Artificial Intelligence (AI) in Occupational Safety and Health (OSH) 101: What do OSH professionals need to know?

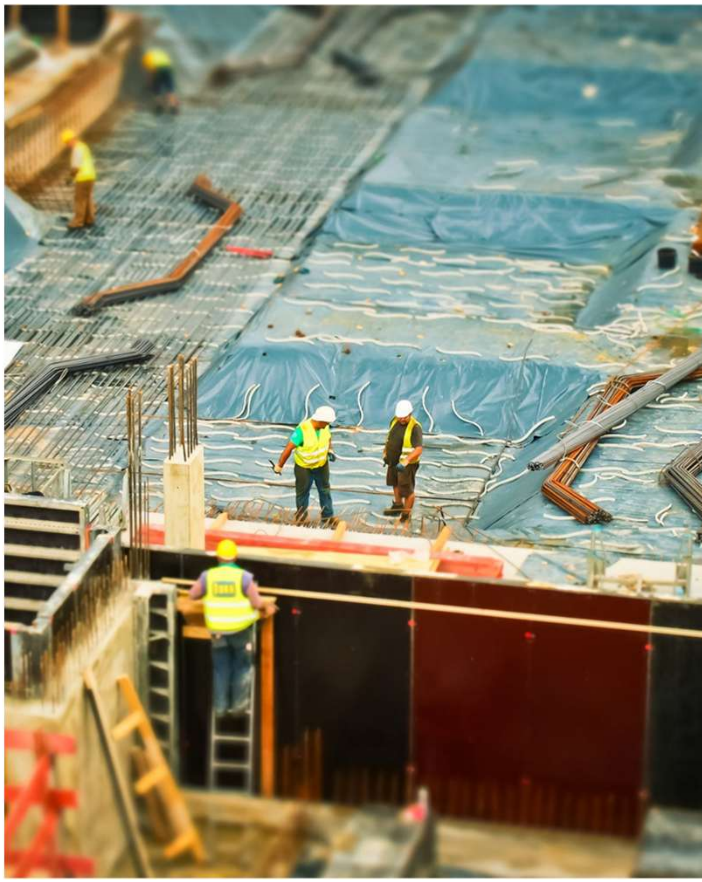
Houshang Darabi, Ph.D.
University of Illinois Chicago
March 12, 2025

Presentation Outline

1. AI systems and where AI can be effective
2. Different types of AI applications in OSH
3. Developing AI systems: Project Pipeline and Personnel
4. What is Machine Learning (ML)?
5. What are the ethical considerations in AI systems?

