Science & Technology in Policing

GLOSSARY OF TERMS FOR ARTIFICAL INTELLIGENCE IN POLICING

June 2025

Office of the Police Chief Scientific Advisor (OPCSA) in partnership with Home Office Accelarated Capability Environment (ACE)

OFFICIAL

1. Context.

This glossary has been developed in accordance with the recommendations of the AI Taxonomy, which was created to address the challenges of compiling taxonomies in this dynamic and multifaceted field. The glossary aims to help policing end users better understand the essential terms relevant to AI, drawing on the five sources outlined below:

1. Policing and Artificial Intelligence[1]
2. International AI Safety Report[2]
3. AI and Policing[3]
4. Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence[4]
5. AI Procurement in a Box[5]

.

2. Glossary of Terms

- **Accountability:** The responsibility and explainability for the actions and decisions made by AI systems, ensuring adherence to ethical standards and legal regulations.
- **AI Agent:** A general-purpose AI which can make plans to achieve goals, adaptively perform tasks involving multiple steps and uncertain outcomes along the way, and interact with its environment with little to no human oversight.

[1] Rick Muir and Felicity O'Connell, Policing and Artificial Intelligence, (The Police Foundation, February 2025) <https://www.police-foundation.org.uk/wp-content/uploads/2010/10/policing-and-ai.pdf> (accessed 10/03/2025)
[2] AI Action Summit, International AI Safety Report: The International Scientific Report on the Safety of Advanced AI' (AI Summit, January 2025) <International AI safety report> (accessed 17/03/2025)
[3] Europol Innovation Lab, AI and Policing – The Benefits and Challenges of Artificial Intelligence for Law Enforcement, (Europol, 2024) < AI-and-policing.pdf> (accessed 17/03/2025)
[4] European Commission, Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence, (European Commission, July 2020) <Sectoral Considerations On The Policy And Investment Recommendations For Trustworthy Artificial Intelligence> (accessed 17/03/2025)
[5] World Economic Forum, AI procurement in a Box, (Word Economic Forum, June 2020) <https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf > (accessed 20/03/25)

- **AI Governance:** The framework and regulations that guide the development, deployment, and use of AI systems. It includes ethical considerations, accountability mechanisms, and compliance with legal standards.
- **AI Office:** The European Commission's function contributing to the implementation, monitoring and supervision of AI systems, general purpose AI models and AI governance.
- **AI System:** A machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.
- **Algorithm:** A set of sequential instructions applied on specific input(s) to reach a predefined goal, ranging from the output of a basic mathematical function to more complex tasks.
- **Algorithmic (training) efficiency:** Measures of how efficiently an algorithm uses computational resources to learn from data, such as the amount of memory used or the time taken for training.
- **Artificial Intelligence (AI):** The development of computer systems that can perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, and language understanding. AI refers to technologies that aim to reproduce or surpass abilities (in computational systems) that would require 'intelligence' if humans were to perform them.
- **Benchmark:** A standardised, often quantitative test or metric used to evaluate and compare the performance of AI systems on a fixed set of tasks designed to represent real-world usage.
- **Bias:** Systematic and unfair preferences or prejudices in the data or algorithms that can lead to discriminatory outcomes. Bias can arise from the data used to train AI models or the design of the algorithms themselves.
- **Biometric Data:** Personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, such as facial images or dactyloscopic data.
- **Capabilities:** The range of tasks or functions that an AI system can perform, and how competently it can perform them.
- **Chain of thought:** A reasoning process in which an AI generates intermediate steps or explanations while solving a problem or answering a question. This approach mimics human logical reasoning and internal deliberation, helping the model break down complex tasks into smaller, sequential steps to improve accuracy and transparency in its outputs.

- **Chatbot:** Software designed to mimic conversation with humans.
- **Classification:** A type of machine learning task where the algorithm assigns predefined categories or labels to input data.
- **Cloud Computing:** The provision of computing services, such as storage, computational power, and software applications, via the internet.
- **Cognitive tasks:** Activities that involve processing information, problem-solving, decision-making, and creative thinking. Examples include research, writing, and programming.
- **Compute:** Shorthand for 'computational resources', which refers to the hardware (e.g. GPUs), software (e.g. data management software) and infrastructure (e.g. data centres) required to train and run AI systems.
- **Computer Vision:** A field of AI that enables machines to interpret and make decisions based on visual data. Computer vision is used for image and video analysis.
- **Conformity Assessment:** The process of demonstrating whether the requirements set out in the EU AI Act relating to a high-risk AI system have been fulfilled.
- **Data minimisation:** The practice of collecting and retaining only the data that is directly necessary for a specific purpose, and deleting it once that purpose is fulfilled.
- **Deep Learning:** Deep learning is a subset of machine learning and AI that mimics human knowledge acquisition to enable models to perform complex tasks like recognising patterns in various data types.
- **Deepfakes:** AI generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful.
- **Deployment:** The process of implementing AI systems into real-world applications, products, or services where they can serve requests and operate within a larger context.
- **Emergent behaviour:** The ability of AI systems to act in ways that were not explicitly programmed or intended by their developers or users.
- **Encryption:** The process of converting information into a code to prevent unauthorised access, particularly important for securing data in AI applications.
- **Evaluations:** Systematic assessments of an AI system's performance, capabilities, vulnerabilities or potential impacts. Evaluations can include benchmarking, red-teaming and audits and can be conducted both before and after model deployment.

