Cisco Principles for Responsible Artificial Intelligence

Our Artificial Intelligence Mission

Artificial intelligence (AI) and subdisciplines such as machine learning offer enormous positive potential for humanity, businesses, and public services that span industry sectors, economies, and societies. These technologies not only raise the bar in terms of the beneficial capabilities they offer; they also create new challenges for customers, users, and other stakeholders. Because AI can automatically generate insights that influence critical decisions and actions, it's imperative to implement clear governance over how we develop, deploy, and operate AI-based solutions.

Realizing AI's significant promise while adhering to standards for **transparency**, **fairness**, **accountability**, **privacy**, **security**, and **reliability** is an ongoing mission at Cisco. To uphold these principles, we scrutinize each of our AI offerings to identify and address potential risks.

Our Responsible AI Principles

Cisco's Responsible AI Principles and approach described below form a broad AI governance framework for anyone who develops, deploys, and uses AI capabilities. We have translated each principle into concrete working practices that appear in italics after each description.

Transparency

AI relies on large datasets and advanced algorithms. Often, it's not clear to users when and how AI is involved in decision-making. As transparency is one of our Trust Principles and core to this framework, we inform customers when AI is being used to make decisions that affect them in material and consequential ways. Customers and users can then inform us of their concerns or let us know when they disagree with decisions. By keeping communications channels open, we intend to build, maintain, and grow the trust that our customers, users, employees, and other stakeholders place in our AI offerings.

Cisco's goal is to provide clarity and consistency in informing users when AI is employed in our technologies; the intent of the AI; the model class; the data demographics; and the security, privacy, and human rights controls applied to the model in a manner that is accessible, transparent, and understandable. We also share how to get more information about our use of AI.

Fairness

AI creates the potential for harmful human bias to become ingrained or amplified by technological systems. At the same time, it presents an opportunity to better understand and mitigate harmful bias and discriminatory results in decision-making and to create technology that promotes inclusion. Achieving better decisions requires assurance that the training data represents the demographics of individuals or groups across the full spectrum of diversity to which AI will be applied.

Cisco strives to identify and remediate any harmful bias within our algorithms, training data, and applications that are directly involved in consequential decisions; that is, decisions that could have a legal or human rights impact on individuals or groups. As an integral component of our responsible AI framework, we have also developed mechanisms for our customers to provide feedback and raise any concerns for review and action by our Incident Response Team. We regularly update these practices to reflect the latest technological advancements, including those in AI.

Accountability

Accountability for AI solutions and the teams that develop them is essential to responsible development and operations throughout the AI lifecycle. AI tools often have more than one application, including unintended use cases and uses that might not have been foreseeable at the time of development. Companies that develop, deploy, and use AI solutions must take responsibility for their work in this area by implementing appropriate governance and controls to ensure that their AI solutions operate as intended and to help prevent inappropriate use.

Cisco is committed to upholding and respecting the human rights of all people, as articulated in our Global Human Rights Policy. The Cisco Responsible AI Framework requires teams to account for privacy, security, and human rights impacts from the very beginning of development through the end of the AI lifecycle. Accountability measures include requiring documentation of AI use cases, conducting impact assessments, and oversight provided by a group of cross-functional leaders.

Privacy

Applications of AI often use personal data that could impact individual privacy and civil liberties if not managed properly. When AI uses personal data or makes decisions for or about a person, privacy controls must be designed into the supporting technology to assure that personal data usage is permitted, purpose-aligned, proportional, and fair. Those controls must be maintained throughout the data and solution's lifecycle.

Cisco has built privacy engineering practices into the Cisco Secure Development Lifecycle (CSDL). These practices help ensure that we design, build, and operate privacy-enhancing features, functionality, and processes into our product and service offerings. When processing personal information, Cisco is committed to following the principles set forth in our Global Personal Data Protection and Privacy Policy, which aligns with applicable international privacy laws and standards.

Security

AI systems must be resilient and protected from malicious actors using similar secure development lifecycle controls as standard software development. Protection against security threats includes testing the resilience of AI systems for conventional as well as adversarial machine-learning attacks; sharing information about vulnerabilities and cyber-attacks; and protecting the privacy, integrity, and confidentiality of personal data.

Cisco builds AI technologies using leading security practices, drawing on our secure development lifecycle to maximize resilience and trustworthiness. To meet the unique characteristics of AI, Cisco has added specific security controls for AI that improve attack resiliency, data protection, privacy, threat modeling, monitoring, and third-party compliance.

Reliability

Across a range of AI applications, the efficacy of AI solutions is measured by how reliably that solution produces a desired output based on the data set on which it has been trained, and the data from which it continuously learns. One of the key offerings of AI solutions is increased accuracy, which can only be achieved if the solutions are systematically tested for and engineered to produce replicable results.

Cisco prioritizes innovation, and we design and test AI systems and their components for reliability. As part of our responsible AI assessment, we review AI-based solutions for embedding controls in their lifecycle to maintain consistency of purpose and intent when operating in varying conditions and use cases. Where we identify that an AI solution has potential impacts on safety, we impose additional integrity controls.

Our Purpose

Cisco has a long history of building network-based solutions as a force for good in society. As we continue to innovate across our product and service offerings with secure solutions that meet our customers' needs and deliver business value, we also strive to meet the highest standards of transparency, fairness, accountability, privacy, security, and reliability. We do this to respect human rights, encourage innovation, and reflect Cisco's purpose to power an inclusive future for all.

Learn more about our approach to Responsible AI at trust.cisco.com.