1. AI Systems and Where AI Can Be Effective

A Natural Intelligence Example: Construction System Loop



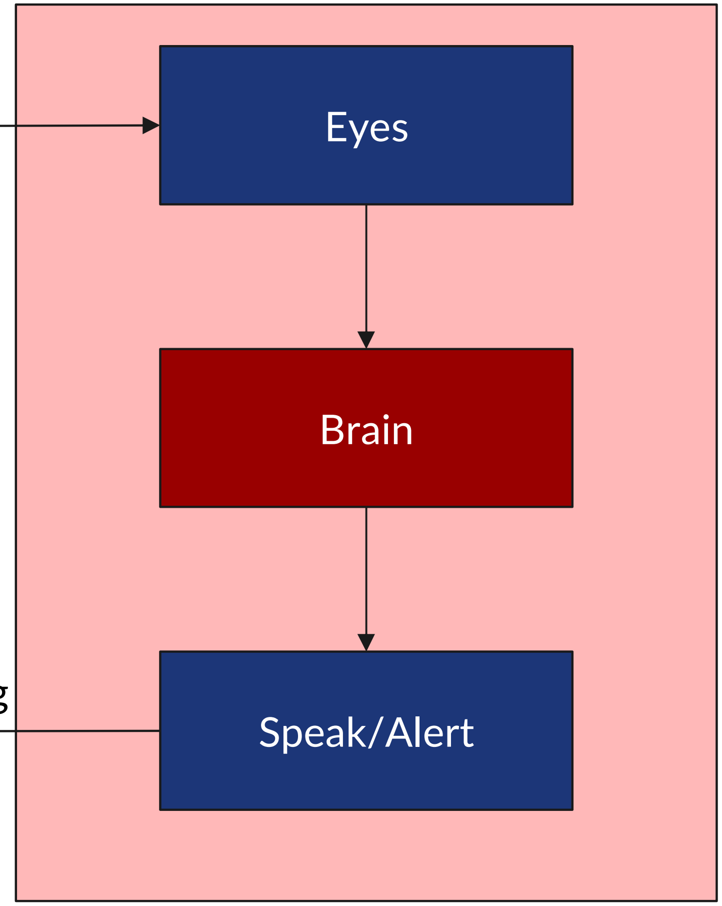
Environment

Percepts

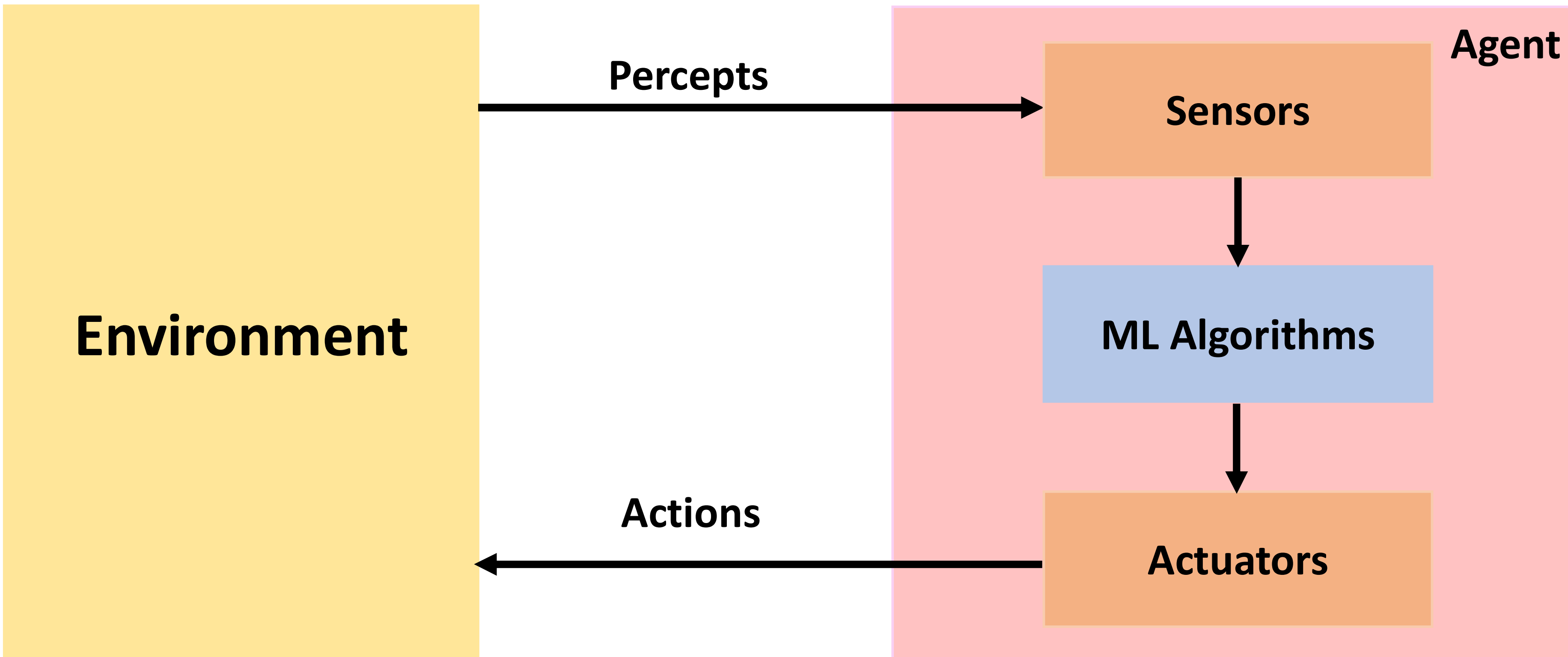
Construction worker, height, close to ledge

Intervention: vocally alert/warn workers they are close to falling

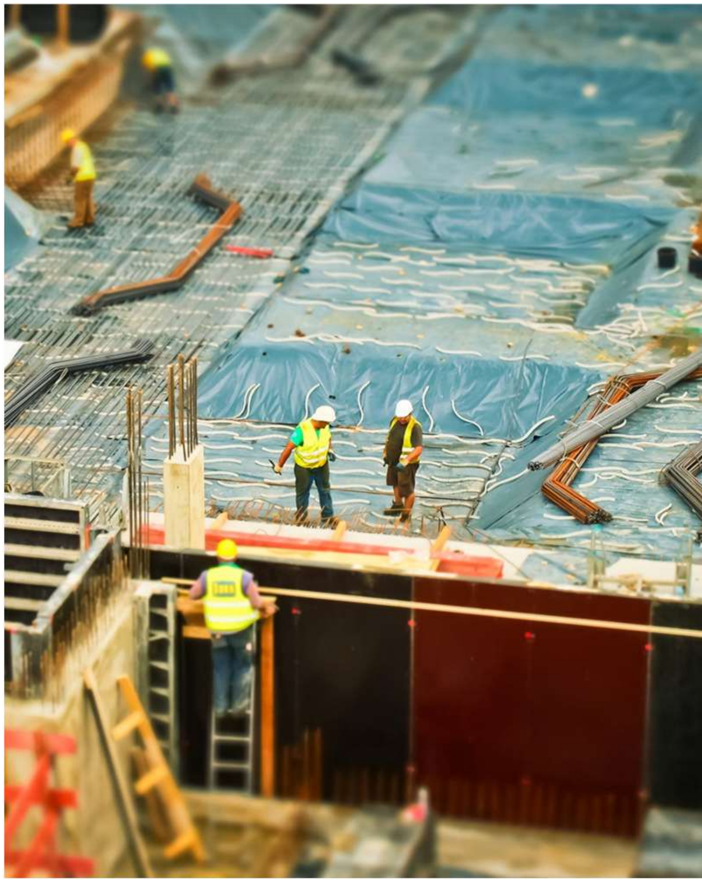
Actions



What is an AI System?