- **Explainability:** The ability to understand and interpret the decisions and actions of AI systems, particularly in complex models like neural networks.
- **Fairness:** Ensuring that AI systems treat all individuals and groups fairly, without introducing biases or discrimination.
- **First-mover advantage:** The competitive benefit gained by being the first to establish a significant market position in an industry.
- **FLOP:** 'Floating point operations' – the number of computational operations performed by a computer program. Often used as a measure for the amount of compute used in training an AI model.
- **Foundation Model:** A pre-trained model that serves as the starting point for various downstream tasks in machine learning.
- **General Purpose AI (GPAI):** An AI system which is based on a general-purpose AI model that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems. 'General-purpose AI' refers to artificial intelligence models or systems that can perform a wide range of tasks rather than being specialised for one specific function.
- **Generative AI:** AI systems that can generate new content, such as images, text, or music, often using general purpose AI models.
- **GHG (greenhouse gas) emissions:** Release of gases such as carbon dioxide (CO_2), methane, nitrous oxide, and hydrofluorocarbons which create a barrier trapping heat in the atmosphere. A key indicator of climate change.
- **Guardrails:** Built-in safety constraints to ensure that an AI system operates as desired and avoids harmful outcomes.
- **High-Risk AI:** AI applications with the potential for significant impact on individuals' rights and safety, requiring stricter regulatory scrutiny.
- **Inference:** The process in which an AI generates outputs based on a given input, thereby applying the knowledge learnt during training.
- **Inference-time enhancements:** Techniques used to improve an AI system's performance after its initial training, without changing the underlying model. This includes clever prompting methods, answer selection methods (e.g. sampling multiple responses and choosing a majority answer), writing long 'chains of thought', agent 'scaffolding', and more.
- **Input (to an AI system):** The data or prompt submitted to an AI system, such as text or an image, which the AI system processes and turns into an output.
- **Interpretability:** The degree to which humans can understand the inner workings of an AI model, including why it generated a particular output or decision. A model is highly interpretable if its mathematical processes can be

translated into concepts that allow humans to trace the specific factors and logic that influenced the model's output.

- **Interpretability research:** The study of how general-purpose AI models function internally, and the development of methods to make this comprehensible to humans.
- **Jailbreaking:** Generating and submitting prompts designed to bypass guardrails and make an AI system produce harmful content, such as instructions for building weapons.
- **Labour market disruption:** Significant and often complex changes in the labour market that affect job availability, required skills, wage distribution, or the nature of work across sectors and occupations.
- **Large Language Model:** A type of AI model, like GPT-4, that is trained on vast amounts of text data and can generate human-like language.
- **Likeness rights:** Rights that protect an individual's image, voice, name, or other identifiable aspects from unauthorised commercial use.
- **Loss of control scenario:** A scenario in which one or more general-purpose AI systems come to operate outside of anyone’s control, with no clear path to regaining control.
- **Machine Learning:** A subset of AI that focuses on the development of algorithms that enable computers to learn and make predictions or decisions based on data.
- **Misalignment:** An AI’s propensity to use its capabilities in ways that conflict with human intentions or values. Depending on the context, this can variously refer to the intentions and values of developers, operators, users, specific communities, or society as a whole.
- **Modalities:** The kinds of data that an AI system can competently receive as input and produce as output, including text (language or code), images, video, and robotic actions.
- **Model:** A computer program, often based on machine learning, designed to process inputs and generate outputs. AI models can perform tasks such as prediction, classification, decision-making, or generation, forming the core of AI applications.
- **Narrow AI:** A kind of AI that is specialised to perform one specific task or a few very similar tasks, such as ranking web search results, classifying species of animals, or playing chess.
- **Natural Language Processing:** A subset of artificial intelligence, focusing on how computers and human languages interact, allowing machines to comprehend, decipher, and generate human language.

- **Neural Networks:** A set of algorithms, modelled loosely after the human brain, designed to recognise patterns. Neural networks are a key component of deep learning.
- **Open-ended domains:** Environments into which AI systems might be deployed which present a very large set of possible scenarios. In open-ended domains, developers typically cannot anticipate and test every possible way that an AI system might be used.
- **Privacy:** A person's or group's right to control how others access or process their sensitive information and activities.
- **Prompt:** An input to an AI system, such as a text-based question or query, that the system processes and responds to.
- **Reliability:** An AI system's ability to consistently perform its intended function.
- **Scaling laws:** Systematic relationships observed between an AI model's size (or the amount of time, data or computational resources used in training or inference) and its performance.
- **Sensitive data:** Information that, if disclosed or mishandled, could result in harm, embarrassment, inconvenience, or unfairness to an individual or organisation.
- **System:** An integrated setup that combines one or more AI models with other components, such as user interfaces or content filters, to produce an application that users can interact with.
- **Weights:** Model parameters that represent the strength of connection between nodes in a neural network. Weights play an important part in determining the output of a model in response to a given input and are iteratively updated during model training to improve its performance.