cisco Live!

Let’s go

#CiscoLive

The bridge to possible

Collaboration's Automated Decisions

Building in Security, Privacy and Human Rights by Design

Lisa Bobbitt, Principal Engineer – Privacy, Regulatory Affairs
@llbobbitt

BRKCOL-2330

Cisco Webex App

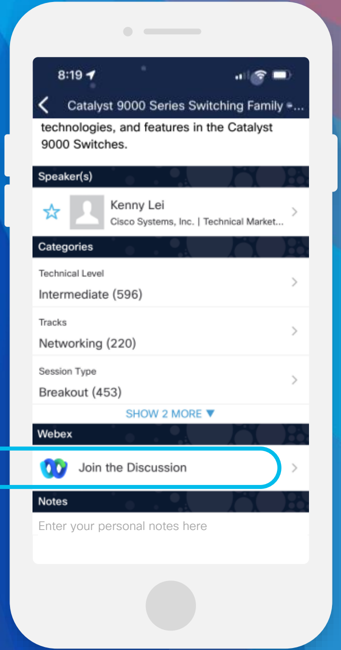
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated
by the speaker until June 9, 2023.



8:19

Catalyst 9000 Series Switching Family ...

technologies, and features in the Catalyst 9000 Switches.

Speaker(s)

Kenny Lei
Cisco Systems, Inc. | Technical Market...

Categories

Technical Level
Intermediate (596)

Tracks
Networking (220)

Session Type
Breakout (453)

SHOW 2 MORE ▼

Webex

Join the Discussion

Notes

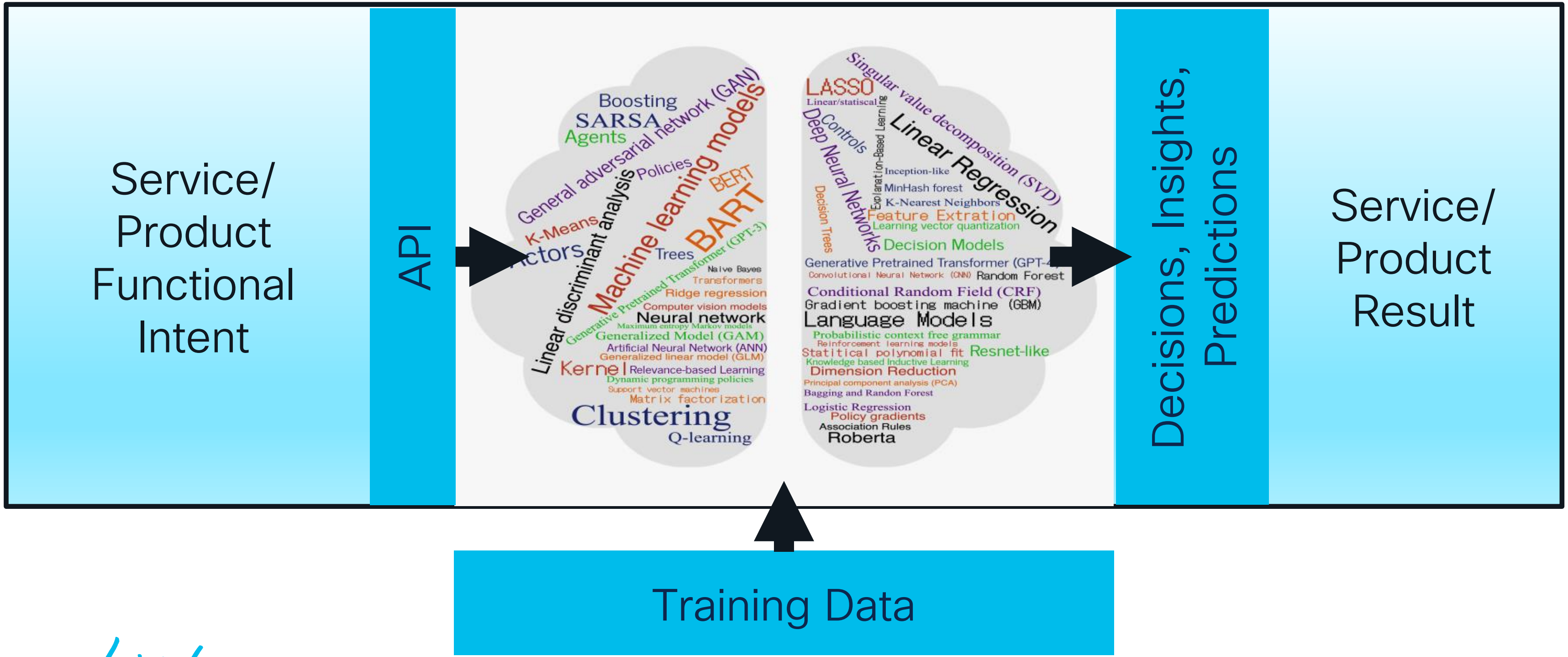
Enter your personal notes here

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKCOL-2330

Agenda

- The Need for Responsible AI
- Security, Privacy, and Human Rights Risks
- Responsible AI Principles & Framework
- Responsible AI By Design
- Applying Responsible AI to Collaboration's Automated Decisions
- The Value of Responsible AI

Artificial Intelligence



Service/ Product Functional Intent

API

Decisions, Insights, Predictions

Service/ Product Result

Training Data

Boosting
SARSA
Agents
General adversarial network (GAN)
Machine learning models
Policies
K-Means
Actors
Linear discriminant analysis
BERT
BART
Trees
Generative Pretrained Transformer (GPT-3)
Naive Bayes
Transformers
Ridge regression
Computer vision models
Neural network
Maximum entropy Markov models
Generalized Model (GAM)
Artificial Neural Network (ANN)
Generalized linear model (GLM)
Kernel
Relevance-based Learning
Dynamic programming policies
Support vector machines
Matrix factorization
Clustering
Q-learning
LASSO
Linear/statiscal
Singular value decomposition (SVD)
Controls
Deep Neural Networks
Explanation-Based Learning
Linear Regression
Inception-like
MinHash forest
K-Nearest Neighbors
Decision Trees
Feature Extraction
Learning vector quantization
Decision Models
Generative Pretrained Transformer (GPT-4)
Convolutional Neural Network (CNN)
Random Forest
Conditional Random Field (CRF)
Gradient boosting machine (GBM)
Language Models
Probabilistic context free grammar
Reinforcement learning models
Statitical polynomial fit
Resnet-like
Knowledge based Inductive Learning
Dimension Reduction
Principal component analysis (PCA)
Bagging and Randon Forest
Logistic Regression
Policy gradients
Association Rules
Roberta

What is Generative AI?

