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 also create new challenges for customers, users, and other stakeholders that deploy, use, or are impacted by them. As part of our ongoing work to maximize the benefits of emerging technologies while addressing challenges and mitigating risks, Cisco has grounded our approach to AI development, deployment, and use in a set of principles and a clear governance framework.

Our Responsible AI Principles

Cisco's Responsible AI Principles and approach described below form the foundation of our AI governance framework – to build safe and trustworthy AI. We have translated each principle into concrete working practices that appear in italics after each description. Realizing AI's significant promise while adhering to standards for **transparency, fairness, accountability, privacy, security,** and **reliability** is an ongoing mission at Cisco.

Transparency

Often, it is not clear to users when and how AI is used or involved in decision-making. Transparent practices, such as informing users when AI is being used to make decisions that affect them in a material and consequential way, build trust in the use of the technology and give users more choice in how they interact with technology that leverages AI.

Cisco's goal is to inform users as appropriate when and how AI is employed in our technologies; the intent of the AI; the model class; data use and demographics; and the security, privacy, and human rights controls applied to the use 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 or stereotypes to become ingrained in 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, resulting in technology that may promote inclusion or drive more equitable outcomes. 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. Testing to promote positive outcomes across demographic, linguistic, and geographic groups is an important step in this process.

Cisco strives to develop and deploy AI systems consistent with our purpose of powering an inclusive future for all. We are committed to upholding and respecting the human rights of all people, as articulated in our Global Human Rights Policy. Cisco seeks 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.

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. Providing communication channels for users to voice their concerns increases the comfort level with AI and serves a useful function for companies to continuously improve their models.

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 providing appropriate oversight by a group of cross-functional leaders. 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. We regularly update these practices to reflect the latest technological advancements, including those in AI.

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 us 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 contain protections from malicious actors by using secure development lifecycle controls similar to those used in standard software development. Protection against security threats includes testing the resilience of AI systems against cyber-attacks; sharing information about vulnerabilities and cyber-attacks; and protecting the privacy, integrity, and confidentiality of personal data.

Cisco is building AI technologies that leverage our secure development lifecycle program to address the unique challenges of AI. We leverage our existing expertise by applying security controls 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 AI Impact Assessment, we review AI-based solutions to determine if adequate controls are embedded 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 or otherwise is potentially a higher risk use, we undertake additional validation and testing, and impose additional controls.

Our Purpose

Cisco has a long history of building technology 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 as part of our commitment 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.