. 19, no. 6, pp. 775–779, Nov. 1983, doi: 10.1016/0005-1098(83)90046-8.

[129] A. B. Skjerve, S. Strand, R. Saarni, and G. Skraaning, “The influence of automation malfunctions and interface design on operator performance,” *Study Plan Prelim. Results HCA-2001 Exp.*, 2002.

[130] G. A. Jamieson and G. Skraaning, “Stumbling Towards a Shared Apprehension of Automation Failure,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 402–423, Dec. 2024, doi: 10.1177/15553434241292400.

[131] C. D. Wickens, L. Onnasch, A. Sebok, and D. Manzey, “Absence of DOA Effect but No Proper Test of the Lumberjack Effect: A Reply to Jamieson and Skraaning (2019),” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 62, no. 4, pp. 530–534, Jun. 2020, doi: 10.1177/0018720820901957.

[132] A. Sebok and C. D. Wickens, “Implementing Lumberjacks and Black Swans Into Model-Based Tools to Support Human–Automation Interaction,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 59, no. 2, pp. 189–203, Mar. 2017, doi: 10.1177/0018720816665201.

[133] R. J. Mumaw, “Not Automation Failures, but Automation Interface Failures,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 333–338, Dec. 2024, doi: 10.1177/15553434241228796.

[134] S. W. A. Dekker and D. D. Woods, “Wrong, Strong, and Silent: What Happens when Automated Systems With High Autonomy and High Authority Misbehave?,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 339–345, Dec. 2024, doi: 10.1177/15553434241240849.

[135] M. M. Van Paassen, A. Landman, C. Borst, and M. Mulder, “Framing Automation and Human Error in the Context of the Skill, Rule and Knowledge Taxonomy,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 318–326, Dec. 2024, doi: 10.1177/15553434241241892.

[136] B. S. Caldwell, “When an Automation Fails in the System, Who Hears? A Response to Skraaning and Jamieson,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 346–352, Dec. 2024, doi: 10.1177/15553434241230727.

[137] B. G. Liptak, “Automation can prevent the next Fukushima,” *Int. Soc. Autom.*, 2014.

[138] G. Skraaning and G. A. Jamieson, “The Failure to Grasp Automation Failure,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 274–285, Dec. 2024, doi: 10.1177/15553434231189375.

[139] D. Kanaan and B. Donmez, “How Are Automation Failures Characterized in the Driving Domain? Insights From a Scoping Review,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 293–301, Dec. 2024, doi: 10.1177/15553434241230604.

[140] S. C. Roberts, “Failures in Driving Automation Systems: Definitions, Taxonomy, and Prevention Mechanisms,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 302–309, Dec. 2024, doi: 10.1177/15553434241227534.

[141] E. K. Chiou, “Failing to Grasp our Failure to Grasp Automation Failure,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 370–376, Dec. 2024, doi: 10.1177/15553434241228799.

[142] N. J. Cooke, “Expanding Human Response to Automation Failures to Sociotechnical Systems,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 360–364, Dec. 2024, doi: 10.1177/15553434241229124.

[143] M. L. Cummings, “A Taxonomy for AI Hazard Analysis,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 327–332, Dec. 2024, doi: 10.1177/15553434231224096.

[144] A. R. Pritchett, “Things Go Wrong and the Captain Has to Handle it,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 365–369, Dec. 2024, doi: 10.1177/15553434241236536.

[145] S. Loft, “Accelerating Understanding of Human Response to Automation Failure,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 377–385, Dec. 2024, doi: 10.1177/15553434241234108.

[146] E. M. Roth, “The Importance of Studying and Guarding Against Systemic Automation Failures,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 353–359, Dec. 2024, doi: 10.1177/15553434241230605.

[147] J. Xing and N. Hughes Green, “A Regulatory Perspective: Have We Done Enough on Grasping Automation Failure?,” *J. Cogn. Eng. Decis. Mak.*, vol. 18, no. 4, pp. 310–317, Dec. 2024, doi: 10.1177/15553434241231056.

[148] P. I. Egbumokei, I. N. Dienagha, W. N. Digitemie, E. C. Onukwulu, and O. T. Oladipo, “Automation and worker safety: Balancing risks and benefits in oil, gas, and renewable energy industries,” *Int. J. Multidiscip. Res. Growth Eval.*, vol. 5, no. 4, pp. 1273–1283, 2024, doi: 10.54660/.ijmrge.2024.5.4.1273-1283.

[149] L. Methnani, A. Aler Tubella, V. Dignum, and A. Theodorou, “Let Me Take Over: Variable Autonomy for Meaningful Human Control,” *Front. Artif. Intell.*, vol. 4, p. 737072, Sep. 2021, doi: 10.3389/frai.2021.737072.

[150] A. Maddula, “Effect of Workload and Trust on Automation Levels in Human-Robot Collaboration,” *La. State Univ. Masters Theses*, 2024, doi: 10.31390/gradschool_theses.6055.

[151] B. C. Lee, J. Park, H. Jeong, and J. Park, “Validation of Trade-Off in Human–Automation Interaction: An Empirical Study of Contrasting Office Automation Effects on Task Performance and Workload,” *Appl. Sci.*, vol. 10, no. 4, p. 1288, Feb. 2020, doi: 10.3390/app10041288.

[152] L. Onnasch, C. D. Wickens, H. Li, and D. Manzey, “Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis,” *Hum. Factors J. Hum.*

Factors Ergon. Soc., vol. 56, no. 3, pp. 476–488, May 2014, doi: 10.1177/0018720813501549.

[153] A. I. Hauptman and N. J. McNeese, "Overcoming the Lumberjack Effect Through Adaptive Autonomy," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 1075–1079, Sep. 2022, doi: 10.1177/1071181322661372.

[154] A. Rühr, B. Berger, and T. Hess, "Can I Control My Robo-Advisor? Trade-Offs in Automation and User Control in (Digital) Investment Management," *Am. Conf. Inf. Syst.*, 2019.

[155] S. Wang, Z. Li, Y. Wang, W. Zhao, and H. Wei, "Quantification of safety improvements and human-machine tradeoffs in the transition to automated driving," *Accid. Anal. Prev.*, vol. 199, p. 107523, May 2024, doi: 10.1016/j.aap.2024.107523.

[156] G. Bansal, B. Nushi, E. Kamar, D. S. Weld, W. S. Lasecki, and E. Horvitz, "Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 2429–2437, Jul. 2019, doi: 10.1609/aaai.v33i01.33012429.