- Generative Artificial Intelligence (GAI) describes algorithms that can be used to create new content, including audio, code, images, text, video
- GAI is a Machine learning (ML) type of AI
  - AI models “learn” from data patterns (training data, inputs) without human direction
- Two common ML models used in GAI:
  1. Diffusion Models
     - For image generation tools like Stable Diffusion and Midjourney
  2. Large Language Models (LLMs)
     - For tools like ChatGPT and Copilot

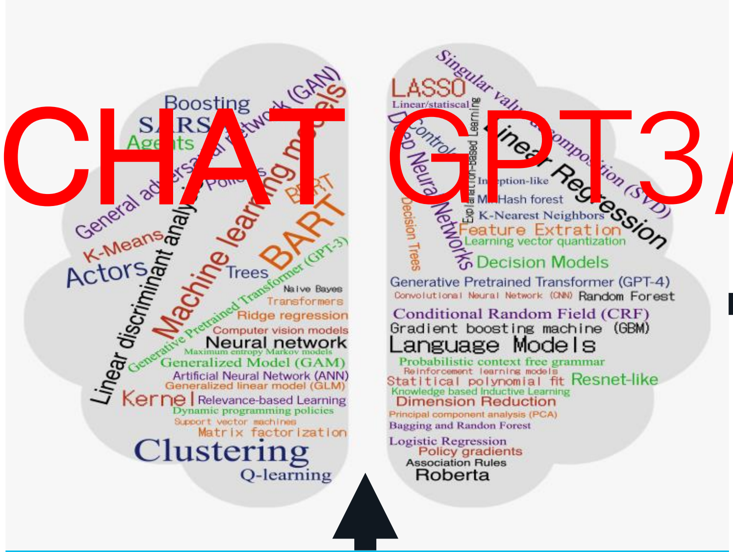
Generative Artificial Intelligence (AI) Models Democratized

UI

PROMPT

Applications

API


Boosting
SARS
Agents
General adversarial network (GAN)
K-Means
Actors
Linear discriminant analysis
Machine learning models
Policies
BERT
BART
Trees
Generative Pretrained Transformer (GPT-3)
Naive Bayes
Transformers
Ridge regression
Computer vision models
Neural network
Maximum entropy Markov models
Generalized Model (GAM)
Artificial Neural Network (ANN)
Generalized linear model (GLM)
Kernel
Relevance-based Learning
Dynamic programming policies
Support vector machines
Matrix factorization
Clustering
Q-learning
LASSO
Linear/statiscal
Singular value decomposition (SVD)
Deep Neural Networks
Control
Decision Trees
Explanation-Based Learning
Linear Regression
Inception-like
MinHash forest
K-Nearest Neighbors
Feature Extration
Learning vector quantization
Decision Models
Generative Pretrained Transformer (GPT-4)
Convolutional Neural Network (CNN)
Random Forest
Conditional Random Field (CRF)
Gradient boosting machine (GBM)
Language Models
Probabilistic context free grammar
Reinforcement learning models
Statitical polynomial fit
Resnet-like
Knowledge based Inductive Learning
Dimension Reduction
Principal component analysis (PCA)
Bagging and Random Forest
Logistic Regression
Policy gradients
Association Rules
Roberta

CHAT GPT3/4

Training Data

Data Becomes Code

010110
110010
001011

01 1 100
000 110

Code

Text

Audio

Images

Decisions Insights, Predictions

What about **Model and Training Data** Risks?

- Poor data quality
- Poor data selection
- Wrong outputs
- Instability
- Lack of reproducibility
- Improper application
- Confirmation bias
- Concept drift/off-label use
- Inadequate consideration of assumptions and limitations

What about **Business** Risks?

- Exposure of customer/partner confidential data
- Exposure of company confidential data
- IP infringement (code)
- Copyright infringement (images and text)
- Loss of patents rights or copyright
- Open-Source contamination

What about **User** Risks?

- Violation of privacy laws
- Risks to Human Rights
- Security vulnerabilities
- Accuracy and safety issues
- Lack of transparency/ understanding
- Lack of accountability

Webex’s Automated Decisions

- Webex has multiple functions that make automated decisions for an individual to enhance the collaborative experience **if** the individual wants to use them

Background Noise Reduction

Benefits: Noise Removal increases user privacy, representation, and comfort in meetings

Webex Assistant and Translation

Benefits: Virtual Assistants can increase meeting accessibility and efficiency in meetings

Virtual Backgrounds

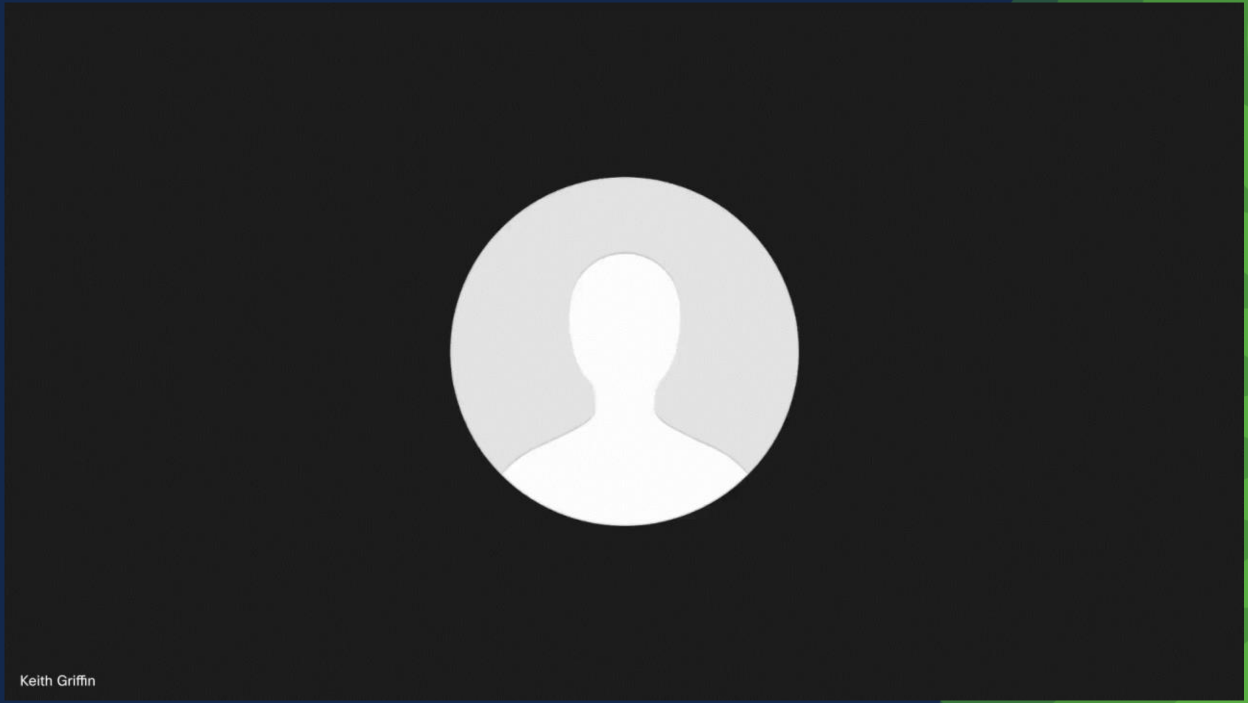
Benefits: Virtual backgrounds can increase user privacy and representation in meetings