Construction AI System Loop



Environment

Percepts

Construction worker, height, close to ledge

Sensors

ML Algorithm (run on a computer)

Actuators

Intervention: vocally alert/warn workers they are close to falling

Actions

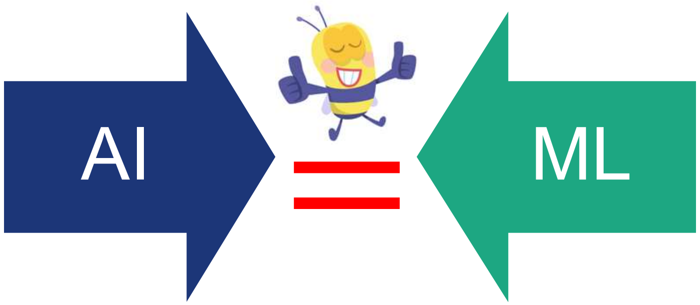
Agent

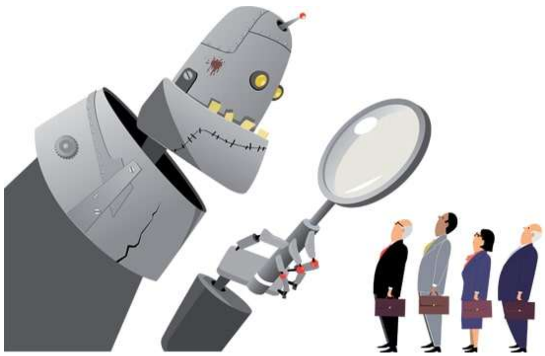
Where can AI be effective (a decision view)?

- Decision cannot be properly made by physics or math-based formulas.
- Decision cannot be made by plain searches of a database.
- Humans are incapable of making the decision as quickly as needed.
- Decision is too complex for a human.
- Decision requires a vast amount of (historical) data that cannot be stored/processed/memorized by the human brain.
- Decision is to be made in an unsafe environment.

