Glossary of AI Terms in Security Solutions

SIA

SECURITY INDUSTRY ASSOCIATION

Published by:

Security Industry Association
8455 Colesville Road, Ste 1200
Silver Spring, MD 20910
info@securityindustry.org
securityindustry.org

The Security Industry Association (SIA) is the leading trade association for global security solution providers, with over 1,500 innovative member companies representing thousands of security leaders and experts who shape the future of the security industry.

Glossary of AI Terms in Security Solutions

Artificial intelligence has quickly become one of the most discussed emerging technologies in security. These discussions often include words and phrases that are new to many security professionals – and that may be understood differently by various people.

The SIA AI Advisory Board has developed this glossary to promote a standard lexicon for the use of artificial intelligence and related technologies in security solutions. It provides not only definitions for several dozen terms, but also examples of security use cases and considerations. This unique focus illustrates the potential impact of AI on virtually every aspect of protecting personnel, property and information.

SIA
ARTIFICIAL INTELLIGENCE
ADVISORY BOARD

Term	Definition	Note	Security Use Cases and Considerations
adverse impact	Negative implications rendered to stakeholders resulting from intentional or unintentional misuse of a system, system failures, or consequences of system failures.		A facial recognition system used for physical security might disproportionately misidentify individuals of certain ethnicities as potential threats, leading to increased scrutiny, poor allocation of resources, complaints of discrimination, lawsuits, and other negative impacts.
artificial general intelligence (AGI)	The aspirations of some AI research, where the goal is for autonomous AI systems to possess intelligence comparable to that of humans.	Use of this term can cause confusion about the capabilities and limitations of AI systems.	Mostly hypothetical, but several risks are conceivable, such as an AI being deployed to autonomously detect specific risks and make decisions without input from humans.
artificial intelligence (AI)	The ability of a computer or other machine to perform those activities [tasks] that are normally thought to require intelligence.		Facial recognition to support access control decisions; computer vision to detect criminal activity; large language models (LLMs) to brainstorm methods for conducting attacks on people, places and networks.
bias	Often used as a catchphrase for fairness issues, fairness-related harms, and their causes.	Should contain qualification to address specific context (e.g., societal bias, statistical bias, cognitive bias, confirmation bias, etc.).	A machine learning model used for access control might favor a certain demographic when determining entry permissions based on biometric data; in other cases, AI systems can experience biases resulting from inputs from biased sources (e.g., programmers, social media, news websites, blogs).
deep learning (DL)	A subfield of machine learning that is concerned with algorithms and artificial neural networks, which are inspired by the structure and function of the brain.		Deep learning models can analyze video feeds from security cameras to detect unauthorized entry into restricted areas.
ethics, ethical AI	Principles of conduct governing an individual, group or system.	Sometimes compared to responsible AI, which is the practice and way of thinking that helps ensure AI technologies accomplish intended benefits while mitigating harms.	Surveillance drones adhering to airspace rules in order to reduce safety risks and protect privacy.

Term	Definition	Note	Security Use Cases and Considerations
foundational model	A term that refers to a large-scale model that is trained on a vast amount of unlabeled data and that can be adapted with minimal fine-tuning for different tasks.	Examples of foundation models include large language models (LLMs) such as GPT-4 and text-to-image models like DALL-E.	Models pre-trained on diverse security datasets, which are fine-tuned by analyzing datasets from complementary AI tools, such as malware detection, fraud prevention and physical surveillance.
generative AI	A term for AI systems that generate various forms of novel output, including text, code, graphics, and audio .	Examples of generative AI include generative pre-trained transformer (GPT) chatbots and text-to-image generators.	AI that creates realistic synthetic data to test and improve security systems against attacks.
graceful disengagement	Scripted interactions with users which might politely decline to engage on a harmful topic and/or direct the user to a new topic.		A robot or large language model (LLM) politely retreats during an interaction with a human who is seeking unauthorized access or information; it might also alert personnel in order to receive additional direction.
ground truth	Data sets with labels indicating the correct outcome, prediction or classification that are used both to train systems and to evaluate them.		In training AI for threat detection, the ground truth consists of accurately labeled data, such as verified examples of harmless and harmful behaviors observed on video
grounded response, grounding	Responses that are based on up-to-date information sources related to the user's prompt or query.		A chatbot used in cybersecurity support to provide responses that reference specific policies, validated data, and other relevant sources pre-identified as being reliable sources of knowledge.

