Digital Chemical Engineering 11 (2024) 100151

Contents lists available at ScienceDirect

Digital Chemical Engineering

journal homepage: www.elsevier.com/locate/dche

Perspectives

# Artificial intelligence – Human intelligence conflict and its impact on process system safety

Rajeevan Arunthavanathan [a,b], Zaman Sajid [a,b], Faisal Khan [a,b,*], Efstratios Pistikopoulos [a,c]

[a] Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, 77843, USA
[b] Mary Kay O'Connor Process Safety Center (MKOPSC), Texas A&M University, College Station TX 77843-3122, USA
[c] Texas A&M Energy Institute, Texas A&M University, College Station, TX, 77843, USA

ARTICLE INFO

*Keywords:*
Artificial intelligence safety
Artificial intelligence conflict
Process System Safety
Process Safety
Automation Safety

ABSTRACT

In the Industry 4.0 revolution, industries are advancing their operations by leveraging Artificial Intelligence (AI). AI-based systems enhance industries by automating repetitive tasks and improving overall efficiency. However, from a safety perspective, operating a system using AI without human interaction raises concerns regarding its reliability. Recent developments have made it imperative to establish a collaborative system between humans and AI, known as Intelligent Augmentation (IA). Industry 5.0 focuses on developing IA-based systems that facilitate collaboration between humans and AI. However, potential conflicts between humans and AI in controlling process plant operations pose a significant challenge in IA systems. Human-AI conflict in IA-based system operation can arise due to differences in observation, interpretation, and control action. Observation conflict may arise when humans and AI disagree with the observed data or information. Interpretation conflicts may occur due to differences in decision-making based on observed data, influenced by the learning ability of human intelligence (HI) and AI. Control action conflicts may arise when AI-driven control action differs from the human operator action. Conflicts between humans and AI may introduce additional risks to the IA-based system operation. Therefore, it is crucial to understand the concept of human-AI conflict and perform a detailed risk analysis before implementing a collaborative system. This paper aims to investigate the following: 1. Human and AI operations in process systems and the possible conflicts during the collaboration. 2. Formulate the concept of observation, interpretation, and action conflict in an IA-based system. 3. Provide a case study to identify the potential risk of human-AI conflict.

## 1. Introduction

As AI technology has advanced over the years, AI-based systems have become more capable of operating autonomously without human interaction. In the future, AI-based autonomous approaches will play a crucial role in controlling industrial plant operations (Arunthavanathan et al., 2023). However, developing AI automated controllers without human interaction raises concerns about the process safety operation. Moreover, in process plants, safety regulations such as the ISA 84 standards require the inclusion of a manual shutdown process by human operators in addition to automated controllers such as safety instrumented systems (SIS) (Wanasinghe et al., 2022).

In industrial plant operations, AI models are developed using available data or existing knowledge. Therefore, these models may not train with all scenarios in a dynamic industrial plant environment. However, ensuring industry safety requires a deep understanding of system dynamics and logical reasoning, especially when the process deviates from its normal operation (Díaz-Rodríguez et al., 2023). Furthermore, AI must adhere to safety regulations and standards when industries are ready to replace humans and operate with automated AI (Badri et al., 2018). Recent studies suggest that neuro-symbolic AI has the capability to train the AI using data (numerical, structural, textual, or graphical), knowledge graphs, logic statements, and multi-model data (Goel et al., 2024; Sheth et al., 2022). These types of AI tools combine traditional knowledge-based models and data-driven deep neural network models to enable complex reasoning tasks and handle diverse types of information effectively (Sheth et al., 2023). In the future, neuro-symbolic models will be a useful tool for developing efficient AI approaches in industrial plant operations. However, these types of AI models highly rely on comprehensive knowledge-based rules developed by humans

* Correspondence.
*E-mail address:* fikhan@tamu.edu (F. Khan).

https://doi.org/10.1016/j.dche.2024.100151
Received 25 February 2024; Received in revised form 4 April 2024; Accepted 5 April 2024
Available online 5 April 2024

and often struggle with unstructured or uncertain operations (Amador-Domínguez et al., 2024). Therefore, implementing automated AI using a model like neuro-symbolic AI to operate in high-risk environments also presents challenges and may not address unknown or unexpected situations in plant operations.

Therefore, developing AI without human intervention may struggle to deal with unknown or unexpected situations during plant operations (Wanasinghe et al., 2021). From an industrial application perspective, the current research in AI clearly recognizes the need to develop trustworthy and safe AI for safer and more efficient operations. Recently, the governments of the United States and the United Kingdom have established safety institutes to create safe and trustworthy AI (Blumenthal, 2024). The focus is on developing technical and regulatory solutions to ensure that AI operates safely and reliably in various applications. The primary goal of these institutes is to develop a socio-technical and human-centered infrastructure to understand the risks of advanced AI (Chamola et al., 2023). Therefore, in advancing the process industries with AI, there is a high possibility of collaborating AI with humans rather than enhancing autonomous operation by replacing humans with AI.

Intelligence augmentation (IA) approaches in process plants can be applied across plant operations, such as process control, safety instrumented systems, process monitoring, maintenance, and process risk mitigations (Chae et al., 2023). IA approaches such as Human in Loop, AI in Loop, and Human on Loop can be used to enhance process plant operations. However, one of the significant challenges in an IA system is when humans and AI collaboratively work and control the actions in a system, which may lead to conflicts and hazardous incidents. For instance, two Boeing 737 Max crashes in October 2018 and March 2019 killed 346 human lives. These incidents resulted in a disagreement between the pilot and the automated system. As a result, the pilot could not take the necessary action within the time frame to prevent the hazardous situation (Kraus, 2019).

From a process safety perspective, the Buncefield and Texas City incidents in 2005 serve as evidence of system failures and human error. The lack of system failures and poor situational awareness led to the largest accidents in process engineering. In the Buncefield explosion, the failure of the automated tank gauge (ATG) system and the independent high-level (IHL) system failed to automatically shut down the operation. On the other hand, the failure of alarms and the operator's unawareness of the situation led to the incident (Sam et al., 2011; Sam et al., 2009). Additionally, the failure in regular inspection and maintenance also did not educate the operators to understand the ATG/IHL device and system malfunctions (Sam et al., 2011). Similarly, in the Texas City incident, level indicator and alarm malfunctions, operator lack of knowledge, and unawareness led to the explosion (Isimite and Rubini, 2016).

Another notable example is that, in recent years, Tesla has developed a self-driving feature that operates without human interaction. In February 2023, over 360,000 vehicles were recalled due to a malfunction in their technology. The recall was initiated because of a software issue that caused the vehicles to disregard traffic signals, stop signs, and speed limits (Sheth et al., 2023). Additionally, according to an analysis by The Washington Post of NHTSA data, autopilot mode was engaged in 736 accidents since 2019 without human interaction, resulting in 17 fatalities (Amador-Domínguez et al., 2024).

This article presents a different perspective on the aforementioned incidents to demonstrate the efficiency of human-machine collaboration. It highlights the importance of IA system operations in future industrial advancements rather than solely relying on AI-based decisions and control action. The article also conceptualizes the types of conflicts that may arise between humans and AI in an intelligence-augmented system. Examples from Boeing and Tesla emphasize the need for industries to cautiously implement automated systems, ensuring that their operation aligns with human expectations. These incidents also highlight the need to educate the operators on the automated system and train them to take necessary actions in abnormal situations. Past incidents such as Buncefield and Texas City clearly demonstrate the importance of collaboration between machines and humans in process industries. These incidents also highlight that relying heavily on automated systems can lead to disastrous consequences due to human carelessness.