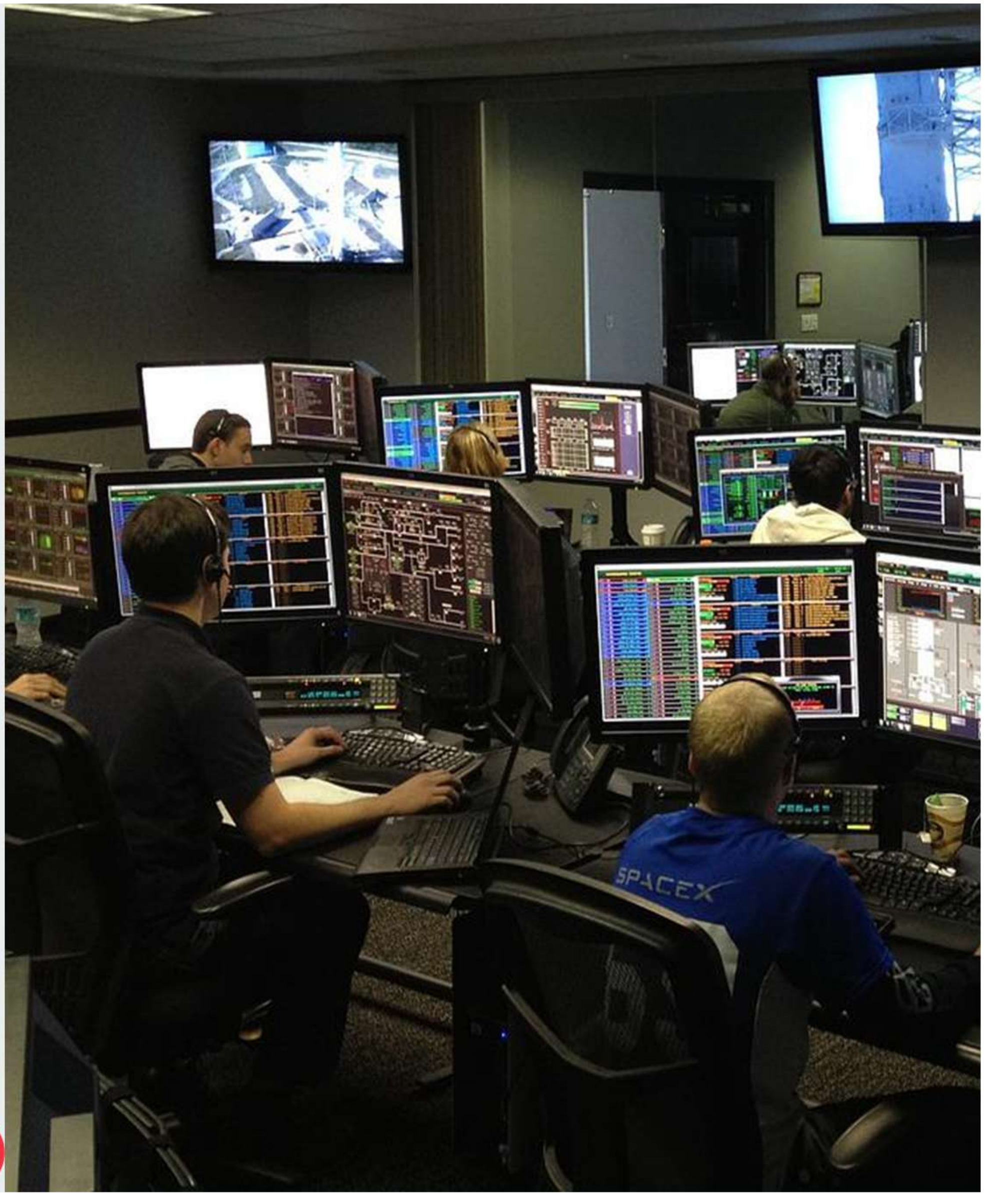
AI Systems Myths

- Experienced computer programmers (or AI experts) are all you need to develop a successful AI application.
- An AI system is just a computer code.
- AI can solve every problem.
- AI and ML are equivalent.
- Decisions made by AI are always correct.
- AI can completely replace humans.



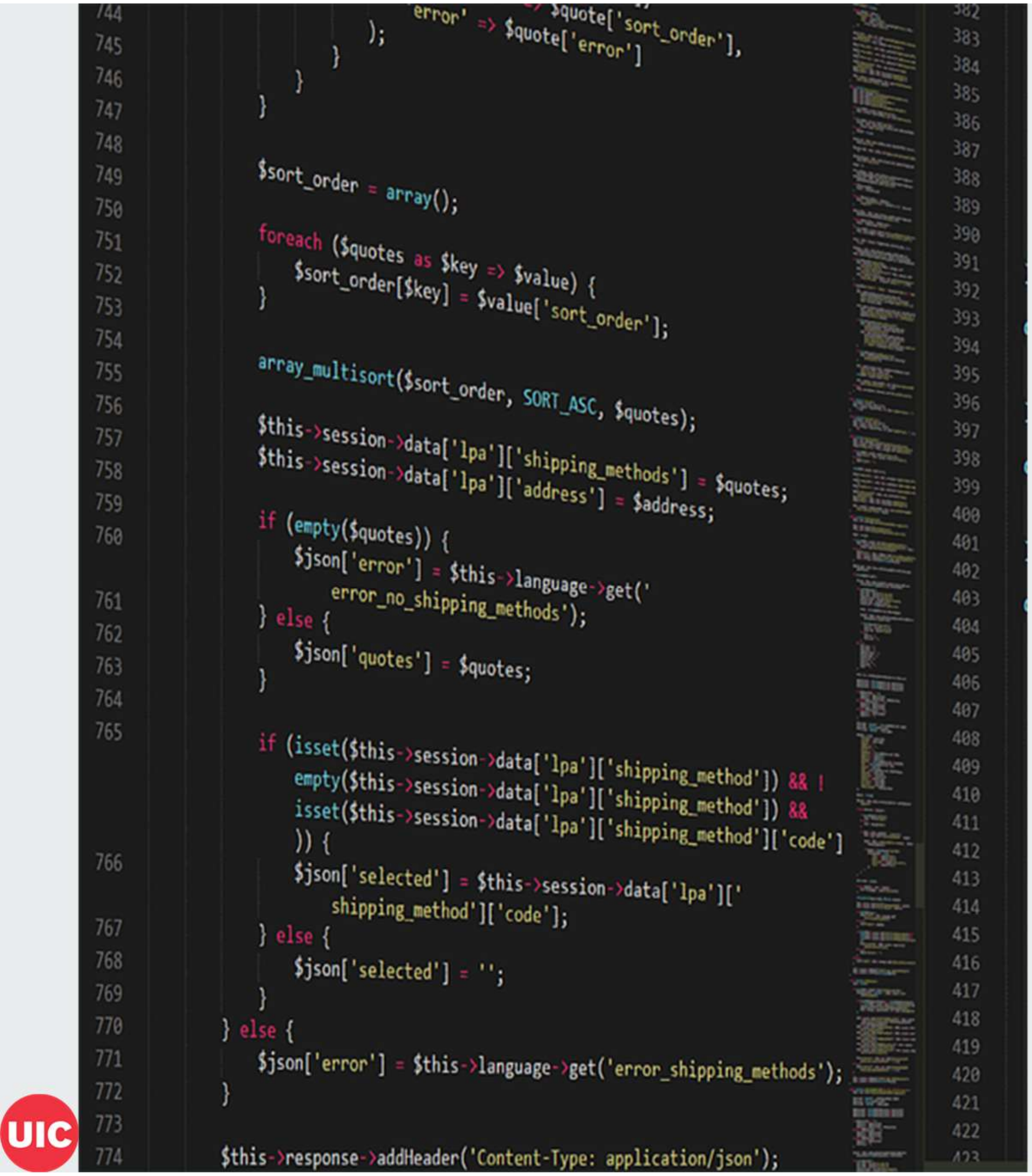


2. Different Types of AI Applications in OSH



1-Real-time Monitoring & Control of Safety Systems

- Use of AI technology to analyze real-time data to detect hazards or reduce the probability of worker injury.
  - Hardware such as sensors, cameras, robots, etc., along with internet connection and specialized software
- Transfer of real-time data allowing AI systems to make prompt decisions, provide notifications/warnings, or, for some advanced systems, stop work completely if a hazard is detected.
- Typically, these AI systems are expected to be very reliable.
- Large Language Models (LLMs) cannot be used for this type of system.