Facial Recognition

Benefits: Facial Recognition can increase identification of the speaker, aiding collaboration and representation in meetings

Webex Contact Center and Connect

Benefits: Actionable Insights, Code Automation, Automated Chat Summaries



For a deeper look at AI in Collaboration, check out Keith’s BRKCOL-1871 Cognitive Collaboration

Potential Risks of Collaboration Automated Decisions

Business

- ✓ Exposure of customer/partner confidential data
- ✓ Exposure of company confidential data
- ✓ Training data copyright infringement (images, video, audio, and text)
- ✓ IP infringement (code)

Model and Training Data

- ✓ Poor data quality
- ✓ Poor data selection
- ✓ Wrong outputs
- ✓ Instability
- ✓ Improper application
- ✓ Confirmation bias
- ✓ Concept drift/off-label use

Webex Usage

- ✓ Risks to privacy
- ✓ Risks to human rights
- ✓ Security vulnerabilities
- ✓ Lack of transparency/ understanding
- ✓ End users combined risks
- ✓ Sensitive data environments: Education, Healthcare, Justice
- ✓ Hybrid experiences – work, family, and home exposure

Responsible AI/ML

Cisco's Responsible AI Principles

Transparency

Fairness

Accountability

Privacy

Security

Reliability

Cisco’s Responsible AI Principles (1/6)

Transparency

- Cisco’s goal is to provide clarity and consistency in informing users about our application of AI in a manner that is accessible, transparent, and understandable. This includes:
  - When AI is employed in our technologies
  - The intent of the AI and its model class
  - The data demographics
  - Security, privacy and human rights controls applied to the model
  - How to get more information about our use of AI

Cisco's Responsible AI Principles (2/6)

Fairness

- Cisco strives to identify and remediate harmful bias within our algorithms, training data, and applications that are directly involved in consequential decisions.
- Consequential decisions are those that could have a legal or human rights impact on individuals or groups.
- We have developed mechanisms for our customers to provide feedback and raise any concerns for review and action by our Incident Response Team.

Cisco's Responsible AI Principles (3/6)

Accountability

- 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.
- Cisco is committed to upholding and respecting the human rights of all people, as articulated in our Global Human Rights Policy.
- Accountability measures include requiring documentation of AI use cases, conducting impact assessments, and oversight provided by a group of cross-functional leaders.

Cisco's Responsible AI Principles (4/6)

Privacy

- Cisco has built privacy engineering practices into the Cisco Secure Development Lifecycle (CSDL) to design, build, and operate privacy-enhancing features, functionality, and processes into our offers. These apply to training data, prompts, and results.
- **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.**

Cisco's Responsible AI Principles (5/6)

Security

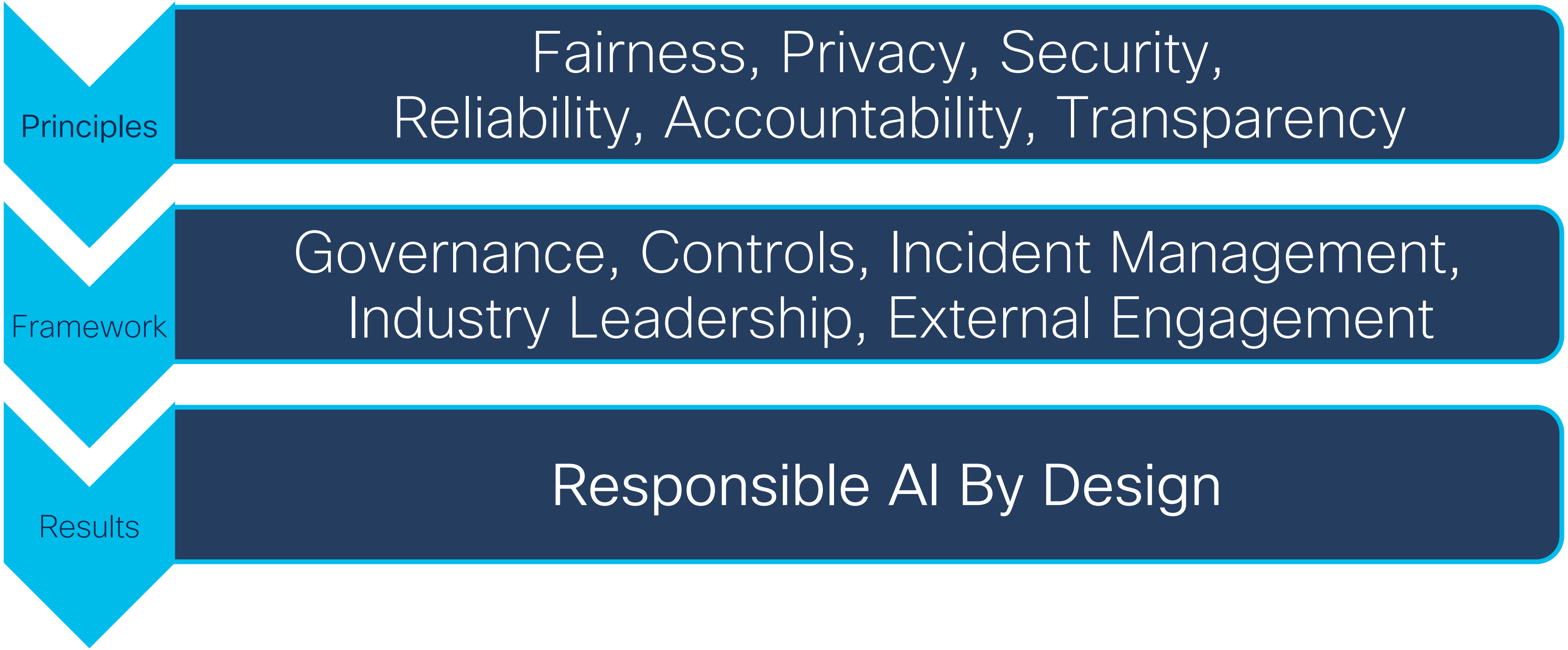
- Cisco builds AI technologies using leading security practices, drawing on our secure development lifecycle to maximize resilience and trustworthiness.
- To meet the unique characteristics of AI, Cisco has added specific security controls for AI that improve attack resiliency, data protection, privacy,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

Cisco’s Responsible AI Principles (6/6)

Reliability

- Cisco designs and tests 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.

Cisco’s Responsible AI



Principles

Fairness, Privacy, Security, Reliability, Accountability, Transparency

Framework

Governance, Controls, Incident Management, Industry Leadership, External Engagement

Results

Responsible AI By Design

Cisco’s Responsible AI Framework

The Responsible AI Framework operationalizes our principles throughout the company.