To expose the conflict between AI and human operation in process safety, this paper first conducts an operational study, which classifies and defines the formation of AI-human conflict in a process system. Furthermore, the paper investigates the impact of learning differences between AI and HI in an interpretation conflict. Finally, the paper uses a relevant case study to demonstrate the definitions of AI-HI conflict. This paper is organized as follows: Section 2 discusses the role of AI and HI in process system control operation and safety, Section 3 focuses on formulating the AI-HI conflict concept in the IA system, Section 4 details the learning differences between AI and Human, Section 5 details the AI-HI conflicts by using a case study and the last section concludes the study.

## 2. Human and artificial intelligence operations in intelligence augmented systems

In an IA system, evaluating the role of humans and AI presents a significant challenge and is crucial for risk assessment and maintaining operational safety. In process plant operations, human-machine involvement in the safety operation has become a standard practice over the years. In current plant operational setups, human operators and automated machines are operated independently and successively to ensure safety. If the basic process control system (BPCS) fails to operate the system, then the human operator is activated to take control. If the human operator also fails, then the automated SIS takes control. However, human operators rely heavily on system responses like active alarms or monitoring of human-machine interface (HMI) dashboards.

To effectively implement IA systems, it is crucial for humans and AI to collaborate and make decisions together. For instance, when developing a BPCS using AI systems, it is essential to involve human operators to monitor the real-time data and update the AI-based controller when an unknown abnormal situation occurs. This section conceptually studies the role of humans and AI in IA-based process plant operations to ensure process safety.

### *2.1. Enhancing safety operations in intelligent augmented-based process systems*

The implementation of Industry 5.0 technology transformation, including human-AI collaboration, is expected to bring significant changes to process industries, with a strong focus on improving system efficiency (Wanasinghe et al., 2022). However, these technological changes may create different potential hazards to the plant when applied to industrial usage. This section compares the traditional process plant with the conceptual IA-based process plant to study the potential technological changes and the involvement of humans and AI in the augmented system.

Fig. 1 illustrates the traditional process plant safety operations in accordance with ISA 84 standards. During plant safety operations, deviation in BPCS triggers an alarm when abnormal conditions are detected. In response, human operators promptly address the issue, either by initiating a shutdown procedure or by ensuring the continued safe operation of the plant (Arunthavanathan et al., 2021). Human operators also continuously monitor plant operations through HMI or workstations and take necessary actions based on the data observation. Beyond manual operation, a SIS is implemented to automatically shut down the process plant upon detecting unsafe conditions (Zhang et al., 2020). The SIS operates independently from the BPCS and uses different sensors, logic controllers, and parameter setpoints. Developing effective safety elements such as alarm systems and SIS requires a deep understanding of process safety principles, engineering expertise, and

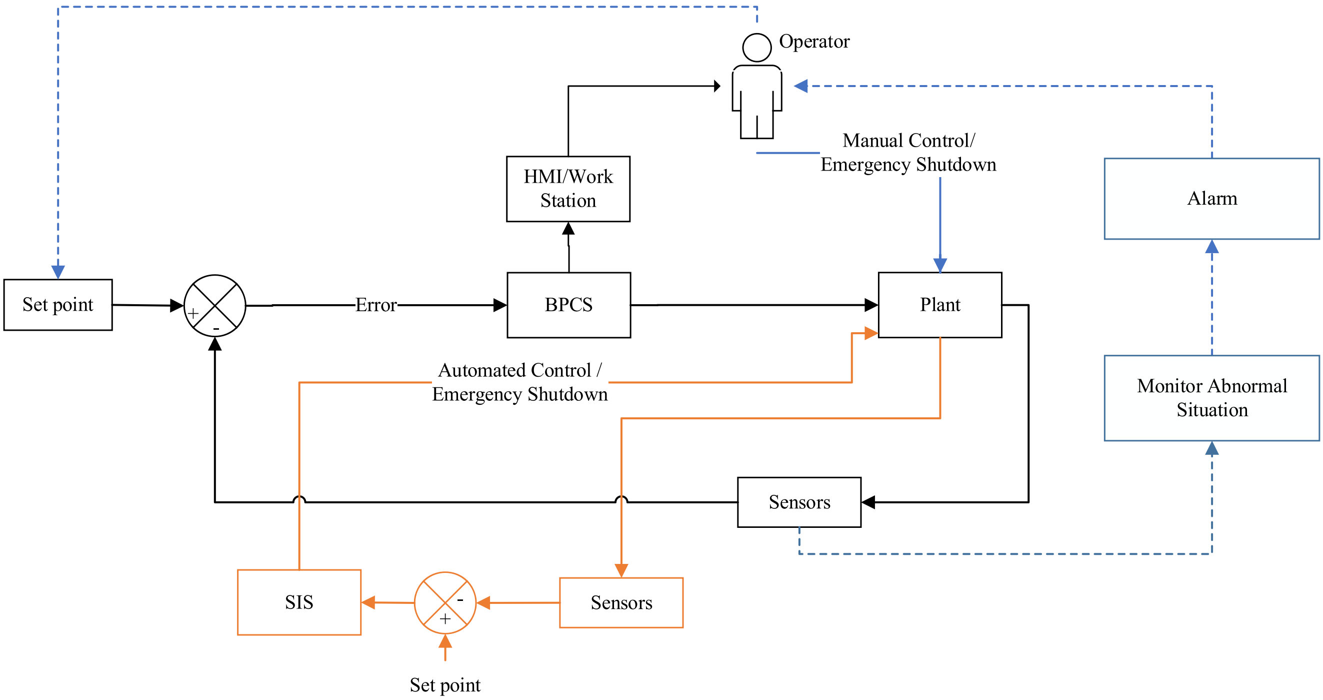


**Fig. 1.** Traditional Process Plant Safety Operation.

adherence to industry standards.

In industrial plant safety, it is crucial to identify the plant's abnormal behavior. Different technologies, such as model-based, knowledge-based, and data-driven methods, have been implemented over the years to monitor the system's abnormal operation (Alauddin et al., 2018). However, developing the most accurate models with minimal occurrence of false alarms or missed alarms is challenging (Hu and Yi, 2015). From the process safety perspective, it is necessary to design an appropriate alarm management system to prevent any missed alarms. Based on the process plant accident surveys, most of the incidents happen due to missed alarms (Wang et al., 2020). The main cause for this missed alarm is identified as a failure in the sensor module to monitor the abnormal action or malfunction in the alarms or actuators. System failure and human error were also identified as a main cause of plant-based accidents.