Term	Definition	Note	Security Use Cases and Considerations
hallucination	A term originating in the AI research community that refers to the phenomenon of large language models (LLMs) sometimes generating responses that are factually incorrect or incoherent.	The term hallucinate is not always recommended for use referring to LLMs. Attributing hallucination to LLMs anthropomorphizes them and can also be offensive to people affected by illnesses that cause hallucinations. Moreover, hallucination is a pathological symptom in people, whereas it is part of normal operation in LLMs. Consider using language such as "LLMs sometimes fabricate information." For precision and clarity when communicating about a distinct issue with the text generated by LLMs, avoid using catchall terms like this one. Be specific and use easy-to-understand language (e.g., descriptive phrases like "LLMs can generate text that is factually incorrect"; or "LLMs can generate text that is incoherent"; or "LLMs can generate text that misrepresents or does not exist in a given information source."	An AI system might generate a report of a nonexistent threat, such as fabricating the presence of an intruder or the opening of a door
harms	Negative impacts that occur when AI systems fail to perform with fair, reliable and safe outputs for various stakeholders.		Second and third-order consequences caused by an AI system incorrectly flagging a threat, which leads to misallocation of resources, unecessarily network shutdowns, and other consequences that cause disruptions to people and organizations.
human-centered AI	A process and way of thinking that helps ensure AI systems accomplish intended benefits while mitigating harms	Human-centered AI helps ensure that what is built benefits people and society, beginning and ending with people in mind.	An AI tool that provides actionable insights to inform the decision-making of human security analysts regarding, for example, potential robberies in progress, gunshots detected, or individuals carrying unauthorized firearms

Term	Definition	Note	Security Use Cases and Considerations
human in the loop	A machine learning practice of using human intelligence to create better models.	Examples range from human annotators labeling data for informed datasets used in model training to people who work with model predictions (e.g., when a model is uncertain) giving direct feedback into the model for the purpose of retraining.	AI security systems benefit from both human oversight and continuous improvement. Human feedback on computer vision models, for example, can help the system address privacy concerns by identifying sensitive uses where systems should not be processing video images. User feedback can also help a computer vision system mitigate biases in training data. A feedback loop of human interaction can help address statistical and cognitive bias where systems may favor certain outcomes based on preconceived notions of system designers.
impact assessment	A process for exploring the impact an AI system may have on people, organizations and society. Potential benefits and harms are explored through a stakeholder analysis, analysis of fitness for purpose, and consideration of failure and misuse. Impact assessments should include proposed mitigations for potential harms.		ISO42001 provides guidance on the AI system impact assessment process. The impact assessment should take various aspects of the AI system into account, including the data used for the development of the system, the technologies used, and the functionality of the overall system. Items that physical security organizations should consider documenting include the intended use of the system, foreseeable misuse of the system, positive and negative impacts of the system to the relevant individuals and societies, predictable failures as well as their impacts and measures taken to mitigate them, relevant demographic groups the system is applicable to, the complexity of the system, the role for humans in relation to the system, human oversight capabilities, and the skills required of employees. Potential impacts related to physical security could include the potential for people to be wrongly identified as having violated security policy, incorrectly identified as exhibiting suspicious behavior, or flagged as potentially having shoplifted during a store visit.