Governance & Oversight

Industry Leadership

Controls

External Engagement

Incident Management

Governance & Oversight

- Establishes a Responsible AI Committee of senior executives
- Advises on Responsible AI practices and oversees Responsible AI Framework adoption
- Reviews high-risk applications of AI proposed by business units and incident reports

Controls

- **Embeds legal, security, privacy, and human rights processes as part of the existing Cisco Secure Development Lifecycle into**
  - **Internally designed AI models**
  - **3rd-party models**
  - **Selection of training data**
  - **Tracking of use**
- Assesses applications embedding AI for adverse impacts to
  - Individuals and/or groups of people
  - Customers
  - Cisco
- Applies to reduce risk of harm, including legal, unintended bias, privacy, model monitoring, 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

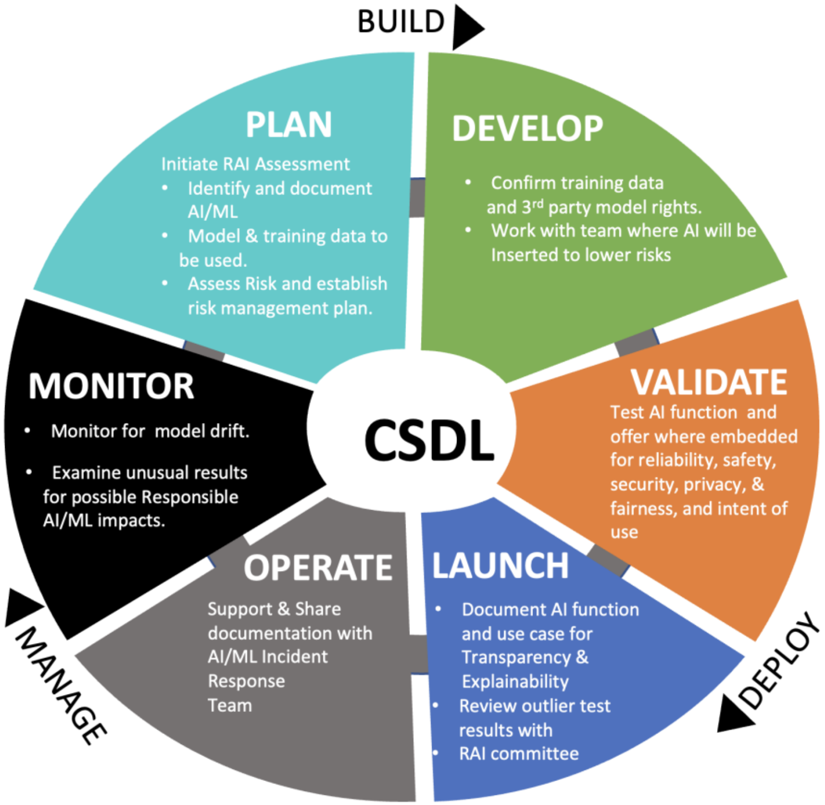
- **Embeds Responsible AI as a focus area for incubation of new technology across Cisco**
- Engages with industry innovation providers focused on delivering Responsible AI
- Participates proactively in industry forums to advance Responsible AI, including the
  - Centre for Information Policy Leadership,
  - Equal AI, and
  - 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

Applying Responsible AI/ML By Design

Responsible AI By Design



PLAN

Initiate RAI Assessment

- Identify and document AI/ML
- Model & training data to be used.
- Assess Risk and establish risk management plan.

DEVELOP

- Confirm training data and 3rd party model rights.
- Work with team where AI will be Inserted to lower risks

VALIDATE

Test AI function and offer where embedded for reliability, safety, security, privacy, & fairness, and intent of use

LAUNCH

- Document AI function and use case for Transparency & Explainability
- Review outlier test results with
- RAI committee

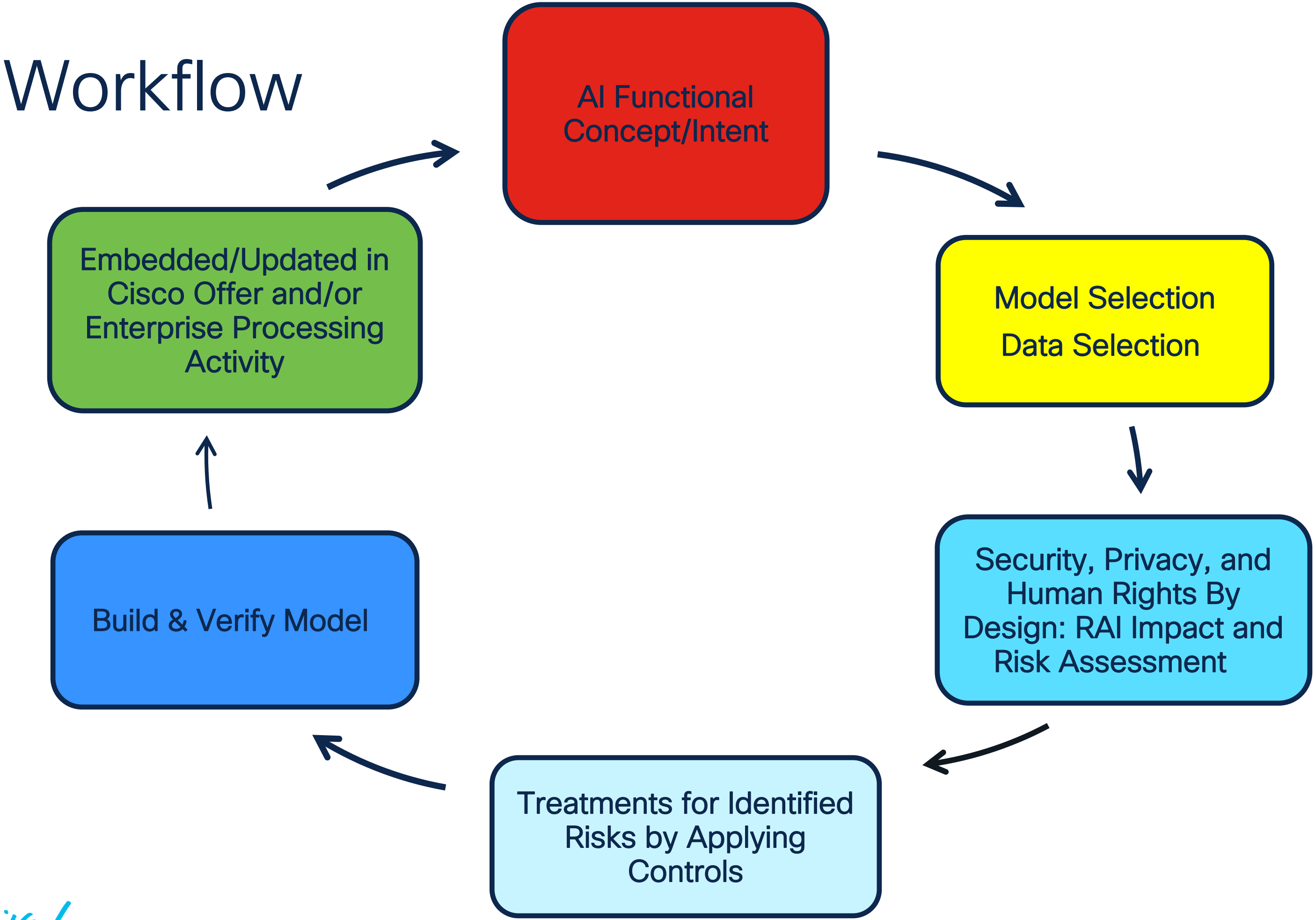
OPERATE

Support & Share documentation with AI/ML Incident Response Team

MONITOR

- Monitor for model drift.
- Examine unusual results for possible Responsible AI/ML impacts.