Fig. 2 shows the industry 5.0-based framework for developing an IA-based process plant. As shown in the framework, BPCS, SIS, and alarm management, including abnormal situation detection, can be enhanced using AI approaches. (Wen et al., 2022). However, due to the nature of the AI algorithm, the quality of data and the algorithm model parameters significantly influence the decision-making and process control. Therefore, to ensure reliable AI operation, human operators frequently

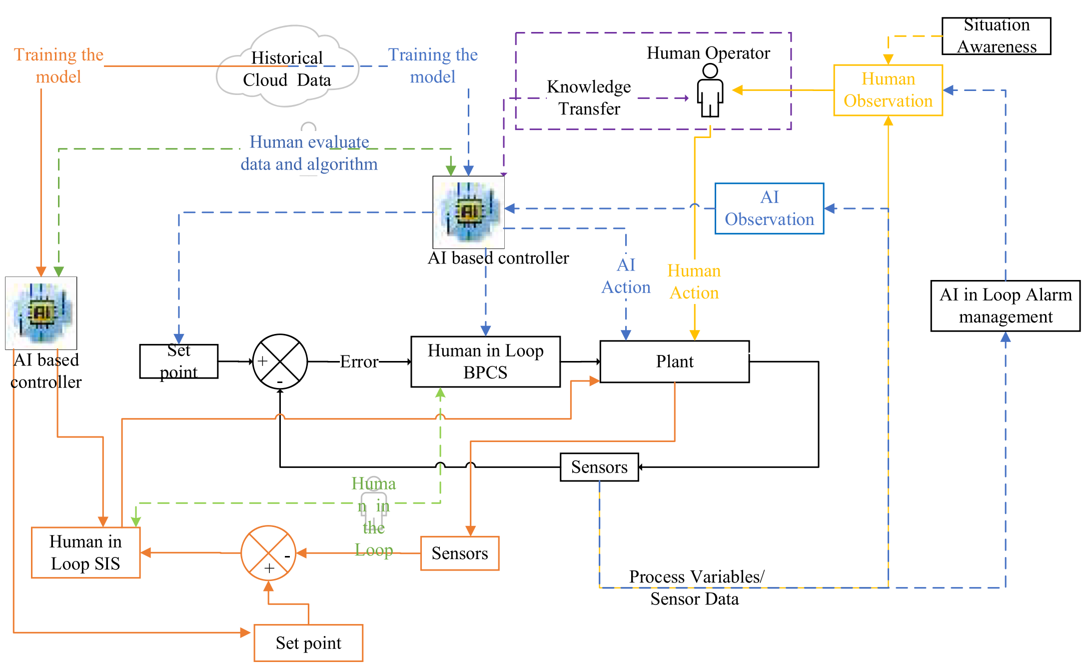


**Fig. 2.** Augmented Intelligence Based Safety Operation (Black Straight line – Basic Process Control, Yellow Straight Line –Manual Shutdown, Orange Straight Line – SIS, Blue Dot Line – AI in and Out, Green Dot Line – Human input, Purple Dot Line – Process Control Knowledge Transfer Between Human and AI).

monitor the quality of data and algorithms during real-time operations, especially in abnormal operational conditions. This approach is known as human-in-the-loop in IA-based systems, where the AI allows humans to provide direct feedback to continuously learn and adapt.

Moreover, in the IA-based process system, AI models and applications can assist human operators in decision-making. Due to the intellectual learning and faster response of the AI models, human operators are able to receive prompt feedback from AI (Chamola et al., 2023). This technology is known as AI in the loop in IA-based systems, where humans receive feedback and knowledge from AI-based systems during the decision-making process.

Therefore, in enhancing traditional process systems with the help of AI approaches, humans and AI collaborate in all the different stages of process plant operations. This ensures the system's efficiency, reliability, and safety operations while creating a learning environment between humans and AI. However, to develop AI in process safety operations apart from plant sensor data, it is also important to train AI with regulatory standards and ethics to ensure appropriate decision-making. However, increasing the sources of data and having AI gather knowledge from various resources may reduce the trustworthiness of AI. Moreover, data-driven AI models lack the ability to filter reliable and unreliable data, which makes them susceptible to misuse or training to operate in abnormal conditions.

However, as illustrated in Fig. 2, the presence of humans in the AI loop to make decisions or AI interruption in human decision-making creates a high potential for conflict in their decision-making process. For instance, in a tank operation, if a human operator decides to shut down the process due to overflow observation, false sensor information may cause disagreement with the AI system. If the control action relies more on the AI system rather than the human operator, it may lead to hazardous operations. The rest of the article details the AI-HI conflict when they collaboratively work in IA-based approaches, such as a human in the loop or AI in the loop.

## 3. Definitions and concepts in AI – HI conflict in process safety

In an AI-based operational process, HI is responsible for developing AI algorithms based on historical data. With real-time sensor data, AI continually observes the current operations and uses the knowledge gathered from the historical data to make control actions. On the other hand, the human operator monitors the real-time sensor data and manages the operation using the intelligence learned from plant operations, experience, and education. Therefore, when considering the overall collaborative control between AI and HI in process operations, conflicts may arise due to the observation, interpretation, and control actions of data. To address and prevent conflicts between AI and humans, it is essential to understand the definitions of observation, interpretation, and action conflicts. The study of the conflict between human operators and the AI control system in the process systems is described in Fig. 3.

In Fig. 3, $O_{AI}$ defines Artificial Intelligence observation, $O_H$ defines Human Intelligence observation, $I_{AI}$ defines AI interpretation, $I_H$ defines Human interpretation, $A_{AI}$ defines AI control action, and $A_H$ defines human control action. As illustrated, AI interprets the observed data from the process plant and takes the necessary action. On the other hand, humans observe the process of plant data and situational awareness to interpret the situation and take the necessary control action to operate the plant safely. Overall, in IA-based process operations, conflicts between AI and HI will be reflected in their control action. Whenever a conflict occurs, and the human operator performs correct decision-making, AI will be able to obtain this knowledge if the decision-making information is digitally available and reliable.

### 3.1. Observation conflict

**Definition 1.** Observation conflict is defined as the disagreement between the data sensed by the AI system and the human operator. Observation conflict may arise due to several reasons, such as human errors in monitoring the sensor data, incorrect or missing sensor data monitoring by the AI algorithms, and sensor faults. Moreover, if humans observe the situation awareness without reflecting in sensor data, it also creates an observation conflict between humans and AI.

$$\text{Observation conflict}: O_C = |O_H - O_{AI}| \tag{1}$$

As described in Fig. 3, $O_H$ described human observations on sensor data, which is reading data from the plant operation control room or direct sensor reading by the humans during the maintenance process. $O_{AI}$ described AI reading from the sensor data and $O_C$ described the conflict between human and AI observed data. To analytically describe the observation conflict, process system variables typically follow a Gaussian distribution with a mean value of μ and a variance of σ, having upper and lower operating control limits (Ge and Song, 2011). These boundaries may vary depending on various operational conditions and the noise level but should not exceed the 3σ bound between the control