[157] N. Bagheri and G. A. Jamieson, "Considering subjective trust and monitoring behavior in assessing automation-induced 'complacency,'" *Hum. Perform. Situat. Aware. Autom. Curr. Res. Trends*, pp. 54–59, 2004.

[158] V. K. Bowden, N. Griffiths, L. Strickland, and S. Loft, "Detecting a Single Automation Failure: The Impact of Expected (But Not Experienced) Automation Reliability," *Hum. Factors*, 2021, doi: 10.1177/00187208211037188.

[159] A. Ladurini, S. Crate, M. Mitchell, S. H. Sánchez, R. Eble, and T. Brockhagen, "Reducing Workload: A Double-Edged Sword," *AIAA DATCIEEE 43rd Digit. Avion. Syst. Conf. DASC*, pp. 1–7, 2024, doi: 10.1109/dasc62030.2024.10749018.

[160] X. Dong and Z. Li, "Automation function and malfunction: effects on human performance in accident handling tasks," *Ergonomics*, vol. 67, no. 6, pp. 866–880, 2024, doi: 10.1080/00140139.2024.2307964.

[161] J. Davis, A. Atchley, H. Smitherman, H. Simon, and N. Tenhundfeld, "Measuring Automation Bias and Complacency in an X-Ray Screening Task," *2020 Syst. Inf. Eng. Des. Symp. SIEDS*, pp. 1–5, 2020, doi: 10.1109/SIEDS49339.2020.9106670.

[162] J. Sanchez, W. A. Rogers, A. D. Fisk, and E. Rovira, "Understanding reliance on automation: effects of error type, error distribution, age and experience," *Theor. Issues Ergon. Sci.*, vol. 15, no. 2, pp. 134–160, doi: 10.1080/1463922X.2011.611269.

[163] C. B. Orellana *et al.*, "The Impact of Automation Conditions on Reliance Dynamics and Decision-Making," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 721–725, 2022, doi: 10.1177/1071181322661477.

[164] M. Yang and J. Zhang, "Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, 2023, doi: 10.14569/IJACSA.2023.0140901.

[165] F. Yang, D. Xiao, and S. L. Shah, "Optimal Sensor Location Design for Reliable Fault Detection in Presence of False Alarms," *Sensors*, vol. 9, no. 11, pp. 8579–8592, Oct. 2009, doi: 10.3390/s91108579.

[166] M. R. Endsley and E. O. Kiris, "The Out-of-the-Loop Performance Problem and Level of Control in Automation," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 2, pp. 381–394, Jun. 1995, doi: 10.1518/001872095779064555.

[167] E. Hollnagel, *Safety-I and safety-II: the past and future of safety management*. Farnham: Ashgate, 2014.

[168] J. Lee, B. Bagheri, and H.-A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” *Manuf. Lett.*, vol. 3, pp. 18–23, Jan. 2015, doi: 10.1016/j.mfglet.2014.12.001.
[169] M. R. Endsley, “Toward a Theory of Situation Awareness in Dynamic Systems,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 32–64, Mar. 1995, doi: 10.1518/001872095779049543.
[170] C. D. Wickens, *Engineering psychology and human performance*, 1st ed. Columbus, OH: Charles Merrill, 1984.
[171] J. D. Lee and K. A. See, “Trust in Automation: Designing for Appropriate Reliance,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 46, no. 1, pp. 50–80, Jan. 2004, doi: 10.1518/hfes.46.1.50_30392.
[172] R. Parasuraman and V. Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 230–253, Jun. 1997, doi: 10.1518/001872097778543886.
[173] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering psychology and human performance*, Fourth edition., vol. Chapter 12 Automation and Human Performance. Boston: Pearson, 2013.
[174] M. R. Endsley, “Supporting Human-AI Teams:Transparency, explainability, and situation awareness,” *Comput. Hum. Behav.*, vol. 140, Mar. 2023, doi: https://doi.org/10.1016/j.chb.2022.107574.
[175] Y. Zhao, T. Li, X. Zhang, and C. Zhang, “Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future,” *Renew. Sustain. Energy Rev.*, vol. 109, pp. 85–101, Jul. 2019, doi: 10.1016/j.rser.2019.04.021.
[176] M. Chang, J.-L. Li, K. H. Chen, Y.-S. Chen, and L. Wang, “Advanced Fault Detection and Diagnosis with AI Techniques,” *37th Eur. Photovolt. Sol. Energy Conf. Exhib. 1388-1391*, p. 4 pages, 8181 kb, 2020, doi: 10.4229/EUPVSEC20202020-5DO.4.3.
[177] C.-C. Wang, “T2-LSTM-Based AI System for Early Detection of Motor Failure in Chemical Plants,” *Mathematics*, vol. 12, no. 17, p. 2652, Aug. 2024, doi: 10.3390/math12172652.
[178] M. Lu, Z. Gao, Y. Zou, Z. Chen, and P. Li, “Three-layer deep learning network random trees for fault detection in chemical production process,” 2024, *arXiv*. doi: 10.48550/ARXIV.2405.00311.
[179] R. J. Simonson, J. R. Keebler, E. L. Blickensderfer, and R. Besuijen, “Impact of alarm management and automation on abnormal operations: A human-in-the-loop simulation study,” *Appl. Ergon.*, vol. 100, p. 103670, Apr. 2022, doi: 10.1016/j.apergo.2021.103670.
[180] G. Jang, S. Suh, S. Kim, Y. Suh, and J. Park, “A proactive alarm reduction method and its human factors validation test for a main control room for SMART,” *Ann. Nucl. Energy*, vol. 51, pp. 125–134, Jan. 2013, doi: 10.1016/j.anucene.2012.07.035.
[181] R. Grimm, “Autonomous I/O-Colour-Screen-System for Process Control with Virtual Keyboards Adapted to the Actual Task,” in *Monitoring Behavior and Supervisory Control*, T. B. Sheridan and G. Johannsen, Eds., Boston, MA: Springer US, 1976, pp. 445–457. doi: 10.1007/978-1-4684-2523-9_37.
[182] International Society of Automation, “ANSI/ISA 18.2 Management of Alarm Systems for the Process Industries.” 2022. [Online]. Available: https://www.isa.org/intech-home/2016/may-june/departments/isa18-alarm-management-standard-updated