RAI Workflow



AI Functional Concept/Intent

Model Selection

Data Selection

Security, Privacy, and Human Rights By Design: RAI Impact and Risk Assessment

Treatments for Identified Risks by Applying Controls

Build & Verify Model

Embedded/Updated in Cisco Offer and/or Enterprise Processing Activity

Responsible AI Assessment

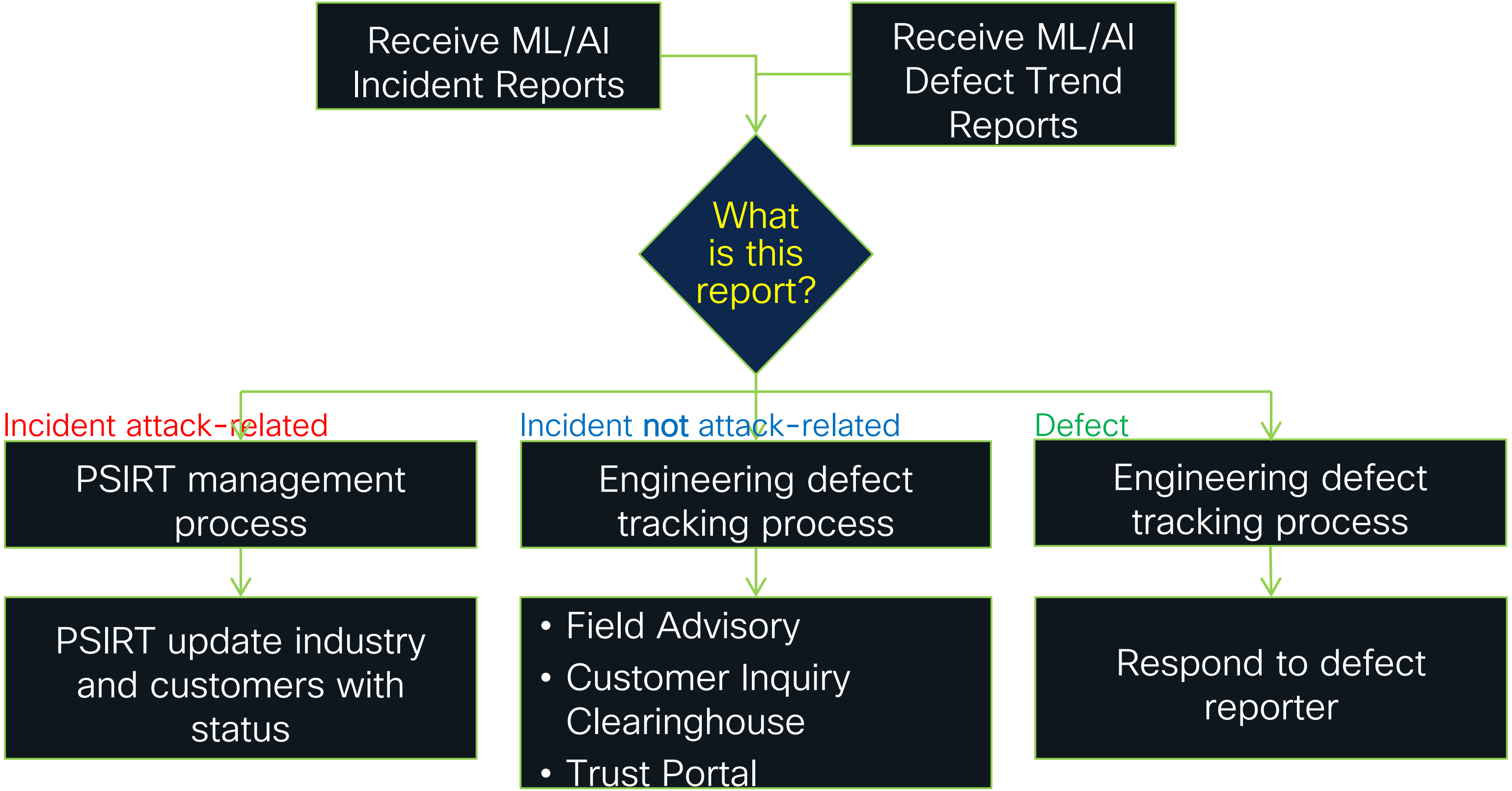
Risk-Based Assessment with Cloud Control Framework RAI Controls to Lower Risk

Use Cases	Intended & Unintended Use
Model Info	Internal, 3rd Party, Rights & Permissions
Training Data	Data Origin, Content, Retention, Aggregation, Labeling
Identify Risks & Apply Treatments	Legal, Privacy, Fairness, Security, Reliability, Transparency, Accountability

Sample Questions from the Assessment

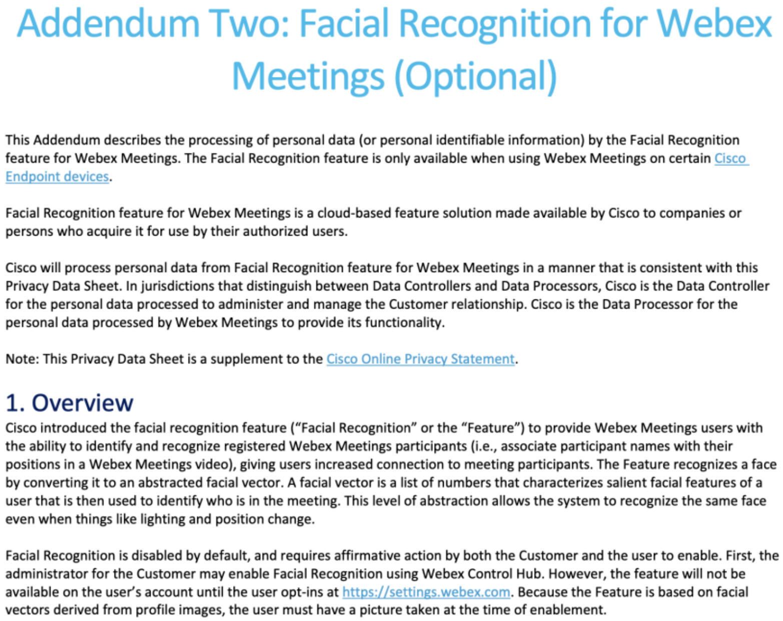
1. What is your intent for this function?
2. Do you have legal and commercial use rights?
3. What use cases are explicitly out of scope?
4. Does this model generate output that results in a consequential decision affecting a user or a certain group of users?
5. Has this model been tested for differing outcomes by demographic category?
6. Does this model include a mechanism or process that enables feedback from a user?

