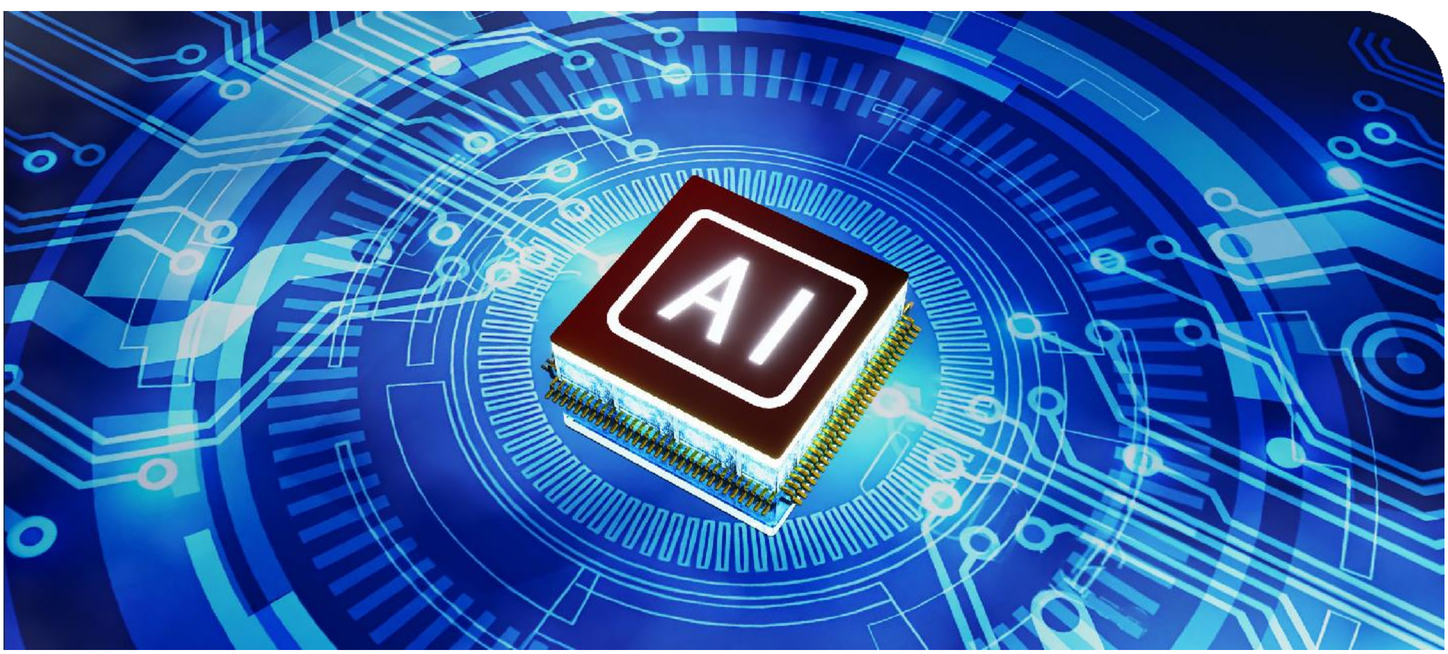
AI Glossary for Safety & Health Professionals

Artificial intelligence (AI) is increasingly valuable in improving worker health and safety by enabling organizations to predict hazards, monitor workplace conditions in real time, and automate high-risk tasks that could otherwise expose workers to harm. From wearable sensors that detect fatigue to predictive analytics that identify injury trends, AI can enhance decision-making and help prevent incidents before they occur. For safety and health professionals, understanding key AI concepts is an important first step toward effectively integrating these tools into their practice. However, any list of AI-related terms should be viewed as a starting point rather than a comprehensive guide; the field is rapidly evolving, and continued learning is essential to keep pace with emerging technologies and their applications in workplace safety.



National Safety Council

Core Artificial Intelligence Concepts

Artificial Intelligence (AI)

What it is:

A broad term for computer systems capable of performing tasks normally requiring human intelligence, such as recognizing patterns, making decisions, learning from data, or generating content.