[183] "Corys Dynamic Simulation." 2015. [Online]. Available: https://www.corys.com/en/simulator-range
[184] S. K. Long, J. Lee, Y. Yamani, J. Unverricht, and M. Itoh, "Does automation trust evolve from a leap of faith? An analysis using a reprogrammed pasteurizer simulation task," *Appl. Ergon.*, vol. 100, p. 103674, Apr. 2022, doi: 10.1016/j.apergo.2021.103674.
[185] X. Chen, "Tennessee Eastman simulation dataset." IEEE DataPort, Jun. 09, 2019. doi: 10.21227/4519-Z502.
[186] The MathWorks Inc., "App Designer: Create desktop and web apps in MATLAB." 2022. [Online]. Available: https://www.mathworks.com/products/matlab/app-designer.html
[187] A. Bathelt, N. L. Ricker, and M. Jelali, "Revision of the Tennessee Eastman Process Model," *IFAC-Pap.*, vol. 48, no. 8, pp. 309–314, Jan. 2015, doi: 10.1016/j.ifacol.2015.08.199.
[188] I. A. Udugama, K. V. Gernaey, M. A. Taube, and C. Bayer, "A novel use for an old problem: The Tennessee Eastman challenge process as an activating teaching tool," *Educ. Chem. Eng.*, vol. 30, pp. 20–31, Jan. 2020, doi: 10.1016/j.ece.2019.09.002.
[189] Manca, Gianluca, "'Tennessee-Eastman-Process' Alarm Management Case Study." IEEE DataPort, Nov. 01, 2020. doi: 10.21227/326K-QR90.
[190] L. Ma, J. Dong, and K. Peng, "A novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes," *ISA Trans.*, vol. 96, pp. 1–13, Jan. 2020, doi: 10.1016/j.isatra.2019.06.004.
[191] International Society of Automation, "ISA101 Human-Machine Interfaces." 2022. [Online]. Available: https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa101
[192] K. Raghunandan, *Supervisory Control and Data Acquisition (SCADA)*. in Introduction to Wireless Communications and Networks. Cham: Springer, 2022.
[193] T. Macaulay and B. Singer, *Cybersecurity for industrial control systems: SCADA, DCS, PLC, HMI, and SIS*. Boca Raton, FL: CRC Press, 2012.
[194] T. Us, N. Jensen, M. Lind, and S. bay Jørgensen, "Fundamental Principles of Alarm Design," *Int. J. Nucl. Saf. Simul.*, vol. 2, no. 1, pp. 44–51, 2011.
[195] J. Wang, F. Yang, T. Chen, and S. L. Shah, "An Overview of Industrial Alarm Systems: Main Causes for Alarm Overloading, Research Status, and Open Problems," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 1045–1061, Apr. 2016, doi: 10.1109/TASE.2015.2464234.
[196] J. Miller, "Global precedence in attention and decision.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 7, no. 6, pp. 1161–1174, 1981, doi: 10.1037/0096-1523.7.6.1161.
[197] N. Dadashi, D. Golightly, and S. Sharples, "Seeing the woods for the trees: the problem of information inefficiency and information overload on operator performance," *Cogn. Technol. Work*, vol. 19, no. 4, pp. 561–570, Nov. 2017, doi: 10.1007/s10111-017-0451-1.
[198] D. A. Strobhar, "Human factors in process plant operation," *Momentum Press*, 2014.
[199] D. H. Rothenberg, *Alarm management for process control: a best-practice guide for design, implementation, and use of industrial alarm systems*. New York: Momentum Press, 2009.
[200] C. Aldrich and L. Auret, *Unsupervised process monitoring and fault diagnosis with machine learning methods*. New York: Springer, 2013.
[201] O. Foong, S. Sulaiman, D. R. B. A. Rambli, and N. Abdullah, "ALAP: Alarm prioritization system for oil refinery," *Proc World Congr. Eng. Comput. Sci.*, vol. 2, p. 2009.

[202] F. Higuchi, I. Yamamoto, T. Takai, M. Noda, and H. Nishitani, “Use of Event Correlation Analysis to Reduce Number of Alarms,” in *Computer Aided Chemical Engineering*, vol. 27, Elsevier, 2009, pp. 1521–1526. doi: 10.1016/S1570-7946(09)70644-3.

[203] Y. Cheng, I. Izadi, and T. Chen, “Pattern matching of alarm flood sequences by a modified Smith–Waterman algorithm,” *Chem. Eng. Res. Des.*, vol. 91, no. 6, pp. 1085–1094, Jun. 2013, doi: 10.1016/j.cherd.2012.11.001.

[204] Rockwell Automation, “Alarm Rationalization and Implementation,” *Publ. PROCES-WP015B-EN-P*, 2017, [Online]. Available: https://literature.rockwellautomation.com/idc/groups/literature/documents/wp/proces-wp015_-en-p.pdf

[205] J. M. Noyes and M. Bransby, Eds., *People in control: human factors in control room design*. in IEE control engineering series, no. 60. London: Institution of Electrical Engineers, 2001.

[206] NIST/SEMATECH e-Handbook of Statistical Methods, “6.3.2. What are variables control charts?,” 2013. [Online]. Available: https://www.itl.nist.gov/div898/handbook/pmc/section3/pmc32.htm#Trend

[207] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 30, no. 3, pp. 286–297, May 2000, doi: 10.1109/3468.844354.

[208] J. Sauer and A. Chavaillaz, “How operators make use of wide-choice adaptable automation: observations from a series of experimental studies,” *Theor. Issues Ergon. Sci.*, vol. 19, no. 2, pp. 135–155, Mar. 2018, doi: 10.1080/1463922X.2017.1297866.

[209] B. Lorenz, F. Di Nocera, S. Röttger, and R. Parasuraman, “The Effects of Level of Automation on the Out-of-the-Loop Unfamiliarity in a Complex Dynamic Fault-Management Task during Simulated Spaceflight Operations,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 45, no. 2, pp. 44–48, Oct. 2001, doi: 10.1177/154193120104500209.

[210] B. R. Mehta and Y. J. Reddy, *Industrial process automation systems: design and implementation*. Kidlington, Oxford: Butterworth-Heinemann, 2015.

[211] P. Anke and J. F. Krems, *Automation and situation awareness*. in The Handbook of Human-Machine Interaction. Boca Raton, FL: CRC Press, 2017.

[212] N. Cross, *The automated architect*. in Research in planning and design ; 4. London: Pion, 1977.

[213] S. Paul and M. Rosala, “The Wizard of Oz Method in UX,” Nielsen Norman Group, 2024.

[214] Laurel D. Riek, “Wizard of Oz studies in HRI: a systematic review and new reporting guidelines,” *J. Hum.-Robot Interact.*, vol. 1, no. 1, pp. 119–136, Jul. 2012, doi: https://doi.org/10.5898/JHRI.1.1.Riek.