AI/ML Incident Response


Receive ML/AI Incident Reports
Receive ML/AI Defect Trend Reports
What is this report?
Incident attack-related
Incident **not** attack-related
Defect
PSIRT management process
Engineering defect tracking process
Engineering defect tracking process
PSIRT update industry and customers with status
• Field Advisory
• Customer Inquiry Clearinghouse
• Trust Portal
Respond to defect reporter

Applying Responsible AI/ML into Collaboration

Transparency -Webex AI Addendum to Privacy Data Sheet on trustportal.cisco.com



Addendum Two: Facial Recognition for Webex Meetings (Optional)

This Addendum describes the processing of personal data (or personal identifiable information) by the Facial Recognition feature for Webex Meetings. The Facial Recognition feature is only available when using Webex Meetings on certain Cisco Endpoint devices.

Facial Recognition feature for Webex Meetings is a cloud-based feature solution made available by Cisco to companies or persons who acquire it for use by their authorized users.

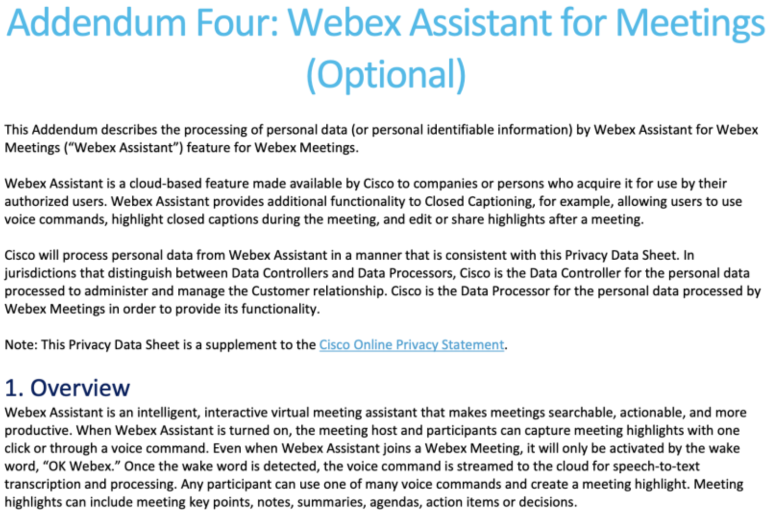
Cisco will process personal data from Facial Recognition feature for Webex Meetings in a manner that is consistent with this Privacy Data Sheet. In jurisdictions that distinguish between Data Controllers and Data Processors, Cisco is the Data Controller for the personal data processed to administer and manage the Customer relationship. Cisco is the Data Processor for the personal data processed by Webex Meetings to provide its functionality.

Note: This Privacy Data Sheet is a supplement to the Cisco Online Privacy Statement.

1. Overview

Cisco introduced the facial recognition feature ("Facial Recognition" or the "Feature") to provide Webex Meetings users with the ability to identify and recognize registered Webex Meetings participants (i.e., associate participant names with their positions in a Webex Meetings video), giving users increased connection to meeting participants. The Feature recognizes a face by converting it to an abstracted facial vector. A facial vector is a list of numbers that characterizes salient facial features of a user that is then used to identify who is in the meeting. This level of abstraction allows the system to recognize the same face even when things like lighting and position change.

Facial Recognition is disabled by default, and requires affirmative action by both the Customer and the user to enable. First, the administrator for the Customer may enable Facial Recognition using Webex Control Hub. However, the feature will not be available on the user's account until the user opt-ins at https://settings.webex.com. Because the Feature is based on facial vectors derived from profile images, the user must have a picture taken at the time of enablement.



Addendum Four: Webex Assistant for Meetings (Optional)

This Addendum describes the processing of personal data (or personal identifiable information) by Webex Assistant for Webex Meetings ("Webex Assistant") feature for Webex Meetings.

Webex Assistant is a cloud-based feature made available by Cisco to companies or persons who acquire it for use by their authorized users. Webex Assistant provides additional functionality to Closed Captioning, for example, allowing users to use voice commands, highlight closed captions during the meeting, and edit or share highlights after a meeting.

Cisco will process personal data from Webex Assistant in a manner that is consistent with this Privacy Data Sheet. In jurisdictions that distinguish between Data Controllers and Data Processors, Cisco is the Data Controller for the personal data processed to administer and manage the Customer relationship. Cisco is the Data Processor for the personal data processed by Webex Meetings in order to provide its functionality.

Note: This Privacy Data Sheet is a supplement to the Cisco Online Privacy Statement.

1. Overview

Webex Assistant is an intelligent, interactive virtual meeting assistant that makes meetings searchable, actionable, and more productive. When Webex Assistant is turned on, the meeting host and participants can capture meeting highlights with one click or through a voice command. Even when Webex Assistant joins a Webex Meeting, it will only be activated by the wake word, "OK Webex." Once the wake word is detected, the voice command is streamed to the cloud for speech-to-text transcription and processing. Any participant can use one of many voice commands and create a meeting highlight. Meeting highlights can include meeting key points, notes, summaries, agendas, action items or decisions.

Fairness Applied to Collaboration

- Assessments of the model, its development, and its production environment for consequential decisions and result affecting human rights and privacy
  - Men’s versus women’s voice range
  - Head coverings and hair styles
  - Culturally inappropriate results from generative prompts

- User control of the user of the AI functions
  - Company determines if the capability is turned on
  - End user turns it on for themselves

Webex AI Accountability

Design Accountability

- Webex AI Facial Recognition is only used in the Collaboration Products as the end-user has control of its use
- Each data set used for training of Webex AI/ML go through a review with both the Privacy Office and Product Legal

Operational Accountability

Webex AI responses to feedback when end-user experiences do not align to their expectations

Legal and Privacy Review of Training Data

Webex Example

- Review includes
  - Legal use of the data
  - Commercial use of the data
  - Review of PII that is in the data set
    - Recommendations of minimization of PII through De-Identification, Anonymization, and Deletion
    - Review of security and access of the training data

Webex Security

- Webex is designed and developed via the Cisco Secure Development Lifecycle
- Review includes
  - Operational Security
  - Platform Security
  - Secure Data Storage
  - Secure Data In Transit
  - Access management of customers use including all AI functions

For security details:
www.cisco.com/c/en/us/products/collateral/conferencing/webex-meeting-center/white-paper-c11-737588.html

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.