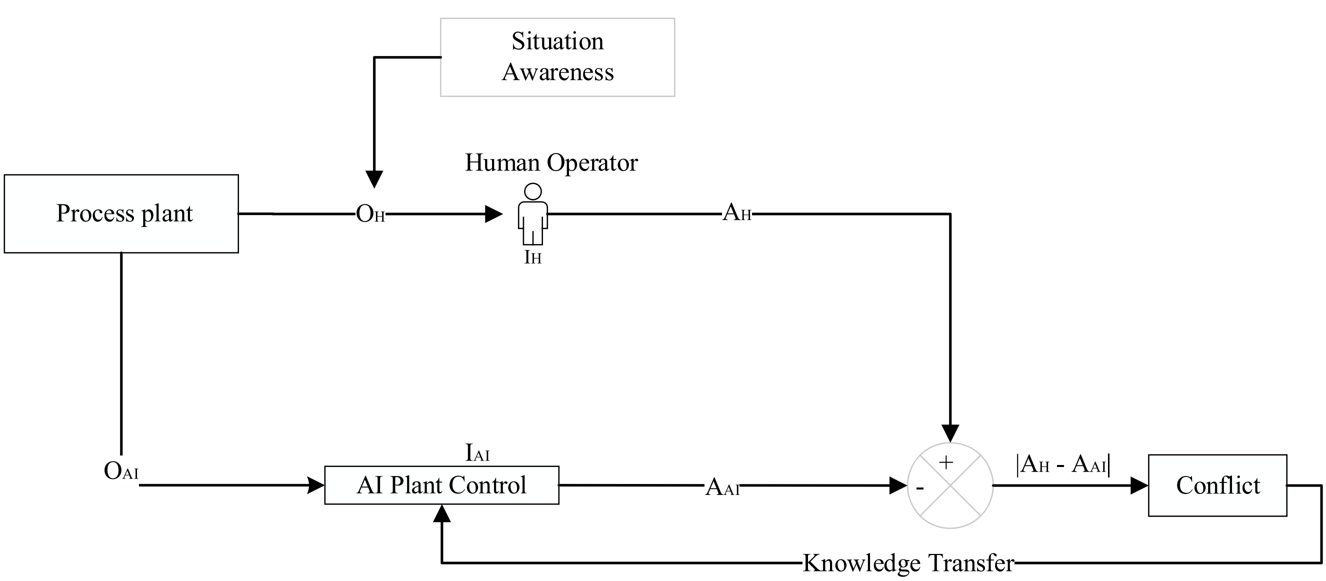


**Fig. 3.** AI-HI conflict in process.

limits. For instance, in a process system, the operational variable is set to function within $\mu \pm 3\sigma$ sigma condition. Under such circumstances, observation conflict may arise,

$$\text{Conflict Scenario 1} : O_{AI} \in \mu \pm 3\sigma, \text{and } O_H \notin \mu \pm 3\sigma \tag{2}$$

$$\text{Conflict Scenario 2} : O_{AI} \notin \mu \pm 3\sigma, \text{and } O_H \in \mu \pm 3\sigma \tag{3}$$

If both the human and AI observations ($O_{AI}$ and $O_H$) $\in \mu \pm 3\sigma$, then the system is considered to be operating in a normal condition. Conversely, if both human and AI observations, ($O_{AI}$ and $O_H$) $\notin \mu \pm 3\sigma$ then the system observed as an abnormal condition. These agreed observations between humans and AI lead to the system operating in a normal condition or initiating a shutdown. However, as shown in Eqs. (2) and 3, if one observation falls outside the range and disagrees with the other, it results in observation conflict arias.

Moreover, in certain situations, such as sensor fault scenarios, both humans and AI may observe incorrect data that falls outside the $3\sigma$. This can lead to both the human and AI making incorrect observations without any conflict. This indicates that even though both AI and humans agree on the operation control, there might be a possibility of incorrect control action.

### 3.2. Interpretation conflict

**Definition 2.** Conflicts between HI and AI can arise when they have differing interpretations of data, leading to divergent decision-making outcomes. Factors contributing to the interpretation conflicts include differences in observations, disparities in learning capabilities between AI and humans, and AI system failure or human errors.

$$\text{Interpretation Conflict} : I_C = |I_H - I_{AI}| \tag{4}$$

Modeling interpretation conflict is a major challenge due to the complexities involved in appropriately modeling HI. To address this issue, as illustrated in Fig. 3, a direct analysis of the control action conflict between HI and AI can be used. However, action conflict may also be impacted by actuator faults or human error in action. Moreover, conflict in data observation between AI and HI will directly impact the interpretation and decision-making. Therefore, if observation conflict arises, there is a high probability of interpretation conflict.

For instance, if a process variable follows a normal distribution and the data for normal operation conditions is bounded within the $\sigma$ boundary level, then the detailed description of observed data statistical evaluation, AI interpretation, and HI interpretation can be described as follows.

To statistically observe the interpretation of sensor data within binary decision-making, the sensor data can be described as,

$$P\left(Normal_{operation} \middle| O_S\right) = \theta_S\left(\frac{O_S - \mu}{\sigma}\right)$$

$$Decesion_{Statistical} = \begin{cases} P\left(Normal_{operation} \middle| O_S\right) > \alpha_S \; Normal\ Operation \\ \quad else\ Abnormal\ Operation \end{cases} \tag{5}$$

Where $O_S$ denotes as an observed data value, $\theta_S$ denote as a probability density function and $\alpha_S$ use as the threshold to classify as normal operation.

The Interpretation of AI to make a binary decision using AI observed data can be described as,

$$I_{AI} = f\ (O_{AI},\ \theta_{AI}) \tag{6}$$

$$\theta_{AI(t+1)} = \theta_{AI(t)} - \eta\nabla\text{L} \tag{7}$$

Where $O_{AI}$ denote observed variable, $\theta_{AI}$ denote model parameter, $f$ represent the activation function that maps the input features and AI decision making. $\eta$ denote learning rate, $\nabla$L denote loss function.

In a AI training mode, parameter $\theta_{AI}$ calculated using observed data, number of training samples, learning rate and loss function. The general relationship between model parameter loss function and learning rate is shown in Eq. (7). The number of iterations in learning will update the model parameter to high accuracy; however, in most of the cases, it may lead to the model overfitting. Also, running the large iteration to achieve better accuracy requires high computation time. In real-time running mode, the decision is based on the trained model Eq. (6), and the activation function is generally selected by human developers based on the targeting AI decisions. For instance, to make a normal or abnormal binary decision, a sigmoid activation function can be used.

This will show that AI decision-making completely depends on historical data and AI learning. AI learning highly relies on the optimization algorithm and loss function.

On the other hand, The Interpretation of HI to make a binary decision using HI-observed data can be described as,

$$I_{HI} = \sum_{hi=1}^{m} \omega_{hi} * f_{hi} \tag{8}$$

Where $f_{hi}$ denotes different factors involved in human decision-making, including observation $O_{AI}$, perception, memory, intellectual ability, and emotion. $\omega_{hi}$ denotes weights assigned to each feature. For instance, As shown in Eq. (8), in normal operational conditions, observation and perception may have a high weight in monitoring the system condition without any impactful decision-making. However, if the system deviates from its normal operating conditions, intellectual ability, experience, memory, and emotions are given a high weight in making impactful and effective decisions.

As shown in Eqs. (5), 6, and 8, binary decision-making using observed data differs in the way statistical models, AI models, and HI interpret the data. The interpretation of statistical models reveals that noise or disturbance in observed data may affect the decision. In AI models, the quality of the data model accuracy and the presence of noise affect the decision. In HI, data observation is considered along with human factors ($f_{hi}$) that impact the decision-making process.

Conclusively, this highlights that during the process of plant operation, AI interpretation heavily relies on observed data, compared to human interpretation. In recent years, data used to train AI models has been limited to sensor data. However, as digitalization advances, more data sources such as reports, experiences, and emotions can be transformed into digital forms to train AI models. This may enhance AI's learning ability and make decisions closer to HI. This technological advancement raises questions about the superiority of AI over HI, which could also lead to conflicts in interpretation. The learning ability of HI and AI causes this conflict to be more detailed in Section 4.

### 3.3. Action conflict

**Definition 3.** The Action conflict defines the difference between AI and HI operation control action. The factors that are contributes to the action conflict includes differences in HI and A decision making, misleading the action operations, and actuator faults.

$$\text{Action Conflict} : A_C = |A_H - A_{AI}| \tag{9}$$

The modeling of an action conflict between an AI system and HI is highly linked with the interpretation conflict. Therefore, control action is described as,

$$A_{AI} = g_{AI}.I_{AI} \tag{10}$$

$$A_{HI} = g_{HI}.I_{HI} \tag{11}$$

Where $g_{AI}$ and $g_{HI}$ denote interpretation to action conflict function of AI system and HI. $g_{AI}$ highly relies on automated actuator failure, $g_{HI}$ relies on actuator failure as well as human mislead the control action.