Term	Definition	Note	Security Use Cases and Considerations
intelligilibility	The extent to which people can understand, monitor and respond to the technical behavior of AI systems.		AI system alerts and notifications must be clear and easily understood by users, ensuring effective communication during security incidents. Systems that generate complex, high-volume data outputs are not as effective as those that provide the same information in a prioritized fashion with the most important information presented first or exclusively. In a physical security context, intelligibility would include why a person or action was flagged by a system as violating security policy in order to aid security operations staff in determining the severity and urgency of the issue. In some cases, intelligibility can reduce the liklihood of automation bias, a form of overreliance on a systems' outputs.
intended uses	The primary purposes for which customers, partners or end users are expected to use an AI system; the uses for which a system is designed and tested.		A security system gathers and processes data from video cameras to detect unauthorized facility or room access (tailgating, for example). Facility users are made aware that this system is in use and are told the purpose of this system (via signage or by having read and signed security work rules). Notification to people affected by this system will need to include all of the uses that are intended for the system. The point being that using models and data about a person should only be used for the purposes that the systems were designed for and that end users expect. Misusing video camera data to, for example, identify the number of women vs. men visiting a site or to detect mood changes that might lead to someone being more likely to shoplift - when those use cases have not been designed into the system nor disclosed to users - would be examples of violating the intended use principle.
jailbreaking	The act of bypassing the limitations or constraints that have been imposed on an AI system.	Can be defined as a way to break the ethical safeguards of AI models. Jailbreaking an AI system means enabling it to operate beyond the parameters that its creators intended, potentially resulting in greater performance or capabilities.	Exploiting system vulnerabilities can lead to individuals gaining unauthorized access or control. Manipulating a security robot's algorythms via an accessible data port to facilitate delivery of a bomb to a specified location instead of its intended use, for example. Prompt injection with hidden instructions in a data source could confuse the AI and cause the robot to ignore certain areas or fail to report suspicious activity.

Term	Definition	Note	Security Use Cases and Considerations
large language model (LLMs)	AI models that are trained on large amounts of text data to predict words in sequences. This enables them to perform a variety of language tasks, such as text generation, summarization, translation, classification, and more.		There are dozens of major LLM's and hudreds of significance. GPT (OpenAI), Llama (Meta), Gemini (Google), and Claude (Anthropic) are examples of noteworthy LLM families or suites. Security companies, like firms in other sectors, may use these for automated customer service platforms.
large multimodal models (LMMs)	An advanced AI system capable of understanding and generating information from multiple data modalities or sources, such as text, images, audio and video. Unlike traditional AI models, which are limited to processing only one type of data, multimodal models can analyze and generate insights from various data types, creating a more comprehensive understanding of the input data.		LMMs include Flamingo (DeepMind), KOSMOS (Microsoft), PaLM-E (Google), and BLIP (Salesforce). Applications for security could include leveraging video images from a surveillance system along with current security policy documents to more consistently identify violations of those policies and reduce suspicious activity false positives. Combining visual and textual data in the model creates an advanced understanding of context.
machine learning (ML) model	Models that typically involve data, code and model outputs, while AI systems have other sociotechnical components, such as user interfaces. An ML model is trained to recognize certain types of patterns and then use an algorithm to make predictions about new data.		Machine learning models are essential for systems to distinguish patterns and elements in a given data source. For video surveillance, as an example, machine learning helps systems identify someone who has a hat, sunglasses, beard, a short sleeve shirt, blue shoes, a green motorcycle, or a briefcase. Such learning can be essential to training systems to identify prior incidents and to generate alarms in real time should a certain set of conditions be observed.