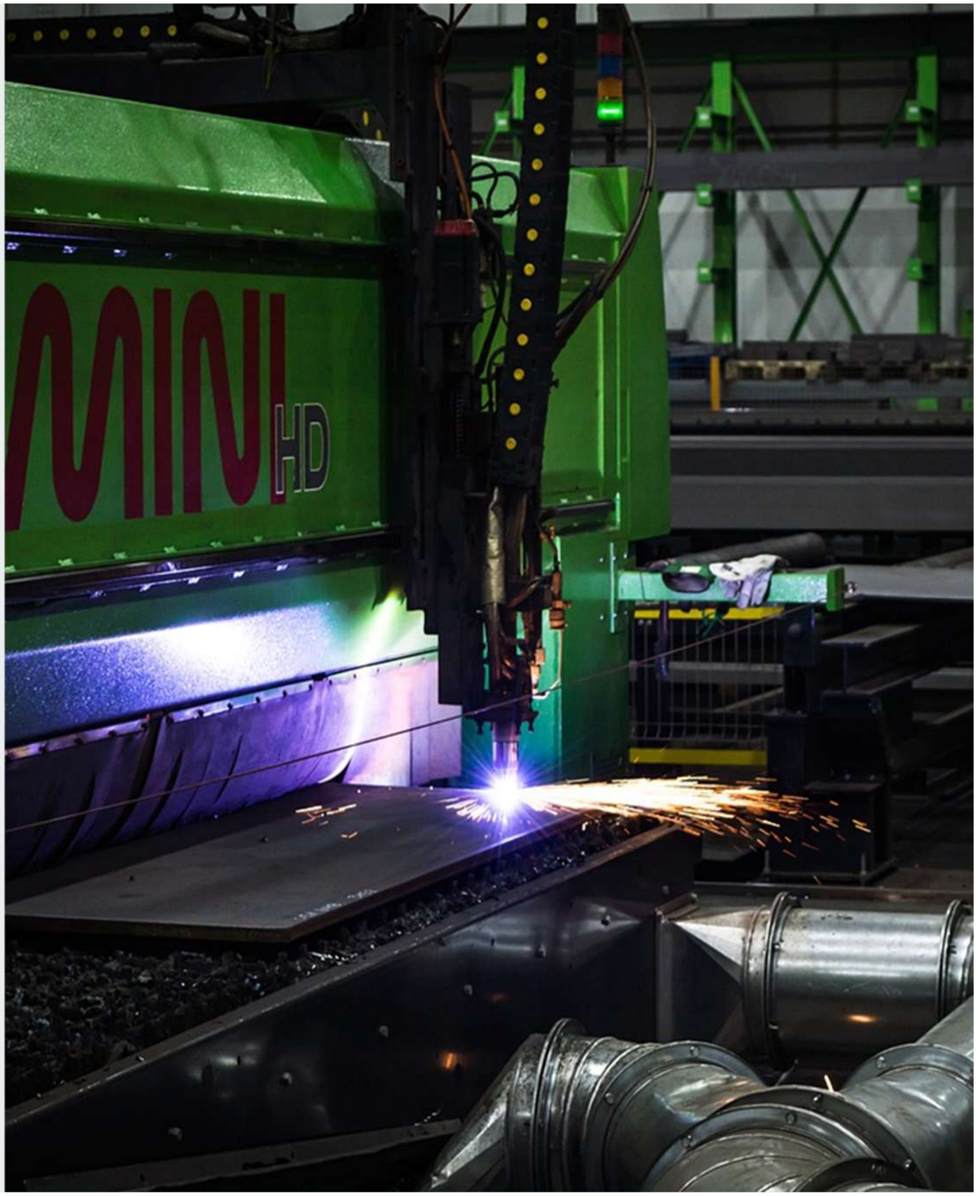


2-Planning and Decision Making (not real-time)

- Use of an AI system where a data set is collected to build a prediction on what actions the AI system will take based on each combination of inputs coming from sensors.
  - Do not need to transfer real-time data or information
- Uses machine learning algorithms designed based on existing data to predict system outcomes.
- Technology can operate with or without an internet connection.
- LLMs can be helpful in such systems.

3-Educational AI

- AI technologies that have been catered for training or teaching.
- Beneficial costwise to help train new users without wasting any human instructor resources.
- Can limit injury for jobs or tasks that are extremely dangerous.
- Example systems: Stimulations, Virtual Reality, Augmented Reality
- Example application: Training medical students in surgery
- LLMs can be used for these systems.