When assessing the conflict between AI and HI in the context of process system operation, action conflicts will be the direct cause of incidents. However, the root cause for the action conflict will be the observation and interpretation conflicts. Therefore, before developing IA-based process system plant operations, it is crucial to analyze the risk associated with the type of conflicts. Also, when an incident occurs, reassessing the scenario based on the action, interpretation, and observation may lead to the actual cause of the incident.

## 4. Learning ability of AI and HI that can lead to the interpretation conflict

In the collaboration between humans and AI within an IA-based system, maintaining comparable levels of learning ability between both entities is crucial. Recent advancements in AI have not yet endowed it with the technical capability to surpass human learning ability. However, the learning ability of AI is constrained, it may still learn faster than humans within its limited resources. Therefore, human learning is characterized by slow, broad, and steady progress, while AI learning is faster yet confined.

### *4.1. Comparing AI and HI learning ability*

The term "Artificial Intelligence" was first coined at the Dartmouth Conference in 1956 (Jarrahi et al., 2022). After the advancements of AI in recent decades, the prominence and excitement surrounding AI have led to misconceptions about its nature and complexity. Also, recent findings suggest that the idea of AI replacing human intelligence is often overhyped in terms of operational safety (Joshi, 2024). According to a recent report by IBM, there is a significant gap between the intentions and actions of applying AI in an industrial environment (Kerr et al., 2020). The report suggests that AI systems are mainly focused on completing specific tasks, which are defined by human intelligence (Maadi et al., 2021). Also, with the recent development of ethical arguments from the safety centers of the United States and the United Kingdom, it is currently believed that AI should be bound by HI for safety requirements (Jones, 2023).

A collaborative human-AI augmented framework creates a safer environment than relying solely on manual or AI operations. Therefore, comparing HI operations and AI operations in a system provides a better context to understand the real industrial needs. Applying AI algorithms in systems control and safety operations provides consistent and accurate solutions. Also, AI is capable of performing intellectual tasks, such as learning and analyzing data faster than human intelligence (Koivisto and Grassini, 2023). However, to execute an efficient operation, AI requires high-quality data. On the other hand, advancing beyond those intellectual tasks requires holistic and contextual thinking, which remains a core capability of HI. Humans are experts in using their past experiences to solve new problems and handle unfamiliar situations (Schoenherr et al., 2023). In addition to intellectual solutions, emotional decisions and decisions based on real-time situation awareness can also be helpful in an abnormal situation. This is another core aspect of human intelligence, as humans have the ability to control harmful situations using real-time situational awareness and emotions.

Intelligence is not limited to data processing, analysis, learning, and applying to solve problems. It requires more than an intellectual understanding of the provided information (Hassani et al., 2020). Conceptually, there are no physical laws preventing the development of AI systems that are more intelligent and powerful than the human brain (Montag et al., 2024). However, this improvement can be applicable to AI intellectual ability. As the cognitive level of AI increases and it gains the ability to solve problems using knowledge from various domains, the AI system may make decisions beyond human expectations. However, decision-making based on emotional and real-time situation awareness remains a challenge. This could result in conflicts between AI and HI decision-making and create risks in collaborative environments.

Therefore, AI learning capabilities depend on advancing technology and available data. AI systems learn from data and improve their performance over time by advancing the algorithms. On the other hand, HI involves learning through experience, education, and social interactions (Korteling et al., 2021). This will allow humans to adapt to the new situations and acquire knowledge (Dong et al., 2020).

### *4.2. AI singularity and future advancement prediction*

Enhanced AI technologies and advancements in digitalization are expected to elevate the intellectual ability of AI in the future. The learning capability of AI exponentially increase in past decades, driven by advancements in deep learning technologies, digitalization and computation power AI learning abilities keep enhancing. However, this increase learning ability enhance the intellectual level of AI while still lacking in other human intelligence factors such as emotions, learning from new environment.

Fig. 4 shows the evolution of learning models over the decades and the advancements in AI intellectual and AI singularity. In the early era, rule-based methods were introduced to develop and implement process control systems. These systems included critical components such as BCPS and SIS, which are essential to maintain process control operation and safety (Koivisto and Grassini, 2023). In the late 90 s, machine-learning approaches were introduced to improve process safety by detecting equipment failures and abnormalities in process operations (Joshi, 2024). Over the past decade, deep learning models have been proposed to analyze sensor data and monitor real-time process systems (Maadi et al., 2021).

In recent years, cognitive AI and time-dependent deep learning models have been developed to detect anomalies, assess the remaining useful life of the system, and create predictive maintenance models for ensuring process plant safety (Kraus, 2019). As per the concept of AI singularity, it is predicted that hypothetically, in 2040, AI's intellectual ability will surpass that of humans (Arunthavanathan et al., 2021). Therefore, in repeated tasks or logical problem-solving, AI decision-making will be more accurate than that of a human.

## 5. Risk in IA: AI-HI conflict a case study

This section formalizes the risks associated with collaborative decision-making between AI and HI in operations. Fault tree analysis is used to identify potential system failures, errors, and hazardous operations that may lead to conflicts between AI and HI. This section also provides a detailed analysis of the Boeing flight crash to help to understand the concept of AI-HI in a system operation, as well as its root cause.

As illustrated in Fig. 5 and Table 1, AI observation failure can be caused by sensor faults, cyber-attacks, and poor data quality. Similarly, HI observation failure can result from sensor faults and poor data quality, as well as human reading errors. Incorrectly observing data during an operation can lead to conflicting observations between AI and humans. This conflict, along with AI failures, human error, and other human factors, may lead to differences in AI-HI decision-making and result in interpretation conflicts. Interpretation conflict may lead to a control action conflict if AI or HI incorrectly operate the system or actuator. Failure to respond to the control action may also lead to AI-HI conflict.

### *5.1. Case study: boeing flight crash*

To demonstrate the potential for conflict between AI and human intelligence in the real world, this article uses the case study of the Boeing flight crashes. Although the case study does not directly point to the conflict as the root cause of the accidents, it is a perfect example to illustrate the disaster that could occur due to an AI-HI conflict.