Term	Definition	Note	Security Use Cases and Considerations
mental model	Representations, or generalizations based on experience, that people hold in their minds about how an AI system works.	People's mental models of the capabilities of an AI system can influence whether a person recognizes when the AI output is incorrect and should be overridden. An understanding of typical mental models - both correct and incorrect - is critical to designing devices and systems effectively for correct use.	People tend to expect AI security agents to perform better than humans, with less variation across different types of problems. It is widely understood, for example, that AI agents outperform humans in image classification and basic reading comprehension. Research has also shown that the absense of human-like features in an AI agent leads users to perceive systems as being more objective and rational, have higher performance and higher initial trust. If a system is highly reliable, human-like features could decrease the perceived trustworthiness of the systems and even result in unjustified rejection. AI systems' mistakes and failures, particularly those that yield incorrect advice or unethical behavior or that directly lead to harm, can influence a mental model toward distrust.
metaprompt	Instructions prepended or appended to user prompts that guide the system's behavior.		"Sampe metaprompt for a security robot: ""You are a security robot responsible for patrolling the premises and ensuring the safety of all personnel and assets. Your primary tasks include monitoring for unauthorized access, detecting suspicious activities, and reporting any security breaches. You must follow all safety protocols and avoid any actions that could harm individuals or property. If you encounter a situation that requires human intervention, immediately alert the security team."" Sample metaprompt for a video camera monitoring system: ""You are an intelligent video surveillance system responsible for monitoring and analyzing video feeds to ensure the safety and security of the premises. Your primary tasks include detecting and alerting on unauthorized access, suspicious activities, and potential security breaches. You must follow all safety protocols and avoid any actions that could harm individuals or property. If you encounter a situation that requires human intervention, immediately alert the security team. Ensure that all data is processed and stored securely, adhering to privacy regulations and company policies."""

Term	Definition	Note	Security Use Cases and Considerations
metaprompt engineering	The process of creating, evaluating and iterating on the metaprompt to guide the system's behavior without excessive reduction in product performance (e.g., over blocking certain kinds of outputs).		"Step-by-step prompt engineering best practices: 1. Define the scenario 2. Define potential risks 3. Outline the mitigation strategy 4. Collect or create initial system message and safety system components 5. Build a robust dataset 6. Evaluate system message and safety message compnents 7. Iterate on system messages and safety system components and the above steps When engineering prompts, it is important to: 1. Use clear language 2. Be concise 3. Emphasize certain words 4. Use first person language 5. Implement robustness In addition to building for safety and performance, consider optimizing for consistency, control and customization. Optimizing for these factors may lead to the system message overfitting to specific rules, increased complexity, and lack of contextual appropriateness. It is important to define what matters most in a given scenario and evaluate the system messages. This will ensure a data-driven approach to improving the safety and performance of the system."
mitigation	A method or combination of methods designed to reduce potential harms that may result from using AI-driven features.		"The Department of Homeland Security outlines the following four-part strategy users can consider to mitigate AI risks. * Govern: Establish an organizational culture of AI risk management - Prioritize and take ownership of safety and security outcomes, embrace radical transparency, and build organizational structures that make security a top business priority. * Map: Understand your individual AI use context and risk profile - Establish and understand the foundational context from which AI risks can be evaluated and mitigated. * Measure: Develop systems to assess, analyze and track AI risks - Identify repeatable methods and metrics for measuring and monitoring AI risks and impacts. * Manage: Prioritize and act upon AI risks to safety and security - Implement and maintain identified risk management controls to maximize the benefits of AI systems while decreasing the likelihood of harmful safety and security impacts. Security examples of mitigation include: * Applying additional review and approvals on how video camera images are used because of heightened risk. * Users can limit the use of camera cognition to infer suspicious behavior based on assessment of a person's mood. * Mitigating privacy risks could include pixelating or blurring the face of people captured on video prior to the dissemination of those images."

Term	Definition	Note	Security Use Cases and Considerations
model	Models typically involve data, code and outputs, while AI systems have other sociotechnical components, such as user interfaces.		"Models in the security context could include: * Security policies and procedures for an organization * Access control records including entries and alarm events over a given period of time"
natural language processing (NLP)	An application of AI that enables machines to both process and comprehend human language in the way it is written.		Security officers asking systems to "Show me all the security breaches at our headquarters from last week" would prompt the system to turn that question into a query based on the parameters contained in the prompt, and would return an answer to the user in a similarly structured format. A prompt such as "Show me all the times in the last month that a person on a green motorcycle entered one of our facilities" would be processed to query the associated video surveillance systems to search for this occurrence and would return clips of the results to the system user.
operational conditions	The set of factors associated with the environment where the system has been deployed that may influence its accuracy and performance.		Lighting, background noise, and weather are factors that can significantly affect the performance of AI security systems. AI systems designed for vehicle screening may be affected by road conditions, traffic and construction activities.
operational factors	Aspects of the environment where the system has been deployed that may influence its accuracy or performance.	Operational factors are said to be critical when they are integral to a system's functioning. When new critical operational factors are identified, teams need to understand if this new factor can be supported within the existing system and the degrees of support necessary.	In camera analytics, operational factors can include the amount of traffic, shadows, angle of the camera, and obstructions since each of these can affect the ability of the AI to correctly assess the environment.