[215] P. Ganesh, S. Widrow, J. Radadiya, C. D. Fitzpatrick, M. Knodler, and A. K. Pradhan, “A Wizard-of-Oz experimental approach to study the human factors of automated vehicles: Platform and methods evaluation,” *Taylor Francis*, vol. 21, pp. S140–S144, 2020, doi: https://doi.org/10.1080/15389588.2020.1810243.

[216] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, “Wizard of Oz studies: why and how,” *Proc. 1st Int. Conf. Intell. User Interfaces*, no. IUI ’93, pp. 193–200, 1993, doi: https://doi.org/10.1145/169891.169968.

[217] A. Trutnev, A. Rozenknop, and M. Rajman, “Speech Recognition Simulation and its Application for Wizard-of-Oz Experiments,” vol. Proceedings of the Fourth International Conference on Language Resources and Evaluation, no. LREC’04, 2004.

[218] T. Hitron, I. Wald, H. Erel, and O. Zuckerman, “Introducing children to machine learning concepts through hands-on experience,” *Proc. 17th ACM Conf. Interact. Des. Child.*, no. IDC ’18, pp. 563–568, 2018.
[219] J. T. Browne, “Wizard of Oz Prototyping for Machine Learning Experiences,” *Ext. Abstr. 2019 CHI Conf. Hum. Factors Comput. Syst.*, no. CHI EA ’19, pp. 1–6, 2019.
[220] T. A. J. Schoonderwoerd, E. M. van Zoelen, K. van den Bosch, and M. A. Neerincx, “Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task,” *Int. J. Hum.-Comput. Stud.*, vol. 164, no. 102831, 2022, doi: https://doi.org/10.1016/j.ijhcs.2022.102831.
[221] K. Ung, O. Nemer, Krishna, M. Chioua, and P. Doyon-Poulin, “PER4Mance Prototyping environment for research onhuman-machine interactions for alarm floods management: thecase study of a chemical plant process control,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 430–434, 2022, doi: https://doi.org/10.1177/1071181322661248.
[222] C. K. Lau, K. Ghosh, M. A. Hussain, and C. R. Che Hassan, “Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS,” *Chemom. Intell. Lab. Syst.*, vol. 120, pp. 1–14, Jan. 2013, doi: 10.1016/j.chemolab.2012.10.005.
[223] E. Galy, J. Paxion, and C. Berthelon, “Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving,” *Ergonomics*, vol. 61, no. 4, pp. 517–527, Apr. 2018, doi: 10.1080/00140139.2017.1369583.
[224] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” in *Advances in Psychology*, vol. 52, Elsevier, 1988, pp. 139–183. doi: 10.1016/S0166-4115(08)62386-9.
[225] A. Holm, K. Lukander, J. Korpela, M. Sallinen, and K. M. I. Müller, “Estimating Brain Load from the EEG,” *Sci. World J.*, vol. 9, pp. 639–651, 2009, doi: 10.1100/tsw.2009.83.
[226] R. A. Grier, “How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 59, no. 1, pp. 1727–1731, 2016, doi: https://doi.org/10.1177/1541931215591373.
[227] M. R. Endsley and D. J. Garland, *Situation awareness: analysis and measurement*. Mahwah, N.J.: L. Erlbaum Associates, 2000. Accessed: Aug. 03, 2021. [Online]. Available: http://site.ebrary.com/id/10346741
[228] M. R. Endsley and Jones, Debra G., *Designing for situation awareness: an approach to user-centered design*, 2nd ed., vol. Chapter 14 Evaluating Design Concepts for SA. Boca Raton, FL: CRC Press, 2011.
[229] M. R. Endsley, “A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 63, no. 1, pp. 124–150, Feb. 2021, doi: 10.1177/0018720819875376.
[230] E. Coolen, J. Draaisma, and J. Loeffen, “Measuring situation awareness and team effectiveness in pediatric acute care by using the situation global assessment technique,” *Eur. J. Pediatr.*, vol. 178, Jun. 2019, doi: 10.1007/s00431-019-03358-z.
[231] E. Fabris, V. Abichequer Sangalli, L. Soares, and M. Pinho, “Immersive telepresence on the operation of unmanned vehicles,” *Int. J. Adv. Robot. Syst.*, vol. 18, Feb. 2021, doi: 10.1177/1729881420978544.
[232] J. C. F. De Winter, Y. B. Eisma, C. D. D. Cabrall, P. A. Hancock, and N. A. Stanton, “Situation awareness based on eye movements in relation to the task environment,” *Cogn. Technol. Work*, vol. 21, no. 1, pp. 99–111, Feb. 2019, doi: 10.1007/s10111-018-0527-6.

[233] M. R. Endsley, “Direct Measurement of Situation Awareness: Validity and Use of SAGAT,” in *Situational Awareness*, 1st ed., E. Salas, Ed., Routledge, 2017, pp. 129–156. doi: 10.4324/9781315087924-9.

[234] L. Tao, Q. Wang, D. Liu, J. Wang, Z. Zhu, and L. Feng, “Eye tracking metrics to screen and assess cognitive impairment in patients with neurological disorders,” *Neurol. Sci.*, vol. 41, no. 7, pp. 1697–1704, Jul. 2020, doi: 10.1007/s10072-020-04310-y.

[235] M. A. Just and P. A. Carpenter, “Eye fixations and cognitive processes,” *Cognit. Psychol.*, vol. 8, no. 4, pp. 441–480, Oct. 1976, doi: 10.1016/0010-0285(76)90015-3.

[236] G. G. Menekse Dalveren and N. E. Cagiltay, “Insights from surgeons’ eye-movement data in a virtual simulation surgical training environment: effect of experience level and hand conditions,” *Behav. Inf. Technol.*, vol. 37, no. 5, pp. 517–537, May 2018, doi: 10.1080/0144929X.2018.1460399.

[237] J.-C. Liu, K.-A. Li, S.-L. Yeh, and S.-Y. Chien, “Assessing Perceptual Load and Cognitive Load by Fixation-Related Information of Eye Movements,” *Wearable Technol. Appl. Eye Track.*, 2022, doi: 10.3390/s22031187.

[238] K. Holmqvist and R. Andersson, *Eye tracking: a comprehensive guide to methods, paradigms, and measures*, 2nd edition. Lund, Sweden: Lund Eye-Tracking Research Institute, 2017.

[239] McKinsey & Company, “The state of AI in 2023: Generative AI’s breakout year,” 2023. [Online]. Available: https://www.mckinsey.com/mgi/overview/in-the-news/ai-could-increase-corporate-profits-by-4-trillion-a-year-according-to-new-research

[240] National Academies of Sciences, Engineering, and Medicine, *Artificial Intelligence and the Future of Work*. Washington, D.C.: National Academies Press, 2024, p. 27644. doi: 10.17226/27644.