Background Noise Reduction

- **Benefits**: Noise Removal increases user privacy, representation, and comfort in meetings
- **Risks**: Early 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**: Early models did not perform as well for all hair textures, hairstyles, skin tones or lighting conditions
- **Remediation**: Added more hair textures, styles, skin tones, and lighting conditions to training data

Responsible AI/ML in Webex

Responsible AI Impact Assessments focus on the risks of the AI

Webex Assistant

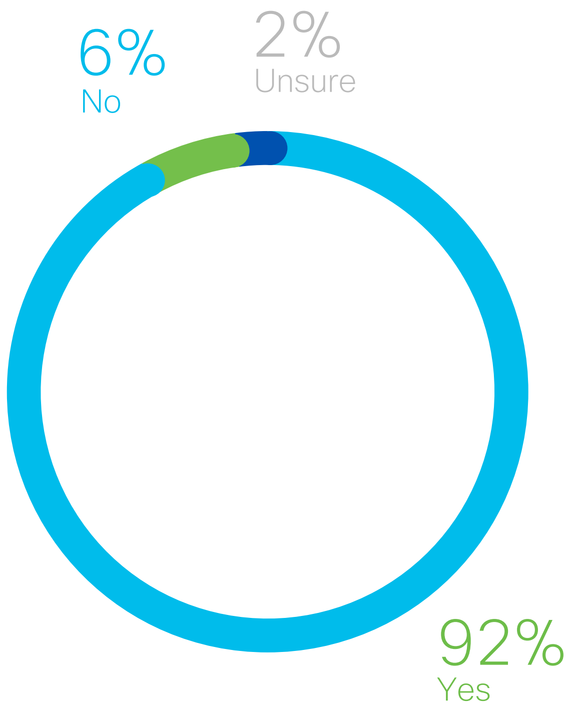
- **Benefits**: Virtual Assistants can increase meeting accessibility and efficiency in meetings
- **Risks**: Virtual Assistants may not perform as well for all languages, dialects, accents, or pitches for transcription into captions and translation. Poor transcription contributes to product inaccessibility.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

Facial Recognition

- **Benefits**: Individuals can be identified in a meeting without maintaining their image
- **Risks**: Use of facial recognition for other purposes
- **Remediation**: Limited access to the function to only Webex

The Value of Applying Responsible AI/ML

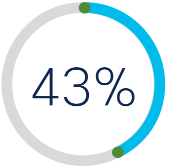
Does your organization need to do more to reassure customers about their data for AI?



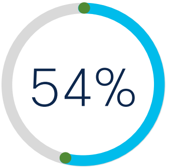
Source: Cisco 2023 Data Privacy Benchmark Study

What about using AI and personal data?

Support for AI Use

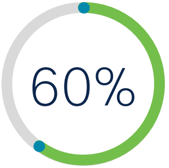


Believe AI can be useful in improving our lives

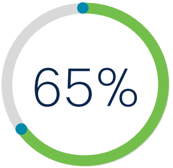


Willing to share anonymized personal data to improve AI products

Concerns About Current AI Use



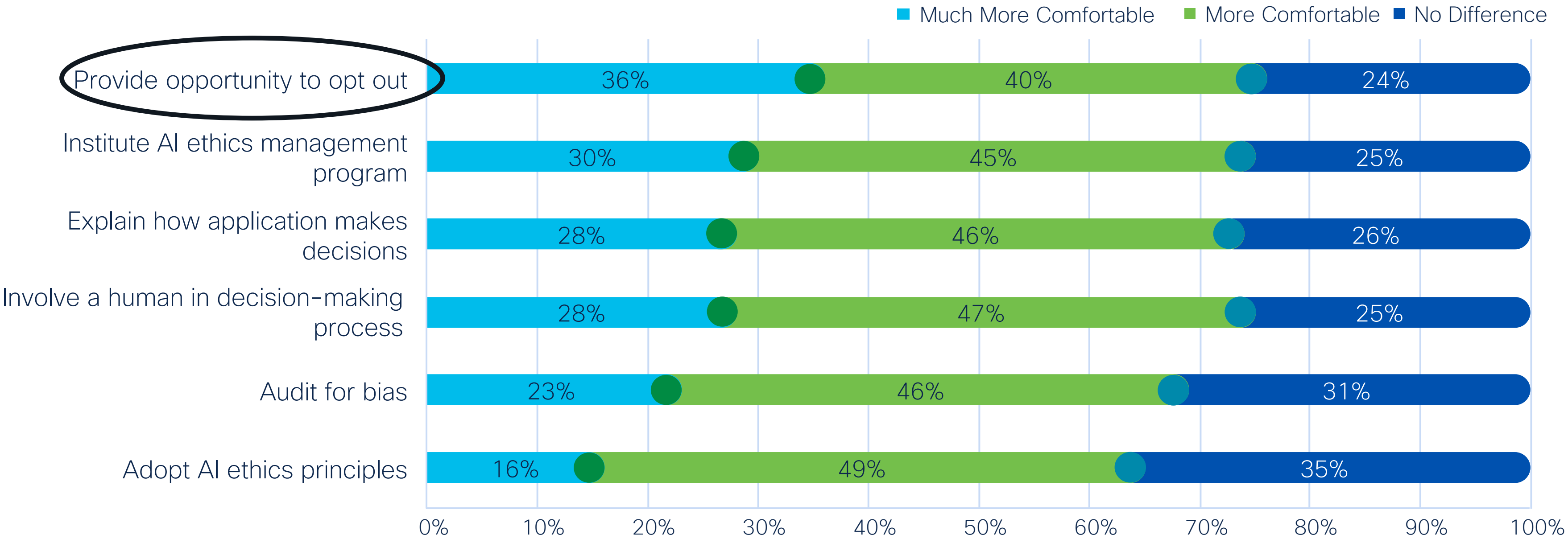
Concerned about the business use of AI today



Use of AI by organizations has already eroded trust in them

Source: Cisco 2022 Consumer Privacy Survey

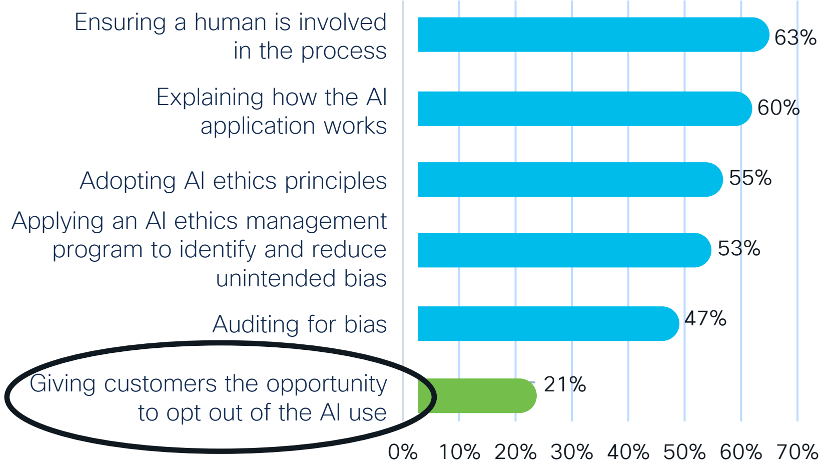
What makes consumers more comfortable with AI