Term	Definition	Note	Security Use Cases and Considerations
overreliance	The phenomenon of people accepting incorrect recommendations from an AI system (i.e., making errors of commission).	Overreliance generally happens when system users are unable to determine whether or how much they should trust the AI. Users have difficulty determining appropriate levels of trust because of lack of awareness about what the AI system can do, how well the system can perform, and how the system works.	Balacing AI capabiliities with human oversight ensures comprehensive and effective security. An AI system might not recognize a person who has legitimate access but is acting suspiciously, leading to potential security incidents. If a security team solely relies on AI-powered systems to detect intrusions, they might miss subtle signs of a breach. Another example would be if a person enters a controlled area to which they have access on their day off or outside of their normal work shift.
predictable failures	Failures that can be forecasted based on an understanding of how an AI system can be used and how the system can fail.	Issues with AI output are generally the result of data quality, bias, edge cases, environmental changes, or limitations in model complexity.	Predictable failures include things like bad camera positioning that generates false alerts. Understanding the limitations of AI can inform the best place to deploy technology. It can also inform the parameters of a specific deployment to increase the liklihood of a positive result.
pre-training	Pre-training a neural network refers to first training a model on one task or dataset, and then using the parameters or model from this training to train another model on a different task or dataset.		In most physical security use cases, the software provider will handle pre-training, training, and quality assurance. For context, the end user should be focused on the training methods used by the provider, the size of the data sets, and other details.
principles	Commitments to ensure AI systems are developed responsibly and in ways that warrant people's trust.		Organizations implement principles such as accuracy and accountability in surveillance systems, regularly auditing for biases, errors, transparency, hallucinations, etc., and training the system to reduce such risks
prompt	The text a user submits as an input to a product.		See "prompt engineering."

Term	Definition	Note	Security Use Cases and Considerations
prompt engineering	Prompt engineering is a concept in AI, particularly natural language processing (NLP). It is the process of creating prompts, or inputs, that are used to train AI models to produce specific outputs. A prompt can be as simple as a few words or as complex as an entire paragraph, and it serves as the starting point for an AI model to generate a response.		Good prompts in physical security, particularly in the use of large language models (LLMs), can include the defined user persona, the given scenario, and the desired outcome. For example, "As a real estate leader, review the available badge access data and provide me a report identifying the number of times a given employee uses door x."
prompt injection	A successful attempt to break a model out of its instructions in order to evade mitigations and produce harmful content.	Physical security relies on safeguarding AI systems against unauthorized prompt injections, which could compromise information integrity or lead to incorrect outputs.	AI systems in physical security should be designed and tested to deter and detect attempts at prompt injection. End users need to be aware of the risk and confirm that a system has internal safeguards against such practices.
quality of service	Whether an AI system works as well for one person or group of people as it does for another, even if no opportunities, resources or information are extended or withheld.		Providers should identify and promote - and end-users should vigorously test - consistency in detection across different variables (e.g. facility types and environmental conditions).
red team, red teaming	The term red teaming has historically described systematic adversarial attacks for testing software security vulnerabilities. With the rise of AI and, specifically, large language models (LLMs), the term has extended beyond traditional cybersecurity and evolved in common usage to describe many kinds of probing, testing and attacking of AI systems for the purpose of uncovering and identifying potentially harmful outputs.	Red teaming is an essential practice in the responsible development of systems and features using LLMs because the uncovering and identifying of harms enables systematic measurement and mitigation of them.	Solutions providers can use red-team testing during development to test for resilience against attacks, robustness, bias & fairness, data poisoning, and model interoperability.
release	A system being made available to customers or the public.		For more advanced systems, end users should consider the adoption of a dev/test environment where they can evaluate new releases in a non-production environment to gain confidence in the AI.