[241] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, “To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making,” *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–21, Apr. 2021, doi: 10.1145/3449287.

[242] M. R. Endsley, “Ironies of artificial intelligence,” *Ergonomics*, vol. 66, no. 11, pp. 1656–1668, 2023, doi: 10.1080/00140139.2023.2243404.

[243] G. A. Jamieson and G. Skraaning, “The Absence of Degree of Automation Trade-Offs in Complex Work Settings,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 62, no. 4, pp. 516–529, Jun. 2020, doi: 10.1177/0018720819842709.

[244] G. Paaß and D. Hecker, “What Is Intelligent About Artificial Intelligence?,” in *Artificial Intelligence*, Cham: Springer Nature Switzerland, 2024, pp. 1–13. doi: 10.1007/978-3-031-50605-5_1.

[245] T. Gamer, M. Hoernicke, B. Kloepper, R. Bauer, and A. J. Isaksson, “The autonomous industrial plant – future of process engineering, operations and maintenance,” *J. Process Control*, vol. 88, pp. 101–110, Apr. 2020, doi: 10.1016/j.jprocont.2020.01.012.

[246] M. Müller, T. Müller, B. Ashtari Talkhestani, P. Marks, N. Jazdi, and M. Weyrich, “Industrial autonomous systems: a survey on definitions, characteristics and abilities,” - *Autom.*, vol. 69, no. 1, pp. 3–13, Jan. 2021, doi: 10.1515/auto-2020-0131.

[247] A. Isaksson and T. Gamer, “Autonomous systems–five years on,” *ABB Rev.*, pp. 100–105, 2020.

[248] F. Mo *et al.*, “A maturity model for the autonomy of manufacturing systems,” *Int. J. Adv. Manuf. Technol.*, vol. 126, no. 1–2, pp. 405–428, May 2023, doi: 10.1007/s00170-023-10910-7.

[249] M. Paluszek, S. Thomas, and E. Ham, "What Is Deep Learning?," in *Practical MATLAB Deep Learning*, Berkeley, CA: Apress, 2022, pp. 1–24. doi: 10.1007/978-1-4842-7912-0_1.
[250] J. Sifakis, "Autonomous Systems -- An Architectural Characterization," 2018, *arXiv*. doi: 10.48550/ARXIV.1811.10277.
[251] B. Shneiderman, "Human-Centered AI: A New Synthesis," in *Human-Computer Interaction – INTERACT 2021*, vol. 12932, C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, Eds., in Lecture Notes in Computer Science, vol. 12932. , Cham: Springer International Publishing, 2021, pp. 3–8. doi: 10.1007/978-3-030-85623-6_1.
[252] S. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 2016.
[253] S. S. Chanda and D. N. Banerjee, "Omission and commission errors underlying AI failures," *AI Soc.*, vol. 39, no. 3, pp. 937–960, Jun. 2024, doi: 10.1007/s00146-022-01585-x.
[254] K. A. H. Kobbacy, "Artificial Intelligence in Maintenance," in *Complex System Maintenance Handbook*, in Springer Series in Reliability Engineering. , London: Springer London, 2008, pp. 209–231. doi: 10.1007/978-1-84800-011-7_9.
[255] R. Isermann, *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Berlin New York: Springer, 2006.
[256] C. Gattino, E. Ottonello, G. Berselli, and J. Stecki, "Application of AI failure identification techniques in condition monitoring using wavelet analysis," *PHM Soc. Eur. Conf.*, vol. 5, no. 1, p. 12, Jul. 2020, doi: 10.36001/phme.2020.v5i1.1255.
[257] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Hybrid deep fault detection and isolation: Combining deep neural networks and system performance models," Dec. 28, 2019, *arXiv*: arXiv:1908.01529. doi: 10.48550/arXiv.1908.01529.
[258] M. Diwakar, S. Sharma, R. Dhabliya, R. Sonar, S. T. Shirkande, and S. Bhattacharya, "AIdriven Strategy for Predicting Equipment Failure in Manufacturing," in *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*, Jaipur India: ACM, Nov. 2023, pp. 1–5. doi: 10.1145/3647444.3647932.
[259] N. G. Leveson, *Engineering a safer world: systems thinking applied to safety*, New paperback edition. in Engineering systems. Cambridge, Massachusetts London, Englang: The MIT Press, 2017.
[260] G. Yu *et al.*, "A Survey on Failure Analysis and Fault Injection in AI Systems," Jun. 28, 2024, *arXiv*: arXiv:2407.00125. doi: 10.48550/arXiv.2407.00125.
[261] ANSI/ISA-18.2, "Management of Alarm Systems for the Process Industries." 2016.
[262] S. R. Dixon and C. D. Wickens, "On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses?," *Hum. Factors J. Hum. Factors Ergon. Soc.*, p. 49(4):564-72, Sep. 2007, doi: 10.1518/001872007X215656.
[263] CSB, "ExxonMobil Torrance Refinery." U.S. Chemical Safety and Hazard Investigation Board, 2015. [Online]. Available: https://www.csb.gov/file.aspx?DocumentId=6023
[264] D. J. Bruemmer, D. A. Few, R. L. Boring, J. L. Marble, M. C. Walton, and C. W. Nielsen, "Shared Understanding for Collaborative Control," *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 35, no. 4, pp. 494–504, Jul. 2005, doi: 10.1109/TSMCA.2005.850599.
[265] O. Gillath, T. Ai, M. S. Branicky, S. Keshmiri, R. B. Davison, and R. Spaulding, "Attachment and trust in artificial intelligence," *Comput. Hum. Behav.*, vol. 115, p. 106607, Feb. 2021, doi: 10.1016/j.chb.2020.106607.
[266] S. C. Kohn, E. J. De Visser, E. Wiese, Y.-C. Lee, and T. H. Shaw, "Measurement of Trust in Automation: A Narrative Review and Reference Guide," *Front. Psychol.*, vol. 12, p. 604977, Oct. 2021, doi: 10.3389/fpsyg.2021.604977.

[267] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An Integrative Model of Organizational Trust,” *Acad. Manage. Rev.*, vol. 20, no. 3, p. 709, Jul. 1995, doi: 10.2307/258792.

[268] K. De Fine Licht and B. Brülde, “On Defining ‘Reliance’ and ‘Trust’: Purposes, Conditions of Adequacy, and New Definitions,” *Philosophia*, vol. 49, no. 5, pp. 1981–2001, Nov. 2021, doi: 10.1007/s11406-021-00339-1.