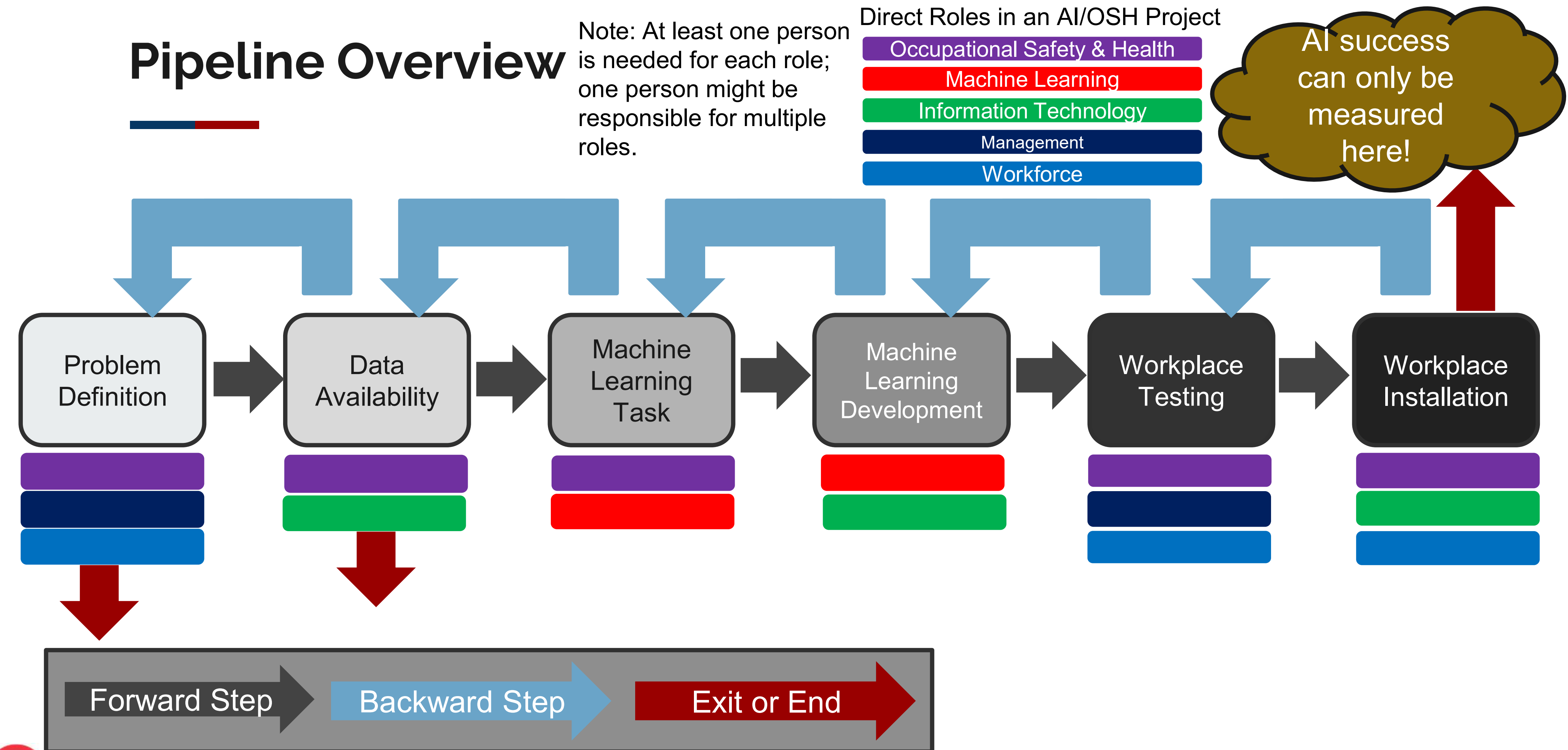
4-Worker Out-of-Loop AI (Fully-Automated)

- This occurs when there is an AI system with no human interaction with the AI technology.
- Removing the human from the AI system loop means no worker safety hazards are experienced.
- Financially, this could save companies more money than paying workers' wages.
- Real-time monitoring or planning and decision-making AI technology
- It is not always possible to fully automate every application and therefore keep the human out-of-loop.
- LLMs cannot be used in this type of system.

3. Developing AI Systems: Project Pipeline and Personnel

Pipeline Overview

Note: At least one person is needed for each role; one person might be responsible for multiple roles.



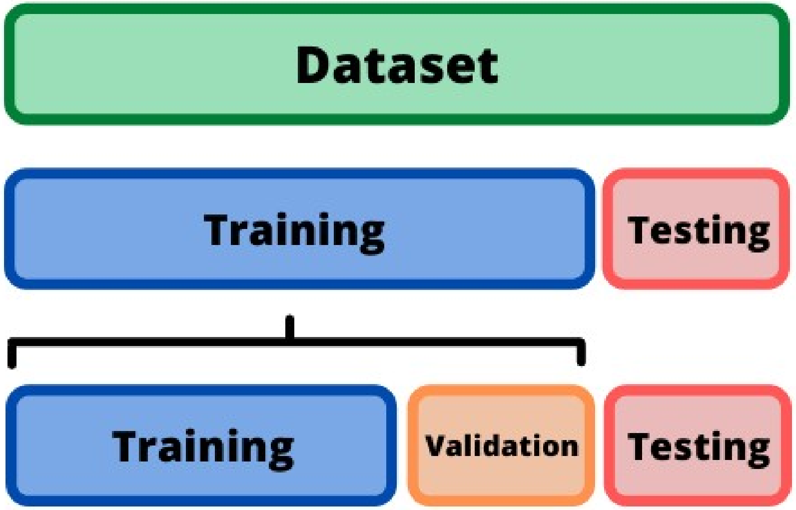
What if the pipeline does not finish successfully: The benefits of failed AI projects