The Airbus A320 features a new, efficient geared turbofan engine

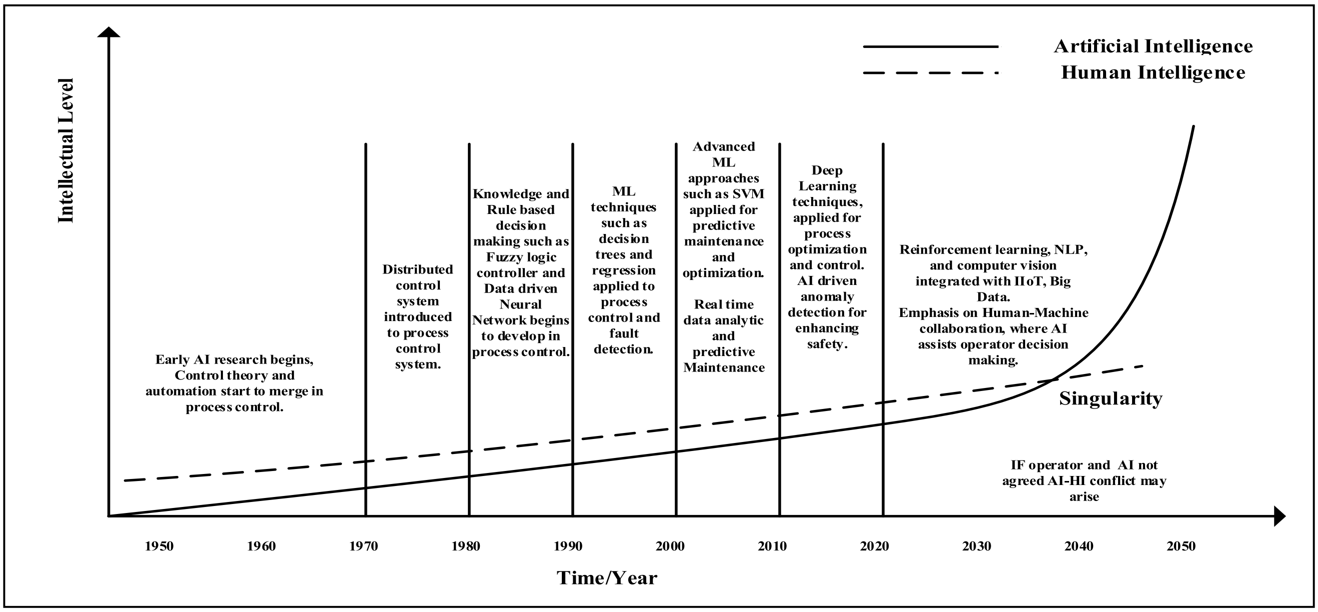


**Fig. 4.** HI and AI Intellectual Level over the years in process application (adapted from (Kurzweil, 2014)]).

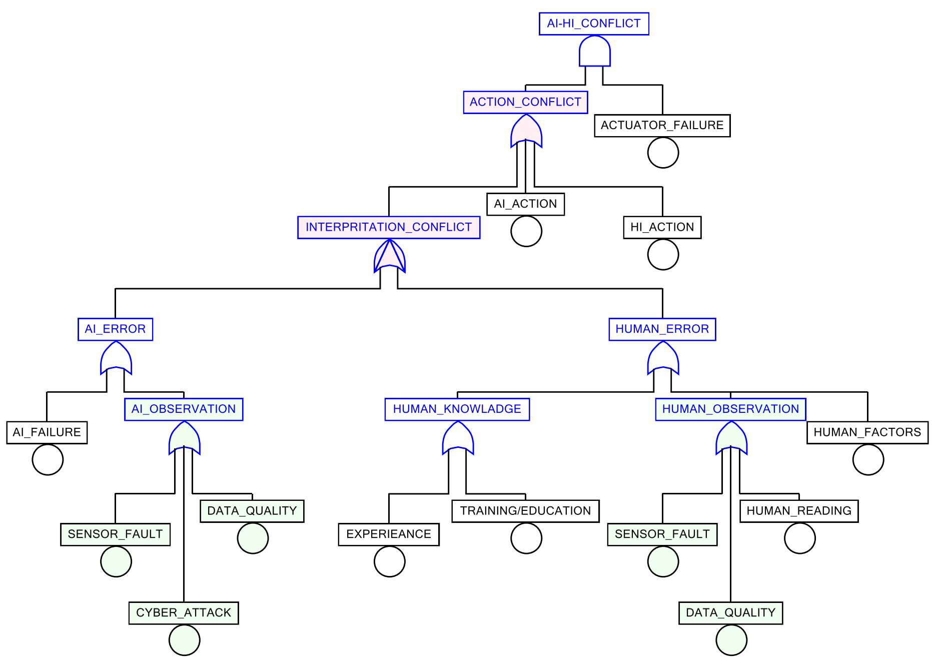


**Fig. 5.** Identified failures and risks that may lead to the AI-HI conflict.

with minimal impact on the aircraft's design and pilot training. Boeing's response with the 737 Max aimed to compete with Airbus but faced challenges due to the larger engine size (Kraus, 2019). To address this, Boeing adjusted the engine placement, inadvertently causing potential safety issues with the aircraft's nose pitch. To mitigate risks, Boeing implemented the Maneuvering Characteristic Augmentation System (MCAS) to monitor and adjust the aircraft's pitch as needed (Naor et al., 2020). Boeing claimed that the design of the 737 Max, with new fuel engines, did not change much, except for the engine placement. However, the engine placement caused issues as the nose tended to point too far upward during takeoff (Travis, 2019). In 2018, several pilots complained to the federal government that the aircraft was suddenly nosing down without any reason. Furthermore, pilots only received a two-hour iPad course before flying the 737 Max for the first time, and the training material did not mention the MCAS software.

There were two major accidents involving the MCAS system - one in Lion Air Flight 610 and the other in Ethiopian Airlines Flight 302. The first incident occurred in 2018 during the Lion Air 610 flight, where the pilot failed to deactivate the electric trim tab stabilizer and was only able to control the manual operation (Sterman and Quinn, 2023). In the Ethiopian Airlines incident that occurred in 2019, the pilots were able to recognize the need to disable the system, but not in time, as the aircraft had gained too much speed (Herkert et al., 2020). Both instances highlight the fact that pilots cannot be solely relied upon to overcome issues raised by MCAS. Additionally, in both incidents, the pilots identified that the flight was out of control and that the MCAS system kept pulling down the nose and taking the plane toward the ground (Sterman and Quinn, 2023; Herkert et al., 2020).

**Table 1**
AI-HI Conflict Fault Tree Events and their Description

| Fault Tree Event | Description | Fault Tree Event | Description |
|---|---|---|---|
| AI-HI_CONFLICT | Human-AI conflict led to the incident. | DATA_QUALITY | Noisy or missing data points in the data. |
| ACTION_CONFLICT | Conflict in action between human operator and AI. | HUMAN_ERROR | Humans make wrong decision |
| INTERPRITATION_CONFLICT | Conflict in Human and AI Interpretation/Decision-making. | HUMAN_KNOWLADGE | Lack of human knowledge to deal with the situations |
| AI_ACTION | AI fails to take action based on the decision. | EXPERIENCE | Lack of Experience to Handle the Situation |
| HI_ACTION | HI fails to take action based on the decision-making. | TRAINING/ EDUCATION | Lack of Training/Education |
| AI_ERROR | AI makes wrong decisions. | HUMAN_OBSERVATION | Human Observation Error |
| AI_FAILURE | AI system failure due to a Hardware/Software issue. | HUMAN_READING | Error in human reading |
| AI_OBSERVATION | AI observation Error. | HUMAN_FACTORS | Human Intellectual ability, Memory, perception, and emotion. |
| SENSOR_FAULT | Fault/Failure in Sensor. | ACTUATOR_FAILURE | Failure in Electromechanical actuators or manual valves |
| CYBER_ATTACK | Cyber-attack and inject false data. | | |

In the fault tree analysis shown in Fig. 6 and Table 2, it is evaluated that the failure of the AoA sensor leads to incorrect decision-making by the MCAS system. This results in the system taking the wrong action of nosing down the aircraft. Meanwhile, the pilots are monitoring the situation and decide to pull up the nose to stabilize the flight. However, this action conflict between the pilot and the system could not be resolved in a time, and the pilot did not have enough time to correct the autonomous MCAS decision-making, ultimately resulting in the aircraft crash.

Moreover, detailed studies indicate that pilots heavily relied on the MCAS decision-making system. Unfortunately, due to a lack of training and insufficient information, they failed to handle the situation effectively. This highlights that pilots' interpretation based on their experience and available information realized the situation by disagreeing with the MCAS decision-making and control action. However, the pilot was unable to take control due to a lack of knowledge on how to disconnect the MCAS operation and take manual control within a specific time frame.

This case study clearly demonstrates that when humans and machines collaboratively work, conflict in control action may arise in an abnormal situation. If a proper mechanism is offered to identify the situation and verify the conflicting observations and interpretations within the time frame, it may help to prevent hazardous situations.