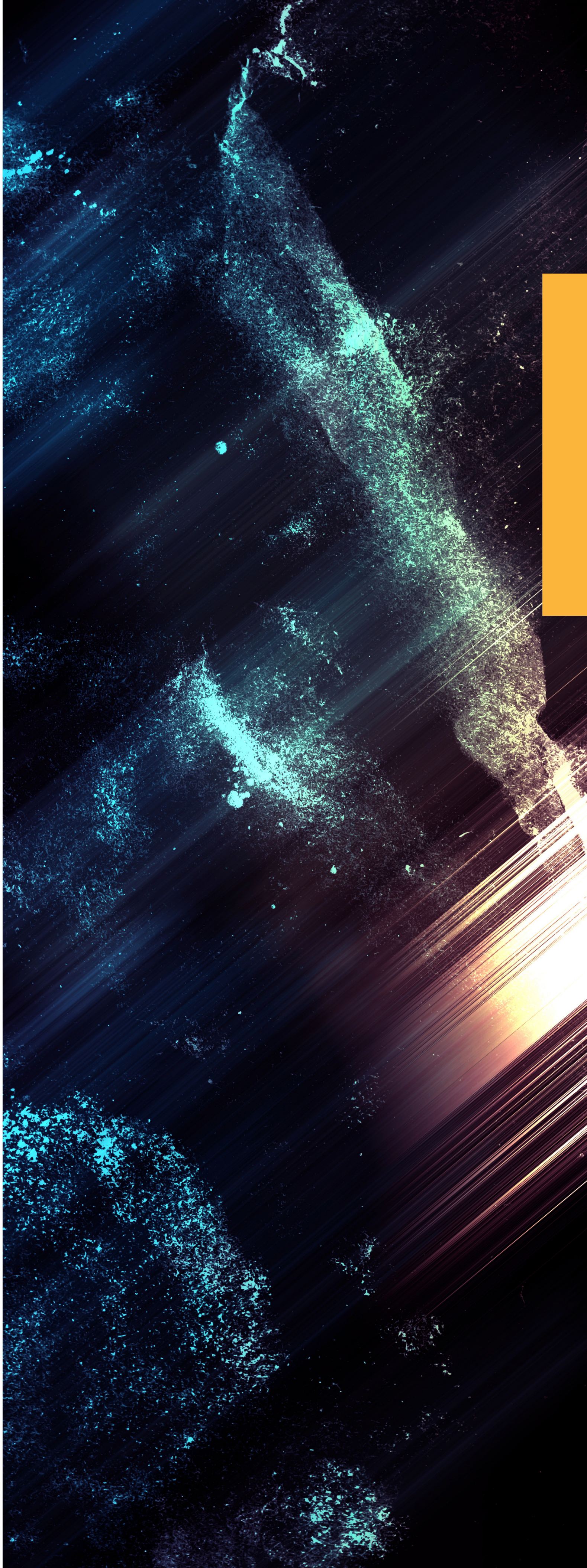
Term	Definition	Note	Security Use Cases and Considerations
reliability	The ability of a system to provide intended behavior and results.		AI systems should be able to provide statistics on False-Positive (FP), True-Positive (TP), False-Negative (FN) and True-Negative (TN). In most security settings, a false-negative would be an unacceptable outcome as it would be a missed behavior resulting in a vulnerability. TP/TN is the successful application of AI to determine the given behavior. Alert on a true door forced alarm (TP) versus resolving a nusiance alarm (TN). FP is the standard noise while FN is an actual security incident inadvertently resolved by the AI.
remediation	The steps or actions taken in response to an AI system failure.		Examples include a system designed for an end user to report and provide specific feedback on false-positives or false-negatives in the output of an AI system, (e.g., camera AI falsely detecting a "person" that it is actually a shadow or animal).
response	The text output in response to a prompt or query. Synonyms for "response" include "completion," "generation" and "answer."		The information provided by an AI-powered system to answer a query/ prompt.
responsible AI	A human-centered approach, the goal of responsible AI is to create trustworthy AI systems that benefit people while mitigating harms by using research-driven best practices.	Responsible AI is becoming the de facto way of managing the AI system lifecycle through the implementation of the ethical principles at the heart of all national and international AI regulation, as originated by the Organisation for Economic Co-operation and Development (OECD). The principles include accountability, transparency and explainability, robustness, security and safety, human rights and democratic values, including fairness and privacy, inclusive growth, sustainable development and well-being.	Without the implementation of responsible AI principles, technologies such as facial recognition can suffer from bias, lack of transparency, lack of consent, overreliance on machine "automation bias," lack of human oversight, and issues with data poisoning arising from cybersecurity issues. Remedial actions are rooted in industry guidance and clarifying cases in law, as well as new regulation, the EU AI Act which sets this software as high-risk, and mandates the use of responsible AI principles through certification. British Standard BS 9347 includes internationally interoperable processes, while the Confederation of European Security Services (CoESS) has published a charter on the use of AI in security services.