Why it matters for safety & health:

AI can improve hazard detection, monitoring, and prevention, but also introduce new risks related to reliability, transparency, bias, and over reliance.

Narrow AI (Weak AI)

What it is:

AI systems designed to perform specific, limited tasks within a defined scope. Narrow AI does not possess general understanding or consciousness.

Examples:

- Computer vision models detecting PPE
- Predictive analytics forecasting injury risk
- Speech recognition systems
- Generative AI tools trained for text or image creation

Why it matters for safety & health:

All AI currently used in workplaces is narrow AI. These systems may appear intelligent but can fail outside their training conditions.

General AI (Artificial General Intelligence, AGI)

What it is:

A hypothetical form of AI capable of understanding, learning, and performing any intellectual task a human can, across domains and contexts.

Current status:

General AI does not exist today.

Why it matters for safety discussions:

The term is often used to describe the potential future of artificial intelligence capabilities and while useful in theory, it may cause inflated expectations, misplaced fears or confusion in policy and risk assessment.

Key clarification for professionals:

Workplace safety concerns today relate only to narrow AI, not general AI.

Machine Learning (ML)

What it is:

A subset of artificial intelligence (AI) that enables computer systems to learn, identify patterns, and make decisions from data without being explicitly programmed for every specific task.

Why it matters:

ML systems may change behavior over time, making their performance less predictable than traditional software—an important consideration for safety critical uses.

Deep Learning

What it is:

A subset of machine learning that uses multi layered neural networks to identify complex patterns in large datasets.

Common uses:

- Computer vision
- Speech recognition
- Natural language processing
- Some predictive analytics systems

Why it matters for safety & health:

Deep learning models often outperform traditional systems but are harder to understand and explain.

AI Tools and Technologies

Computer Vision

What it is:

AI that interprets images or video to identify objects, people, activities, or conditions.

Safety & health applications:

- Detecting PPE use or non use
- Monitoring restricted areas
- Identifying unsafe postures or movements
- Spotting spills, obstructions, or fire risks

Key risks:

- False positives or negatives
- Privacy and surveillance concerns
- Performance degradation in poor lighting or complex environments

---

Generative AI

What it is:

AI that creates new content (text, images, video, audio, or code) based on learned patterns.

Safety & health applications:

- Drafting safety procedures or training materials
- Generating incident reports
- Producing simulations for hazard training

Key risks:

- “Hallucinated” or incorrect information
- Overconfidence in AI generated guidance
- Lack of traceability to authoritative sources

Predictive Analytics

What it is:

The use of data, statistics, and AI to forecast future events or trends.

Safety & health applications:

- Predicting injury risk
- Identifying high risk tasks or locations
- Anticipating equipment failures

Key risks:

- Oversimplification of complex work conditions
- Shifting focus from prevention to prediction
- Reinforcing past biases in incident data

Agentic AI (Autonomous or Semi Autonomous Agents)

What it is:

AI systems that can plan, make decisions, and take actions toward goals with limited human input.

Safety & health applications:

- Automated scheduling of inspections
- Coordinated response systems
- Smart safety management platforms

Key risks:

- Reduced human oversight
- Unclear accountability for decisions
- Errors propagating quickly across systems

Robotics

What it is:
Physical machines that perform tasks, often using AI for perception and decision making.

Safety & health applications:

- Removing workers from hazardous environments
- Precision handling of dangerous materials
- Emergency response and inspection

Key risks:

- Human robot interaction hazards
- Unexpected movements or behaviors
- Maintenance and lockout/tagout challenges

Automation

What it is:

The use of technology to perform tasks with minimal human involvement.

Safety & health applications:

- Can reduce exposure to hazards
- Can also increase task pace, monitoring, or workload

Key risk concept:

Automation may shift risk rather than eliminate it, moving hazards from physical to cognitive or organizational domains.

Edge Computing

What it is:

Processing data locally on devices (e.g., cameras, sensors, robots) instead of sending it to the cloud.

Safety & health applications:

- Faster response times for hazard detection
- Reduced reliance on network connectivity
- Enhanced data privacy

Key risks:

- Limited update and oversight capabilities
- Inconsistent performance across devices

AI and Work Organization and Human Factors

Work Intensification

What it is:

The increase in work pace, workload, or performance pressure, often enabled by digital tools.