## 6. Conclusion

This article conceptualizes and discusses the conflicts between human intelligence and artificial intelligence from the perspective of intelligence-augmented systems for processing safe operations. It investigates the observation, interpretation, and action of conflicts that can arise in process plant control operations and provides clear definitions for each type. Observation conflict may depend on conflicts that arise in human observation, including situation awareness with AI observation, which is purely based on sensor data. Interpretation conflict may arise due to the conflict in observation and learning ability of individuals. The focus is mainly on action conflicts, which are the most evident and outlines conflict to evaluate the associated risks. The article also highlights the importance of considering process control faults in assessing risk due to the AI-HI conflict and investigates the cause of risk approaches that can be applied. Finally, a case study describes the different conflicts and how they lead to the accident scenario.

## CRediT authorship contribution statement

**Rajeevan Arunthavanathan:** Conceptualization, Formal analysis, Writing – original draft, Data curation, Methodology, Validation. **Zaman Sajid:** Data curation, Formal analysis, Investigation, Validation, Writing – review & editing. **Faisal Khan:** Conceptualization, Formal

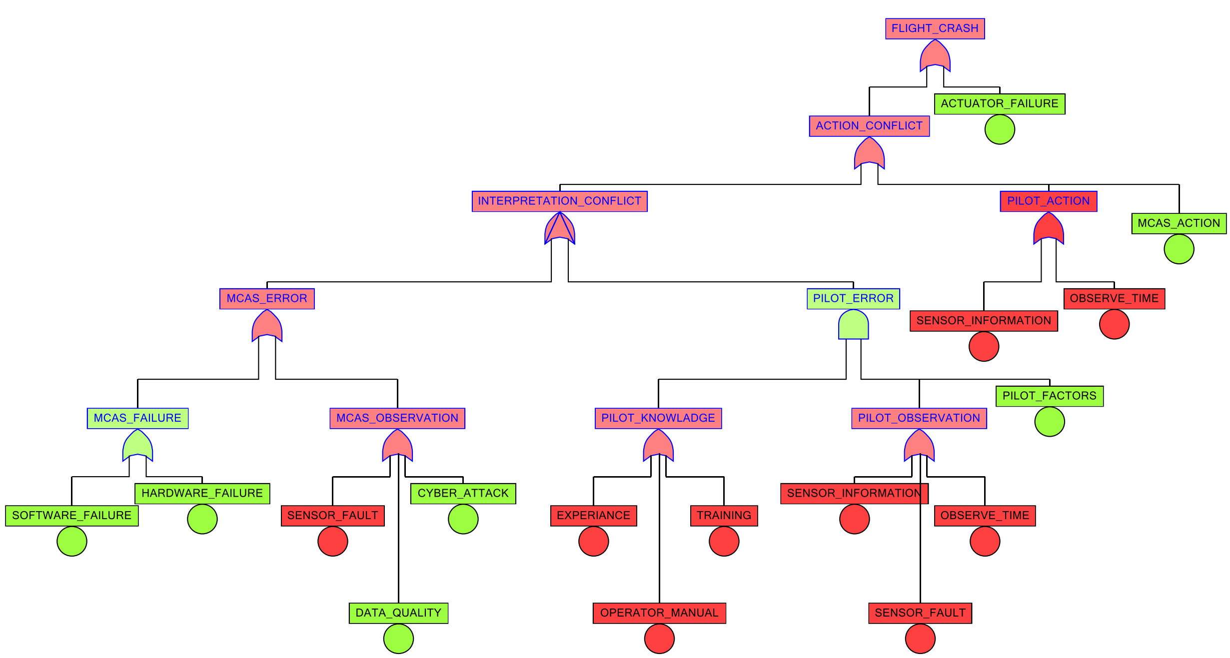


**Fig. 6.** Boing 737 Flight Accident Scenario.

**Table 2**
Boing 737 Flight Accident Scenario Fault Event and Description.

| Fault Tree Event | Description | Fault Tree Event | Description |
|---|---|---|---|
| FLIGHT_CRASH | Boeing 737–8 Incident | CYBER_ATTACK | Cyberattack to change the data or process set values. |
| ACTION_CONFLICT | Conflict in Nosel Action (Nose-(Down/Up)) | PILOT_ERROR | Pilot observation Error |
| ACTUATOR_FAILURE | Stabilizer Failure | PILOT_KNOWLADGE | |
| INTERPRETATION_CONFLICT | The conflict between MCAS - Human to operate nose | EXPERIENCE | The pilot did not have a similar experience with disconnecting the MCAS operation. |
| MCAS_ERROR | Error in MCAS Decision-Making | OPERATOR_MANUAL | The Operation Manual did not include the MCAS operation |
| MCAS_FAILURE | MCAS Failure due to software/hardware | TRAINING | Lack of training in MCAS. |
| MCAS_OBSERVATION | MCAS Observation Error | PILOT_OBSERVATION | |
| SOFTWARE_FAILURE | MCAS Software Failure | SENSOR_INFORMATION | The pilot was unaware of the AoA sensor data and MCAS installation. |
| HARDWARE_FAILURE | MCAS Hardware Failure | OBSERVE_TIME | Situation observed after manageable control time period. |
| SENSOR_FAULT | Fault in AoA sensor | PILOT_ACTION | Pilots fail to operate within the Time Frame |
| DATA_QUALITY | The quality of the data includes noise level and other factors | MCAS_ACTION | The system fails to act |

analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Efstratios Pistikopoulos:** Conceptualization, Formal analysis, Supervision, Validation, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The authors thankfully acknowledge the financial support provided by the National Science Foundation through project RETRO – Toward Safe and Smart Operations via Real-time Risk-based Optimization and the Mary Kay O'Connor Process Safety Centre at Texas A&M University, USA.

## References

Alauddin, M., Khan, F., Imtiaz, S., Ahmed, S., 2018. A bibliometric review and analysis of data-driven fault detection and diagnosis methods for process systems. Ind. Eng. Chem. Res. 57 (32), 10719–10735. https://doi.org/10.1021/ACS.IECR.8B00936/ASSET/IMAGES/LARGE/IE-2018-00936N_0017.JPEG. Aug.

Amador-Domínguez, E., Serrano, E., Manrique, D., 2024. Neurosymbolic system profiling: a template-based approach. Knowl.-Based Syst. 287, 111441. https://doi.org/10.1016/j.knosys.2024.111441.

Arunthavanathan, R., Khan, F., Ahmed, S., Imtiaz, S., 2021. An analysis of process fault diagnosis methods from safety perspectives. Comput. Chem. Eng. 145, 107197 https://doi.org/10.1016/j.compchemeng.2020.107197.

Arunthavanathan, R., Ahmed, S., Khan, F., Imtiaz, S., 2023. Machine Learning for Process Fault Detection and Diagnosis. In: Machine Learning in Chemical Safety and Health: Fundamentals with Applications, pp. 113–137. https://doi.org/10.1002/9781119817512.ch6.

Badri, A., Boudreau-Trudel, B., Souissi, A.S.. Occupational health and safety in the industry 4.0 era: A cause for major concern? https://doi.org/10.1016/j.ssci.2018.06.012.

Blumenthal, D., 2024. The U.S. president's executive order on artificial intelligence. NEJM AI 1 (2). https://doi.org/10.1056/AIPC2300296. Jan.

Chae, J., Lee, S., Jang, J., Hong, S., Park, K.J., 2023. A survey and perspective on industrial cyber-physical systems (ICPS): from ICPS to AI-augmented ICPS. IEEE Transact. Industrial Cyber-Phys. Syst. 1, 257–272. https://doi.org/10.1109/TICPS.2023.3323600. Oct.