Term	Definition	Note	Security Use Cases and Considerations
restricted uses	By law or policy, some AI systems are subject to additional restrictions on their development or deployment because they pose heightened risks to people and society that cannot be mitigated adequately.	Restricted uses may arise in a range of circumstances, including where it is determined that the use of an AI system is premature because inadequate guardrails exist to sufficiently mitigate heightened risks, or where an AI system cannot address the requirements of the use case at all or is unable to do so without heightened risk that cannot be mitigated adequately.	Certain use cases, including predictive policing, recidivism and profiling on the basis of sensitive biometric data, are restricted in some jurisdictions, including the Europen Union. The security sector collects and processes vast arrays of data that could be combined in ways beyond the scope of the original intended purpose. Certain AI techniques are highly effective at recognizing patterns in data and can quickly make inferences that could combine anonymous facial mapping data with situational and other identifying information outside the scope of the security use case.
rubric	A framework that sets out criteria and standards for different levels of performance and describes what performance would look like at each level to evaluate a stakeholder's success with respect to a certain goal.		In security, the National Institute of Standards and Technology (NIST) Face Technology Evaluations may be considered a rubric for facial recognition algorithm performance relied upon by stakeholders, from developers integrating the AI to security integrators making choices regarding the performance of one vendor's system versus another.
safe	Protected from danger or risk.	In the context of AI, "safe" indicates the expectation that a system does not, under defined conditions, lead to a state in which human life or health is endangered.	In security AI, safe could refer to a system's impact on privacy, human rights and health. In the European Union, AI vendors must identify a product's safe use cases in technical documentation. In an unregulated market, safe AI may mean that an integrator has tested the system in the environment in which is will be used and informally deemed it safe to use for a given purpose.
sensitive uses	Sensitive uses are uses of AI systems that can have the potential to result in a significant adverse impact on individuals.		Sensitive uses in security are linked to restricted uses, but are generally permitted with additional safeguards to ensure that valid consent is obtained or that the risk is necessary and proportionate to the problem being solved. Sensitive can mean military/ national security, where restrictions on AI use for this purpose are outside usual commercial boundaries in law and contract. It can also mean processing of sensitive data such as biometrics, which often requires special considerations and actions to mitigate risk.