- Data collection is always a good practice. It may reveal missing, inconsistent, and inefficient data systems.
- Automation components (sensors, communication, actuators, decision-making) may increase the efficiency of the data collection and enforcement processes and improve the effectiveness of the (non-AI) decisions made based on the data.
- AI projects force users to think about their decision rules and policies and create a unified set of rules. Sometimes, sharing the unified set among users will help them improve their decisions.
- Failed AI projects could increase the chance of success in completing the redefined AI problem.

4. What is Machine Learning (ML)?

What is a data set and how is an ML model developed?

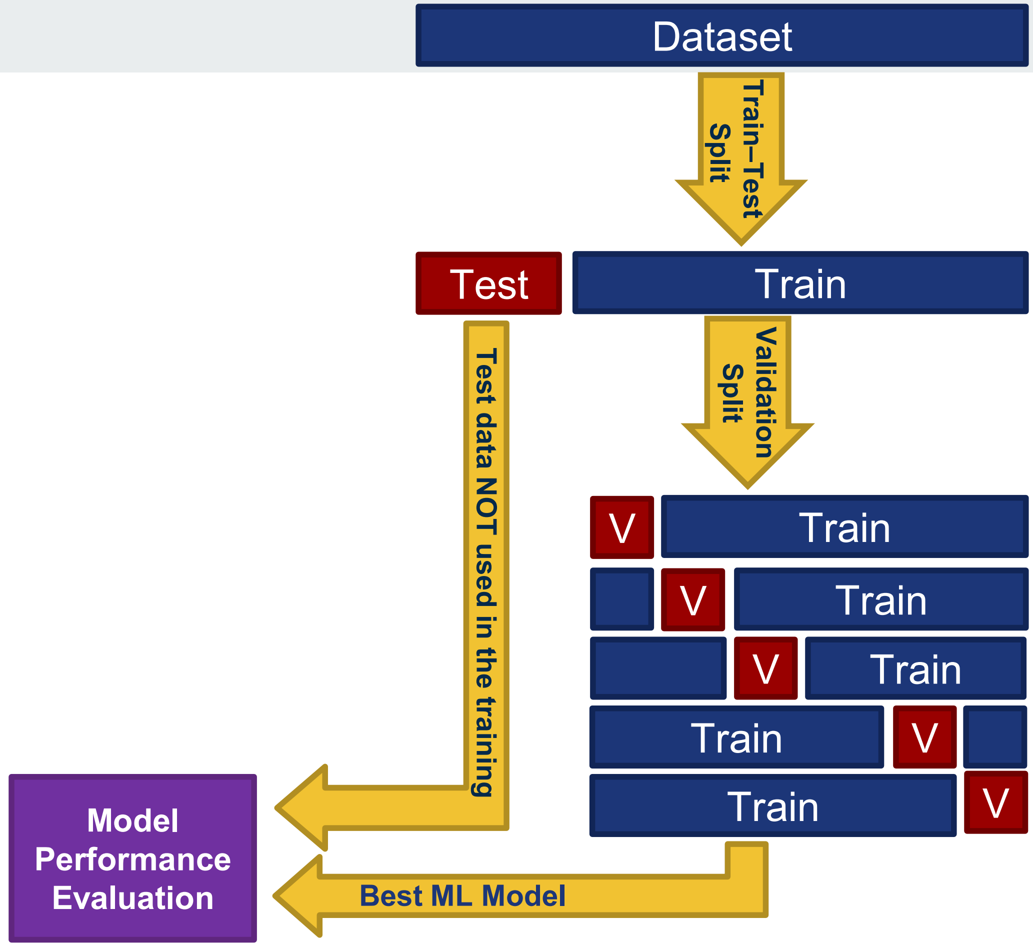
- A dataset is a set of rows. Each row shows the **actual** decision for a given set of actual inputs
- The goal is to develop a model (computer code) to estimate the decision (output) for each given set of inputs.
- The model is successful if it works on the actual data it has not seen before.