[269] K. A. Hoff and M. Bashir, “Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 57, no. 3, pp. 407–434, May 2015, doi: 10.1177/0018720814547570.

[270] A. Klingbeil, C. Grützner, and P. Schreck, “Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI,” *Comput. Hum. Behav.*, vol. 160, p. 108352, Nov. 2024, doi: 10.1016/j.chb.2024.108352.

[271] Y. Zhang, Q. V. Liao, and R. K. E. Bellamy, “Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making,” 2020, doi: 10.48550/ARXIV.2001.02114.

[272] J. M. McGuirl and N. B. Sarter, “Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 48, no. 4, pp. 656–665, Dec. 2006, doi: 10.1518/001872006779166334.

[273] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, “A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 58, no. 3, pp. 377–400, May 2016, doi: 10.1177/0018720816634228.

[274] C. J. Pearson, M. Geden, and C. B. Mayhorn, “Who’s the real expert here? Pedigree’s unique bias on trust between human and automated advisers,” *Appl. Ergon.*, vol. 81, p. 102907, Nov. 2019, doi: 10.1016/j.apergo.2019.102907.

[275] W. B. Rouse, *Computing Possible Futures*, 1st ed. Oxford University PressOxford, 2019. doi: 10.1093/oso/9780198846420.001.0001.

[276] M. Miró-Nicolau, G. Moyà-Alcover, A. Jaume-i-Capó, M. González-Hidalgo, M. G. S. Campello, and J. A. P. Sancho, “To Trust or Not to Trust: Towards a novel approach to measure trust for XAI systems,” 2024, *arXiv*. doi: 10.48550/ARXIV.2405.05766.

[277] D. Dalcher, “Why the pilot cannot be blamed: a cautionary note about excessive reliance on technology,” *Int. J. Risk Assess. Manag.*, vol. 7, no. 3, p. 350, 2007, doi: 10.1504/IJRAM.2007.011988.

[278] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, “Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance,” *Front. Comput. Sci.*, vol. 5, p. 1096257, Feb. 2023, doi: 10.3389/fcomp.2023.1096257.

[279] T. Louw, G. Markkula, E. Boer, R. Madigan, O. Carsten, and N. Merat, “Coming back into the loop: Drivers’ perceptual-motor performance in critical events after automated driving,” *Accid. Anal. Prev.*, vol. 108, pp. 9–18, Nov. 2017, doi: 10.1016/j.aap.2017.08.011.

[280] N. Merat *et al.*, “The ‘Out-of-the-Loop’ concept in automated driving: proposed definition, measures and implications,” *Cogn. Technol. Work*, vol. 21, no. 1, pp. 87–98, Feb. 2019, doi: 10.1007/s10111-018-0525-8.

[281] T. B. Sheridan, “Human–Robot Interaction: Status and Challenges,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 58, no. 4, pp. 525–532, Jun. 2016, doi: 10.1177/0018720816644364.

[282] J. Gouraud, A. Delorme, and B. Berberian, “Autopilot, Mind Wandering, and the Out of the Loop Performance Problem,” *Front. Neurosci.*, vol. 11, p. 541, Oct. 2017, doi: 10.3389/fnins.2017.00541.
[283] M. Lewis, K. Sycara, and P. Walker, “The Role of Trust in Human-Robot Interaction,” in *Foundations of Trusted Autonomy*, vol. 117, H. A. Abbass, J. Scholz, and D. J. Reid, Eds., in Studies in Systems, Decision and Control, vol. 117. , Cham: Springer International Publishing, 2018, pp. 135–159. doi: 10.1007/978-3-319-64816-3_8.
[284] I. Rafiq, A. Mahmood, S. Razzaq, S. H. M. Jafri, and I. Aziz, “IoT applications and challenges in smart cities and services,” *J. Eng.*, vol. 2023, no. 4, p. e12262, Apr. 2023, doi: 10.1049/tje2.12262.
[285] D. K. Srivastava, J. M. Lilly, and K. M. Feigh, “Improving Operator Situation Awareness when Working with AI Recommender Systems,” 2023, *arXiv*. doi: 10.48550/ARXIV.2310.11370.
[286] M. Cappelli, A. Gadomski, and M. Sepielli, “Human Factors in Nuclear Power Plant Safety Management: A Socio-Cognitive Modeling Approach using TOGA Meta-Theory.,” *Proc. Int. Congr. Adv. Nucl. Power Plants*, 2011.
[287] O. Okuyelu and O. Adaji, “AI-Driven Real-time Quality Monitoring and Process Optimization for Enhanced Manufacturing Performance,” *J. Adv. Math. Comput. Sci.*, vol. 39, no. 4, pp. 81–89, Mar. 2024, doi: 10.9734/jamcs/2024/v39i41883.
[288] J. Jiang, A. J. Karran, C. K. Coursaris, P.-M. Léger, and J. Beringer, “A Situation Awareness Perspective on Human-AI Interaction: Tensions and Opportunities,” *Int. J. Human–Computer Interact.*, vol. 39, no. 9, pp. 1789–1806, May 2023, doi: 10.1080/10447318.2022.2093863.
[289] A. Bhakte, M. Chakane, and R. Srinivasan, “Alarm-based explanations of process monitoring results from deep neural networks,” *Comput. Chem. Eng.*, vol. 179, p. 108442, Nov. 2023, doi: 10.1016/j.compchemeng.2023.108442.
[290] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” 2017, *arXiv*. doi: 10.48550/ARXIV.1702.08608.
[291] C. Molnar, *Interpretable machine learning: a guide for making black box models explainable*. Victoria, British Columbia: Leanpub, 2020.
[292] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.
[293] R. Sargent, B. Walters, and C. Wickens, “Meta-analysis Qualifying and Quantifying the Benefits of Automation Transparency to Enhance Models of Human Performance,” *Hum.-Comput. Interact. HCII 2023 Lect. Notes Comput. Sci.*, vol. 14011, 2023, doi: https://doi.org/10.1007/978-3-031-35596-7_16.
[294] M. R. Endsley, “From Here to Autonomy: Lessons Learned From Human–Automation Research,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 59, no. 1, pp. 5–27, Feb. 2017, doi: 10.1177/0018720816681350.
[295] S. Malik, K. Muhammad, and Y. Waheed, “Artificial intelligence and industrial applications-A revolution in modern industries,” *Ain Shams Eng. J.*, vol. 15, no. 9, p. 102886, Sep. 2024, doi: 10.1016/j.asej.2024.102886.
[296] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. Barnes, “Situation awareness-based agent transparency and human-autonomy teaming effectiveness,” *Theor. Issues Ergon. Sci.*, vol. 19, no. 3, pp. 259–282, May 2018, doi: 10.1080/1463922X.2017.1315750.