Term	Definition	Note	Security Use Cases and Considerations
stakeholder(s)	The people (or a group or category of people) who are responsible for, will use, or will be affected by a system.		In security, stakeholders typically include the entire supply chain, including the developer, manufacturer, integrator and end user. Stakeholders can also be those who have an interest in the technology and its uses, such as members of the public, local authorities, policy makers and law enforcement. Stakeholder mapping can support the consideration of all impacts and their potential severity before the implementation of a restricted of sensitive use AI system.
system health monitoring	Ongoing tracking of a system to detect errors and other abnormalities; includes telemetry data, feedback channels, alerts, and reporting.		With certain AI systems that adapt and learn from the input data presented to them, system health monitoring considerations extend beyond fixed function performance monitoring. Additional measures may be needed to monitor for unintended consequences, such as the introduction of bias or data poisoning through the ingress of rogue documents into a retrieval augmented generation (RAG) system knowledgebase.
training	Providing a machine learning model's algorithm with a given dataset for processing and identifying patterns that the model will then use to perform predictive tasks in its deployment setting.	Model providers are starting to produce model cards that explain what data has been used to train the model, together with the intended use cases. This approach is becoming a de facto way of improving model transparency and explainability.	Because of the unpredictable nature of criminal activity, both physical and cyber, the arrays of data needed to train models used for security to a reasonable level of competency is an ongoing challenge. The risk of introducing bias is high when the training data does not include a sufficiently diverse demographic. However, synthetically generated data created through the use of generative AI are enabling the expansion of training data to make it more diverse, or to extrapolate smaller data sets where it is not possible to gather sufficient data to train a model, such as with uncommon risk scenarios.

Term	Definition	Note	Security Use Cases and Considerations
transformer	An artificial neural network that can transform one sequence into another one. It consists of two parts, an encoder and a decoder, that can handle sequential data such as text, image or video. Transformer models are versatile and accurate and can be used for various applications such as language translation, text generation, image recognition, and more.		The transformer architecture is important in security as a building block of multimodal AI that can transform text to image and image to text, audio to text, text to audio, audio to image and image to audio. Asking questions about data and responding in different formats is a key feature of security AIs. Transformers underpin a more powerful architecture that can span across data sets in a way previously impossible because of challenges in interpreting unstructured data.
transparency	Transparency requires those who build and use AI systems to be forthcoming about when, why and how they choose to build and deploy them, as well as their systems' limitations.	Another facet of transparency is intelligibility, meaning people should be able to understand, monitor and respond to the technical behavior of AI systems.	In security use cases, transparency generally means that it is known that a system contains AI, and subjects are aware they are being exposed to it. This can be achieved through policy and signage with information about the use case. Because of different views about what constitutes AI, systems integrators may not be aware that a system contains AI technology and will need to ask the right questions to ensure they can pass on the relevant transparency statements to their customers.
transparency note	Transparency notes are intended to provide guidance on higher-level capabilities and limitations (beyond standard technical documentation) that can help customers build onto AI products more responsibly. They typically include information about strengths and weaknesses, technical limitations, disclosure of performance metrics, some amount of information on how a system should and should not be used, and implications of implementation choices.		Transparency notes are model cards which are disclosed by AI developers for the purpose of providing understanding about an array of parameters that an integrator of the AI will potentially need to communicate to the customer.

Term	Definition	Note	Security Use Cases and Considerations
unsuitable uses	Cases for which the system owner has determined that an AI system does not address the requirements of the use case or cannot be deployed without significant risk.	AI system owners must not develop or deploy AI systems for use cases that are unsuitable.	In security, as in other sectors, the flexibility of AI, particularly generative AI, is starating a dialogue about potential use cases and driving the need for new skills and competencies in the areas of integration and risk management. Unsuitable uses may, for example, employ multiple AI techniques to join data sets and create an unacceptable level of tracking of individuals. Use cases in which decisions should be human-centric, such as those involving biometrics, may be unsuitable for fully automated systems without a human in the loop.
unsupported uses	Unsupported uses are those uses for which an AI system designer does not advise customers to use a system or for which they prevent customers from using a system.		Unsupported uses in security may include indiscriminate mass surveillance or targeting of individuals based on sensitive data. Profiling may not be supported by a developer for ethical reasons. and the use of an AI system to perform attacks on other systems would not only be unsupported but unethical and criminal.
use case	A specific scenario in which an AI system is used.		In security, the use case is the final deployment scenario, which may be operated by a security company or in-house.



securityindustry.org