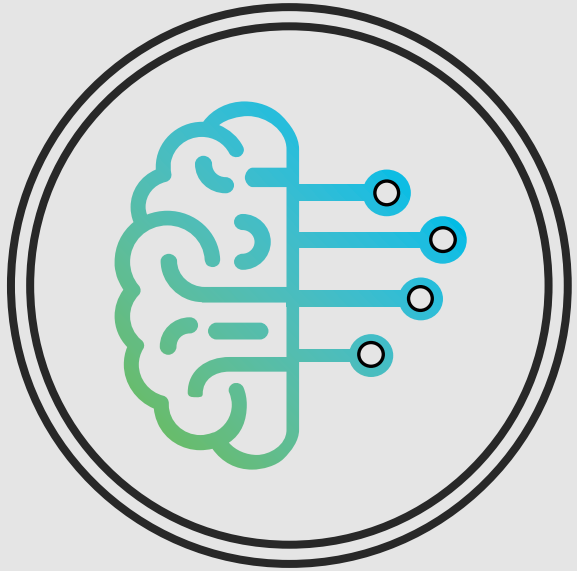
Responsible AI/ML Initiative

February 2023

Katie Shay, Associate General Counsel, Director, Human Rights

Cisco Legal

Responsible AI/ML at Cisco



Cisco's Responsible AI Principles

Artificial Intelligence (AI) can be leveraged to power an inclusive future for all. By applying this technology, we have a responsibility to mitigate its potential harms.

We translate our Responsible AI Principles into controls that can be applied to model creation and the selection of training data.

These controls embed Security, Privacy, and Human Rights by Design throughout the model's lifecycle and its application in products, services, and enterprise operations.

Transparency

Fairness

Accountability

Privacy

Security

Reliability

Cisco's Responsible AI Framework

The Responsible AI Framework operationalizes our principles throughout the company.

Governance & Oversight

Establishes a Responsible AI Committee of senior executives across Cisco business units

- Advises on responsible AI practices and oversees Responsible AI Framework adoption
- Reviews high-risk applications of AI proposed by our business units and incident reports

Controls

Embeds security, privacy, and human rights processes into AI design as part of the existing Cisco Secure Development Lifecycle

- Assesses AI applications involved in decisions that could have adverse impacts
- Applies controls to reduce risk of harm, including unintended bias mitigation, model monitoring, fairness, and transparency

Incident Management

Leverages security, data breach, and privacy incident response system to manage reported AI incidents involving bias and discrimination

- Escalates incidents to the Responsible AI Incident Response Team to address
- Tracks and reports AI incidents and remediation to governance board and other relevant stakeholders

Industry Leadership

Embeds Responsible AI as a focus area for incubation of new technology across Cisco

- Engages with industry innovation providers focused on delivering Responsible AI
- Participates proactively in industry forums to advance Responsible AI, including the Centre for Information Policy Leadership, Equal AI, and the Business Roundtable on Human Rights and AI

External Engagement

Works with governments to understand global perspectives on AI's benefits and risks

- Monitors, tracks, and influences AI-related legislation, emerging policy, and regulations
- Partners with and sponsors cutting-edge research institutions, exploring the intersection of ethics and AI from technical, organizational, social, and design perspectives

Responsible AI/ML Impact Assessments

Responsible AI/ML Impact Assessment Map

The assessment sections map to Cisco's Responsible AI Principles and controls.

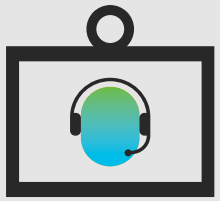
Sample Questions from the Assessment

1. *What use cases are explicitly out of scope for the AI function?*
2. *What are the implications of foreseeable product failure, misuse, or malicious attack?*
3. *Does this model generate output that results in a consequential decision affecting a user or a certain group of users?*
4. *Has this model been tested for differing outcomes by demographic category?*
5. *Does this model include a mechanism that enables appeal, override, or other actionable recourse when developers or users encounter an inaccurate output?*

Case Study: Responsible AI/ML in Webex

Responsible AI/ML in Webex

Responsible AI Impact Assessments focus on the potential impacts of intelligent product components but may not consider the cumulative impacts of those components.



Noise Removal

- **Benefits**: Noise Removal increases user privacy, representation, and comfort in meetings
- **Risks**: Pre-release models did not perform as well for higher-pitched voices
- **Remediation**: Created pitch-balanced test sets, added more high-pitch voices to training data, and expanded the subjective test suite

Virtual Backgrounds

- **Benefits**: Virtual backgrounds can increase user privacy and representation in meetings
- **Risks**: Pre-release models did not perform as well for all hair textures, hairstyles, or lighting conditions
- **Remediation**: Added more hair textures, styles, skin tones, and lighting conditions to training data

Webex Assistant

- **Benefits**: Virtual Assistants can increase meeting accessibility and efficiency in meetings
- **Risks**: Virtual Assistants may not perform as well for all languages, dialects, accents, or pitches for transcription into captions and translation. Poor transcription contributes to product inaccessibility.
- **Remediation**: Include diverse, high-quality training data appropriate for Webex's use cases

Responsible AI/ML Resources

On Cisco's Approach:

- The Cisco Responsible AI/ML Framework
- Cisco Principles for Responsible AI

On Rights Respecting AI/ML:

- Weapons of Math Destruction, Cathy O'Neil
- Sex, Race, and Robots, Dr. Ayanna Howard
- Tools and Weapons, the Promise and Peril of the Digital Age, Brad Smith

On Responsible Innovation:

- TTC Labs, Responsible Innovation Workshop Toolkits
- All Tech is Human, Responsible Tech Guide
- COMPASS EU, Responsible Innovation Self-Check Tool


CISCO