[297] G. Skraaning and G. A. Jamieson, “Human Performance Benefits of The Automation Transparency Design Principle: Validation and Variation,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 63, no. 3, pp. 379–401, May 2021, doi: 10.1177/0018720819887252.

[298] C. D. Wickens, W. S. Helton, J. G. Hollands, and S. Banbury, *Engineering Psychology and Human Performance*, 5th ed. New York: Routledge, 2021. doi: 10.4324/9781003177616.

[299] S. Ma *et al.*, “Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making,” 2023, *arXiv*. doi: 10.48550/ARXIV.2301.05809.

[300] G. Park, “The effect of level of AI transparency on human-AI teaming performance.” Master’s Thesis, University of Michigan, 2023. [Online]. Available: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/176345/GeeBeum%20Park%20Final%20Thesis.pdf?sequence=1

[301] A. Maged, S. Haridy, and H. Shen, “Explainable Artificial Intelligence Techniques for Accurate Fault Detection and Diagnosis: A Review,” Jun. 10, 2024, *arXiv*: arXiv:2404.11597. doi: 10.48550/arXiv.2404.11597.

[302] P. Agarwal, M. Tamer, and H. Budman, “Explainability: Relevance based dynamic deep learning algorithm for fault detection and diagnosis in chemical processes,” *Comput. Chem. Eng.*, vol. 154, p. 107467, Nov. 2021, doi: 10.1016/j.compchemeng.2021.107467.

[303] C. Belavadi, V. S. Sardar, and S. S. Chaudhari, “Alarm Pattern Recognition in Continuous Process Control Systems using Data Mining,” *Int. J. Comput.*, pp. 333–341, Sep. 2022, doi: 10.47839/ijc.21.3.2689.

[304] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, “A review of process fault detection and diagnosis,” *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, Mar. 2003, doi: 10.1016/S0098-1354(02)00160-6.

[305] D. C. Montgomery, *Introduction to statistical quality control*, Eighth edition, EMEA edition. Hoboken, NJ: Wiley, 2020.

[306] K. Ung, K. Khismetzhan, M. Chioua, and P. Doyon-Poulin, “Automated diagnostic tool supports human performance during alarm floods: a case study in a chemical plant simulator,” *Manuscr. Submitt. Publ.*, 2024.

[307] C. Reinartz, T. Thuesen Enevoldsen, R. Galeazzi, and O. Ravn, “A Causal Model-based Planner for the Reconfiguration of Continuous Processes,” in *2021 European Control Conference (ECC)*, Delft, Netherlands: IEEE, Jun. 2021, pp. 1751–1756. doi: 10.23919/ECC54610.2021.9655016.

[308] J. J. Downs and E. F. Vogel, “A plant-wide industrial process control problem,” *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993, doi: 10.1016/0098-1354(93)80018-I.

[309] F. Rietz, A. Sutherland, S. Bensch, S. Wermter, and T. Hellström, “WoZ4U: An Open-Source Wizard-of-Oz Interface for Easy, Efficient and Robust HRI Experiments,” *Front Robot AI*, vol. 8, no. 668057, 2021, doi: 10.3389/frobt.2021.668057.

[310] F. Diederichs, M. Lesley-Ann, and V. Bopp-Bertenbreiter, “A Wizard-of-Oz vehicle to investigate human interaction with AI-driven automated cars.,” *Proc. DSC*, 2021.

[311] A. Jansen and S. Colombo, “Wizard of Errors: Introducing and Evaluating Machine Learning Errors in Wizard of Oz Studies,” *CHI EA 22 CHI Conf. Hum. Factors Comput. Syst. Ext. Abstr.*, pp. 1–7, 2022, doi: https://doi.org/10.1145/3491101.3519684.

[312] J. B. Lyons and S. Y. Guznov, “Individual differences in human–machine trust: A multi-study look at the perfect automation schema,” *Theor. Issues Ergon. Sci.*, vol. 20, no. 4, pp. 440–458, Jul. 2019, doi: 10.1080/1463922X.2018.1491071.

[313] Z. Li, “Enhancing Human-AI Collaboration through Adaptive Interaction and Explainability,” *Proc. AAAIACM Conf. AI Ethics Soc.*, vol. 7, no. 2, pp. 26–27, Jan. 2025, doi: 10.1609/aies.v7i2.31900.

[314] S. Afroogh, A. Akbari, E. Malone, M. Kargar, and H. Alambeigi, “Trust in AI: progress, challenges, and future directions,” *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, p. 1568, Nov. 2024, doi: 10.1057/s41599-024-04044-8.

[315] S. Mehrotra, C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman, “A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges,” *ACM J. Responsible Comput.*, vol. 1, no. 4, pp. 1–45, Dec. 2024, doi: 10.1145/3696449.

[316] T. Maier, J. Menold, and C. McComb, “The Relationship Between Performance and Trust in AI in E-Finance,” *Front. Artif. Intell.*, vol. 5, p. 891529, Jun. 2022, doi: 10.3389/frai.2022.891529.

[317] A. Kawakami *et al.*, “Training Towards Critical Use: Learning to Situate AI Predictions Relative to Human Knowledge,” Aug. 30, 2023, *arXiv*: arXiv:2308.15700. doi: 10.48550/arXiv.2308.15700.

APPENDIX A NASA TASK LOAD INDEX (NASA-TLX)

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

ID	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

For each of the pairs listed below, circle the dimension that represents the more important contributor to workload during the tasks?

Mental Demand	or	Physical Demand
Mental Demand	or	Temporal Demand
Mental Demand	or	Performance
Mental Demand	or	Effort
Mental Demand	or	Frustration
Physical Demand	or	Temporal Demand
Physical Demand	or	Performance
Physical Demand	or	Effort
Physical Demand	or	Frustration
Temporal Demand	or	Performance
Temporal Demand	or	Effort
Temporal Demand	or	Frustration
Performance	or	Effort
Performance	or	Frustration
Effort	or	Frustration

APPENDIX B SITUATION AWARENESS GLOBAL ASSESSMENT TECHNIQUE (SAGAT)

(a) How many alarms are triggered?	(a) 0-3 (b) 4-6 (c) 7-9 (d) 10+
(b) What is the product concentration of G?	(a) 10% - 25% (b) 26% - 45% (c) 46% - 60% (d) 61% - 80% (e) 81% - 100%
(c) Approximately, what is the current cost of production?	(a) 0$ - 100$ (b) 101$ - 200$ (c) 201$ - 500$ (d) 501$ - 700$ (e) over 700$
(d) Which feed has a red alert? Select all applicable.	(a) A (b) E (c) C (d) D (e) none
(e) Which element(s) are causing the fault?	(a) Feeds (b) Reactor (c) Condenser (d) Stripper (e) Purge (f) There is no fault
(f) What is the projected situation over the next 10 minutes?	(a) Simulator shutdown (b) Need to apply a corrective action (c) Simulation will operate normally