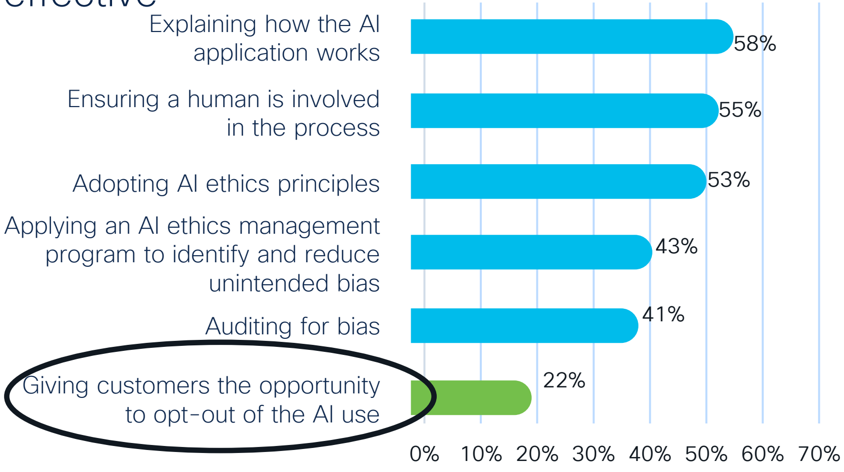
Source: Cisco 2022 Consumer Privacy Survey

What organizations are doing

What organizations have done

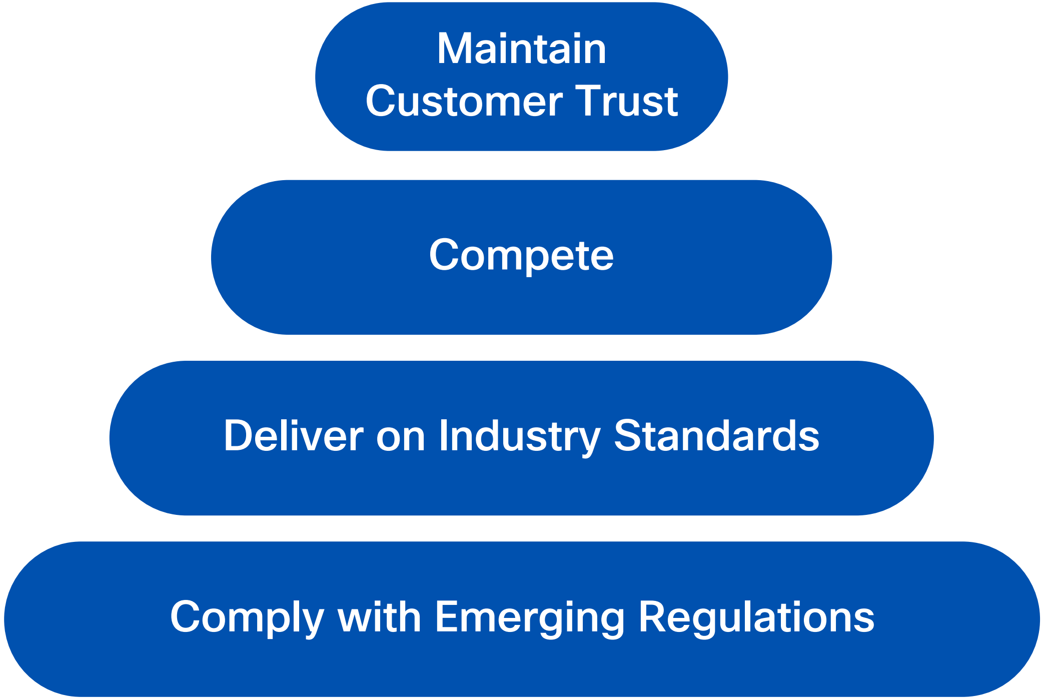


What organizations say would be most effective



Source: Cisco 2023 Data Privacy Benchmark Study

Responsible AI Benefits



Maintain Customer Trust

Compete

Deliver on Industry Standards

Comply with Emerging Regulations

Responsible AI/ML Resources

- The Cisco Responsible AI/ML Framework
  www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf
- Cisco Principles for Responsible AI
  www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-principles.pdf?CCID=cc000742&DTID=odicdc000016
- Cisco 2022 Consumer Privacy Survey
  www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-consumer-privacy-survey-2022.pdf?CCID=cc000160&DTID=esootr000515&OID=wprsc030156
- Transparency Is Key: Introducing Cisco Responsible AI
  www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-introducing-responsible-ai.pdf
- Webex Meeting on Cisco Trust Portal Privacy Data Sheet – Addendum for AI Functions
  trustportal.cisco.com/c/r/ctp/trust-portal.html#/1554085468927155

Fill out your session surveys!

Attendees who fill out a minimum of four session surveys and the overall event survey will get **Cisco Live-branded socks** (while supplies last)!

Attendees will also earn 100 points in the **Cisco Live Game** for every survey completed.

These points help you get on the leaderboard and increase your chances of winning daily and grand prizes

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Thank you

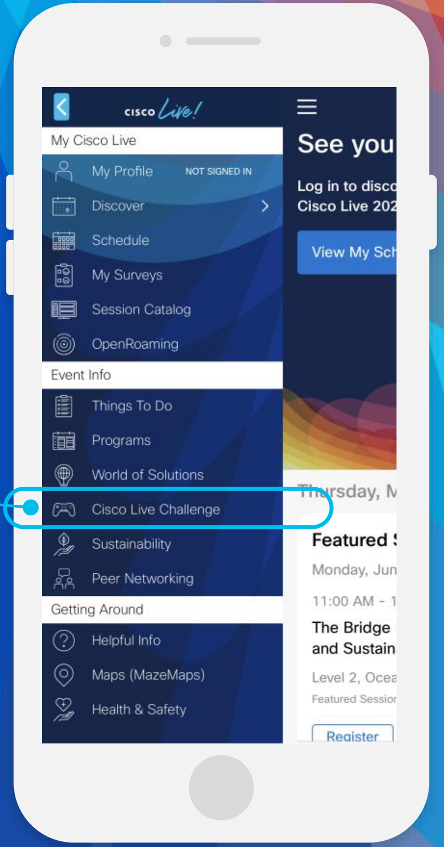
Cisco Live **Challenge**

Gamify your Cisco Live experience!

Get points for attending this session!

How:

1. Open the Cisco Events App.
2. Click on 'Cisco Live Challenge' in the side menu.
3. Click on View Your Badges at the top.
4. Click the + at the bottom of the screen and scan the QR code:



cisco Live!

Let's go

#CiscoLive