AI connection:

AI driven monitoring, optimization, and performance scoring can unintentionally accelerate work and reduce recovery time.

Safety & health concern:

Higher injury risk, mental strain, fatigue, and stress related illness.

Algorithmic Management

What it is:

The use of AI systems to assign tasks, set schedules, evaluate performance, or discipline workers.

Why it matters:

Decisions may be opaque, difficult to challenge, or disconnected from real working conditions.

Human in the Loop

What it is:

AI systems designed so humans review, approve, or intervene in AI decisions.

Safety relevance:

Maintains human judgment in safety critical processes and prevents blind reliance on automation.

Human on the Loop

What it is:

Humans supervise AI systems and can intervene if necessary but are not involved in every decision.

Key challenge:

Workers may be expected to respond immediately without fully understanding system behavior.

Automation Bias

What it is:

The tendency for humans to trust AI outputs even when they are incorrect.

Safety implication:

Workers may ignore their own observations or warning signs due to misplaced trust in AI systems.

AI Risk, Ethics, and Governance

Explainability

What it is:

The ability to understand how an AI system reaches a decision.

Why it matters:

Safety professionals need to explain, justify, and audit AI driven safety decisions, especially after incidents.

Bias

What it is:

Systematic errors that disadvantage certain groups or conditions.

Safety impact:

AI may under or over-detect hazards affecting particular worker groups, job roles, or environments.

Reliability

What it is:

The consistency of AI system performance under real world conditions.

Safety concern:AI that works well in testing but fails in practice can create false confidence.

Fail Safe Design

What it is:

System design that defaults to a safe state if AI fails or behaves unexpectedly.

Critical safety principle:

AI systems should not introduce catastrophic failures when they malfunction.

Cyber Physical Risk

What it is:

Risk arising when digital systems influence physical processes.

Why it matters:

AI errors or cyberattacks can directly cause physical harm in workplaces.

Other Relevant Terms

API (Application Programming Interface)

An **API** is a set of rules and protocols that allows different software systems to communicate with each other.

- It defines how requests are made and responses are returned.
- In AI, APIs are commonly used to access models (e.g., sending a prompt to an AI model and receiving a response).

Example:

A chatbot sends a user query to an API and receives an answer.

Digital Twin

A **digital twin** is a virtual representation of a physical object, system, or process that is continuously updated using real-world data.

- It allows simulation, monitoring, and analysis of the real-world counterpart.
- Often used in industries like manufacturing, healthcare, and smart cities to predict performance, optimize operations, and detect issues early.

Example:

A digital twin of a wind turbine tracks real-time sensor data to predict maintenance needs and improve efficiency.

Hallucinations (in AI)

Hallucinations occur when an AI system generates information that is incorrect, fabricated, or not grounded in real data, but presents it as if it were true.

- Common in large language models when they “fill in gaps” without reliable sources.

Example:

An AI confidently citing a nonexistent research paper.

Leakage (Data Leakage)

Leakage refers to unintended exposure of information, especially sensitive or restricted data. This differs from a data breach which is an intentional cyberattack to extract data.

- In machine learning, it can also mean using data in training that would not be available in real-world predictions, causing overly optimistic results.
- In AI systems, it may involve revealing private data through prompts or outputs.

Example:

A model unintentionally revealing personal data it was trained on.

NLP (Natural Language Processing)

NLP is a branch of AI focused on enabling computers to understand, interpret, and generate human language.

- It combines linguistics, machine learning, and computer science.
- Used in applications like chatbots, translation tools, sentiment analysis, and voice assistants

Example:

Converting spoken speech into text or generating a response to a question.

Prompting

Prompting is the practice of giving input (instructions or questions) to an AI model to guide its output.

- The quality and wording of a prompt greatly influence the response.
- Advanced forms include prompt engineering, which optimizes prompts for better results.

Example:

Writing “Summarize this text in 3 bullet points” to shape how the AI responds.

National Safety Council

1820450 0626