APPENDIX C TRUST AND RELIANCE QUESTIONNAIRE

	Question	(1) Strongly disagree	(2) Disagree	(3) Neither agree nor disagree	(4) Agree	(5) Strongly agree
1	If I had my way, I would NOT let the diagnostic tool have any influence over issues that are important (ex: fix a fault).	☐	☐	☐	☐	☐
2	I would be comfortable giving the diagnostic tool complete responsibility for the monitoring task.	☐	☐	☐	☐	☐
3	I really wish I had a good way to monitor the route decisions of the diagnostic tool.	☐	☐	☐	☐	☐
4	I would be comfortable allowing the diagnostic tool to implement its route decision, even if I could not monitor it.	☐	☐	☐	☐	☐
5	I would rely on the diagnostic tool without hesitation.	☐	☐	☐	☐	☐
6	I think using the diagnostic tool will lead to positive outcomes.	☐	☐	☐	☐	☐
7	I would feel comfortable relying on the diagnostic tool in the future.	☐	☐	☐	☐	☐
8	When the task was hard, I felt like I could depend on the diagnostic tool.	☐	☐	☐	☐	☐
9	If I were facing a very hard task in the future, I would want to have this diagnostic tool with me.	☐	☐	☐	☐	☐
10	I would be comfortable allowing this diagnostic tool to make all decisions.	☐	☐	☐	☐	☐

APPENDIX D CERTIFICAT D'ÉTHIQUE

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

CER-2122-48-D

Montréal, le 2 mars 2022

Objet: Approbation éthique – « Effects of an automated fault diagnostic algorithm on the human performance, trust and workload in a process control » - Projet CER-2122-48-D

Mme Karine Ung,

J'ai le plaisir de vous informer que le Comité d'éthique de la recherche, selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche susmentionné et conclu que ce dernier répond aux normes en vigueur au chapitre de l'éthique de la recherche énoncées dans la *Politique en matière d'éthique de la recherche avec des êtres humains* de Polytechnique Montréal.

Veuillez noter que le présent certificat est valable pour une durée d'un an, soit du 2 mars 2022 au 3 mars 2023, pour le projet tel qu'approuvé au Comité d'éthique de la recherche avec des êtres humains.

Veuillez noter que conformément aux exigences auxquelles l'institution et son personnel sont assujettis afin d'être admissibles aux fonds des organismes subventionnaires, il est de votre responsabilité de déposer au CÉR un rapport annuel ou un rapport final avant l'expiration de la présente approbation éthique afin de l'informer de l'avancement de vos travaux. Le formulaire à remplir est disponible à l'adresse suivante : (http://www.polymtl.ca/recherche/formulaires-et-guides).

De plus, il est de votre responsabilité d'informer le CER de toute modification importante qui pourrait être apportée au protocole expérimental avant sa mise en œuvre, de même que de tout élément ou évènement imprévu pouvant avoir une incidence sur le bien-être ou l'intégrité des participant(e)s impliqué(e)s dans le projet de recherche. Nous vous invitons aussi à nous signaler tout problème susceptible d'avoir une incidence sur les membres de l'équipe de recherche.

Je vous souhaite bonne chance dans la poursuite de vos travaux.

Nous vous prions d'agréer, Madame, l'expression de nos sentiments les meilleurs,

Farida Cheriet, présidente
Comité d'éthique de la recherche
Polytechnique Montréal

c.c. Direction de la formation et de la recherche; Service des Finances
Philippe Doyon-Poulin, Professeur adjoint, Département de mathématiques et de génie industriel
Moncef Chioua, professeur adjoint, Département de génie chimique
Cochercheurs
p.j. Certificat # CER-2122-48-D

Comité d'éthique de la recherche avec des êtres humains
Tél.: 514 340-4711 poste : 3755
Fax : 514 340-4992
Courriel : ethique@polymtl.ca

Adresse postale
C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7

Campus de l'Université de Montréal
2900, boul. Édouard-Montpetit
2500, chemin de Polytechnique
Montréal (Québec) Canada H3T1J4

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

CER-2122-48-D

CERTIFICAT D'APPROBATION ÉTHIQUE

Le Comité d'éthique de la recherche de Polytechnique Montréal, selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche suivant et conclu qu'il respecte les règles d'éthique énoncées dans sa Politique en matière d'éthique de la recherche avec des êtres humains.

Projet	
Titre du projet	**Effects of an automated fault diagnostic algorithm on the human performance, trust and workload in a process control CER-2122-48-D**
Étudiante requérante	**Karine Ung**, Candidate au PhD, Département de mathématiques et de génie industriel
Sous la direction de:	**Philippe Doyon-Poulin, Professeur adjoint**, Département de mathématiques et de génie industriel, Polytechnique Montréal **& Moncef Chioua, professeur adjoint**, Département de génie chimique, Polytechnique Montréal.
Avec la collaboration de:	Jean-Marc Robert (Co-directeur, PolyMTL), Colline Melennec (INSA Lyon)

Financement	
Organisme	Bourse
No de UBR	
Programme	
No d'octroi:	
Titre original de l'octroi:	
Chercheur principal:	

MODALITÉS D'APPLICATION

Toute modification importante qui pourrait être apportée au protocole expérimental doit être transmise au Comité avant sa mise en œuvre.

L'équipe de recherche doit informer le Comité de tout élément ou évènement imprévu pouvant avoir une incidence sur le bien-être ou l'intégrité des participant(e)s impliqué(e)s dans le projet de recherche ainsi que tout problème susceptible d'avoir une incidence sur les membres de l'équipe de recherche.

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du Comité.

Farida Cheriet, présidente
Comité d'éthique de la recherche
Polytechnique Montréal

Date de délivrance :
2 mars 2022

Date de fin de validité :
1er avril 2023

Date du prochain suivi :
3 mars 2023

Comité d'éthique de la recherche avec des êtres humains
Tél : 514 340-4711 poste : 3755
Fax : 514 340-4992
Courriel : ethique@polymtl.ca

Adresse postale
C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7

Campus de l'Université de Montréal
2900, boul. Édouard-Montpetit
2500, chemin de Polytechnique
Montréal (Québec) Canada H3T 1J4