- The dataset must be split into subsets with different purposes.
- Training (typically 26-70%): Used to train a model to predict the outcome you design your model to predict
- Validation (typically 0-20%): Used for model section and parameter tuning (Can be a portion of the Training data)
- Test (typically 20-40%): Used as a final evaluation of model performance on completely unseen data

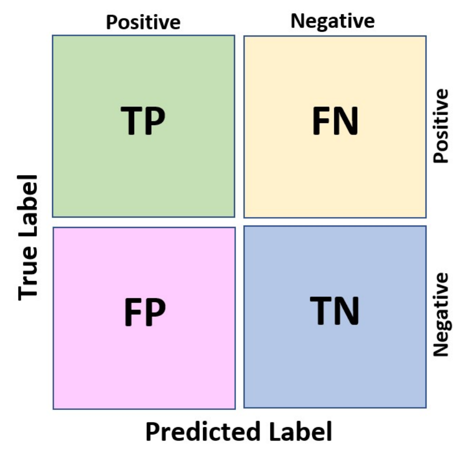
Model Evaluation

- Model evaluation must be done on the data that were not used for training.
- Several evaluation metrics can be used to measure the model's theoretical performance.
- A high-performance model does not necessarily generate an acceptable AI system – everything depends on how the model is used at the workplace.



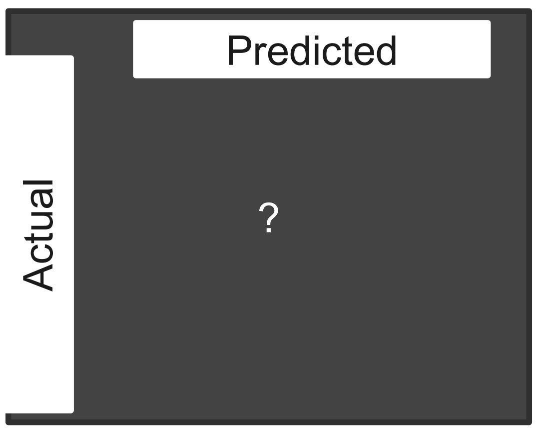
Model Evaluation: Confusion Matrix

Parameters	Abbreviation	Description
True Positives	TP	Represent the instances correctly classified as positive by the model.
True Negatives	TN	Represent the instances correctly classified as negative by the model.
False Positives	FP	Cases where the model predicts positive, but the actual class is negative.
False Negatives	FN	Cases where the model predicts negative, but the true class is positive.



A Confusion Matrix Example

Parameters	Abbreviation	Description
True Positives	TP	Instances correctly classified as positive by the model.
True Negatives	TN	Instances correctly classified as negative by the model.
False Positives	FP	Instances prediction is positive, but the actual is negative.
False Negatives	FN	Instances prediction is negative, but the actual is positive.



Actual Value	Predicted Value
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Negative
Positive	Negative
Positive	Negative
Positive	Negative
Negative	Positive
Negative	Positive
Negative	Negative
Negative	Negative
Negative	Negative
Negative	Negative
Negative	Negative

An Accuracy Example

Accuracy is a common evaluation metric utilized to evaluate model performance.

Accuracy is measured as the number of correctly classified instances divided by the total number of classified.

Predicted

Actual	POSITIVE	NEGATIVE
POSITIVE	8	4
NEGATIVE	2	5

Accuracy = ?

Actual Value	Predicted Value
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Positive
Positive	Negative
Positive	Negative
Positive	Negative
Positive	Negative
Negative	Positive
Negative	Positive
Negative	Negative
Negative	Negative
Negative	Negative
Negative	Negative
Negative	Negative

5. Ethical Concerns with AI

Ethics and AI

- The lack of ethics with AI has led to some major issues regarding the **amount of trust that can be placed in AI**. This is a largely debated topic in the AI community, and some common problem areas discussed are privacy and surveillance, bias and discrimination, and lack of human judgment.
- **Privacy and surveillance:** AI Technology gathers personal data (facial/body images, medical data, personal info, etc.). There is no actual legislation in the US that can protect individuals from the use of personal data in AI applications or software. Not uncommon for data leaks or hacking within AI systems where user data is stolen. Depending on the company and user agreements, this data can be sold off to other companies.
- **Bias** can enter AI algorithms by learning to make decisions based on training data.
- AI technology has intensive algorithms through deep learning, cognitive learning, and predictive analysis to ***mimic*** human judgment to generate a response. AI technology is not perfect and cannot possibly detect every trait/characteristic humans have. **A lack of human judgment** in AI may generate wrong or no response.

Tips on How to Protect Workers Privacy

Privacy and Surveillance Tips[18]	1. Protect data with encryption 2. Collect only necessary data types 3. Use good data sets a. Try to eliminate any historical data that show trends of bias from being used

Worker Relationships[19]	1. Build Relationships with Workers
	    a. Make them feel comfortable with the AI system
	    b. Inform them on how their data is being used
	    c. Allow workers to have some control over the data being used

AI Company[19]	1. Do research on the AI company
	    a. Customer Satisfaction
	    b. Any Data Breaches
	    c. Who they work with
	2. Ensure there is data transparency and disclosure rules between yourself and the AI company

Understanding AI	1. Have an AI expert 2. Learn as much as possible about AI 3. Monitor AI System

[18] *4 ways to preserve privacy in artificial intelligence*. Booz Allen Hamilton. (n.d.). Retrieved March 10, 2023, from https://www.boozallen.com/s/solution/four-ways-to-preserve-privacy-in-ai.html
[19] Kerry, C. F. (2023, March 9). *Protecting privacy in an AI-Driven World*. Brookings. Retrieved March 10, 2023, from https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

Thank you!