Chamola, V., Hassija, V., Sulthana, A.R., Ghosh, D., Dhingra, D., Sikdar, B., 2023. A review of trustworthy and explainable artificial intelligence (XAI). IEEe Access. 11, 78994–79015. https://doi.org/10.1109/ACCESS.2023.3294569.

Díaz-Rodríguez, N., Ser, J.Del, Coeckelbergh, M., López De Prado, M., Herrera-Viedma, E., Herrera, F., 2023. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Inf. Fusion 99, 101896. https://doi.org/10.1016/j.inffus.2023.101896.

Dong, Y., Hou, J., Zhang, N., Zhang, M., 2020. Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. Complexity. 2020 https://doi.org/10.1155/2020/1680845.

Ge, Z., Song, Z., 2011. A distribution-free method for process monitoring. Expert Syst. Appl. 38, 9821–9829. https://doi.org/10.1016/j.eswa.2011.02.048.

Goel, S., Lymperopoulos, P., Thielstrom, R., Krause, E., Feeney, P., Lorang, P., Schneider, S., Wei, Y., Kildebeck, E., Goss, S., Hughes, M.C., Liu, L., Sinapov, J., Scheutz, M., 2024. A neurosymbolic cognitive architecture framework for handling novelties in open worlds. Artif. Intell. 331, 104111. https://doi.org/10.1016/J.ARTINT.2024.104111.

Hassani, H., Silva, E.S., Unger, S., Tajmazinani, M., Feely, S., 2020. Mac, Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? AI 1, 143–155. https://doi.org/10.3390/ai1020008.

Herkert, J., Borenstein, J., Miller, K., 2020. The Boeing 737 MAX: lessons for Engineering Ethics. Sci. Eng. Ethics 26 (6), 2957–2974. https://doi.org/10.1007/S11948-020-00252-Y/FIGURES/2. Dec.

Hu, J., Yi, Y., 2015. A two-level intelligent alarm management framework for process safety. Saf. Sci. 82, 432–444. https://doi.org/10.1016/j.ssci.2015.10.005.

Isimite, J., Rubini, P., 2016. A dynamic HAZOP case study using the Texas City refinery explosion. J. Loss Prev. Process Ind. 40, 496–501. https://doi.org/10.1016/j.jlp.2016.01.025.

Jarrahi, M.H., Lutz, C., Newlands, G., 2022. Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. Big Data Soc. 9.

Jones, N., 2023. The world's week on AI safety: powerful computing efforts launched to boost research. Nature 623 (7986), 229–230. https://doi.org/10.1038/D41586-023-03472-X. Nov.

Joshi, A., 2024. Humans in intelligent environment: near future. Artificial Intelligence. Human Evolut. 147–202. https://doi.org/10.1007/978-1-4842-9807-7_7.

Kerr, A., Barry, M., Kelleher, J.D., 2020. Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. Big Data Soc. 7.

Koivisto, M., Grassini, S., 2023. Best humans still outperform artificial intelligence in a creative divergent thinking task. Scientific Reports 13, 13601. https://doi.org/10.1038/s41598-023-40858-3.

Korteling, J.E.(Hans), van de Boer-Visschedijk, G.C., Blankendaal, R.A.M., Boonekamp, R.C., Eikelboom, A.R., 2021. Human- versus artificial intelligence. Front. Artif. Intell. 4, 622364 https://doi.org/10.3389/FRAI.2021.622364/BIBTEX. Mar.

Kraus, R., 2019. "Aggressive and riskier" A.I. may have caused Boeing 737 Max crashes. URL https://mashable.com/article/boeing-737-max-aggressive-risky-ai (accessed 3.2.24).

Kurzweil, R., 2014. The singularity is near. EthicsEmerg. Techn. 393–406. https://doi.org/10.1057/9781137349088_26.

Maadi, M., Khorshidi, H.A., Aickelin, U., 2021. A Review on Human-AI Interaction in Machine Learning and Insights for Medical Applications. Int. J. Environ. Res. Public Health 18 (4), 1–21. https://doi.org/10.3390/IJERPH18042121. Feb.

Montag, C., Ali, R., Davis, K.L., 2024. Affective neuroscience theory and attitudes towards artificial intelligence. AI Soc. https://doi.org/10.1007/s00146-023-01841-8.

Naor, M., Adler, N., Pinto, G.D., Dumanis, A., 2020. Psychological safety in aviation new product development teams: case study of 737 MAX airplane. Sustainability 12. https://doi.org/10.3390/su12218994.

Sam, M., Mary, M., O'connor, K., 2009. Proceedings of the Hazards XXI. Institution of Chemical Engineers, pp. 662–673.

Sam, M., Mary, M., O'connor, K., 2011. The buncefield explosion and fire-lessons learned. Process Saf. Prog. 30, 138–142. https://doi.org/10.1002/prs.10444.

Schoenherr, J.R., Abbas, R., Michael, K., Rivas, P., Anderson, T.D., 2023. Designing AI using a human-centered approach: explainability and accuracy toward trustworthiness. IEEe Trans. Technol. Soc. 4 (1), 9–23. https://doi.org/10.1109/TTS.2023.3257627. Mar.

Sheth, A., Gaur, M., Roy, K., Venkataraman, R., Khandelwal, V., 2022. Process Knowledge-Infused AI: Toward User-Level Explainability, Interpretability, and Safety. IEEE Internet Comput. 26, 76–84. https://doi.org/10.1109/MIC.2022.3182349.

Sheth, A., Roy, K., Gaur, M., 2023. Neurosymbolic Artificial Intelligence (Why, What, and How). IEEe Intell. Syst. 38 (3), 56–62. https://doi.org/10.1109/MIS.2023.3268724. May.

Sterman, D.J., Quinn, J., 2023. BOEING'S 737 MAX 8 DISASTERS, mitsloan.J. Sterman and J. Quinn, "BOEING'S 737 MAX 8 DISASTERS".

Travis, G., 2019. How the Boeing 737 Max Disaster Looks to a Software Developer. IEEE Spectr. URL https://spectrum.ieee.org/how-the-boeing-737-max-disaster-looks-to-a-software-developer (accessed 2.18.24).

Wanasinghe, T.R., Trinh, T., Nguyen, T., Gosine, R.G., James, L.A., Warrian, P.J., 2021. Human centric digital transformation and operator 4.0 for the oil and gas industry. IEEe Access. 9, 113270–113291. https://doi.org/10.1109/ACCESS.2021.3103680.

Wanasinghe, T.R., Galagedarage Don, M., Arunthavanathan, R., Gosine, R.G., 2022. Industry 4.0 based process data analytics platform. Methods in Chemical Process Safety 6, 101–137. https://doi.org/10.1016/BS.MCPS.2022.04.008.

Wang, J., Fu, G., Yan, M., 2020. Investigation and analysis of a hazardous chemical accident in the process industry: triggers, roots, and lessons learned. Processes 8, 477. https://doi.org/10.3390/PR8040477.

Wen, H., Khan, F., Amin, M.T., Halim, S.Z., 2022. Myths and misconceptions of data-driven methods: applications to process safety analysis. Comput. Chem. Eng. 158 https://doi.org/10.1016/j.compchemeng.2021.107639. Feb.

Zhang, A., Zhang, T., Barros, A., Liu, Y., 2020. Optimization of maintenances following proof tests for the final element of a safety-instrumented system. Reliab. Eng. Syst. Saf. 196, 106779 https://doi.org/10.1016/J.RESS.2019.106779. Apr.