Collaboration AI

From supervised machine learning to agentic AI

Keith Griffin PhD
Cisco Fellow VP, @techkeith

Session ID: BRKCOL-1661

Cisco Webex App

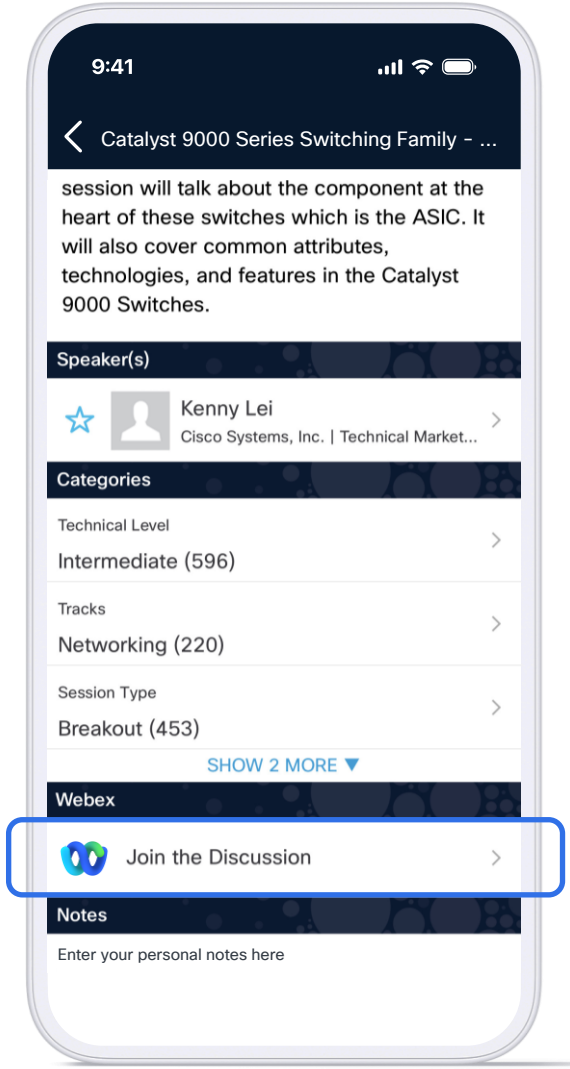
Questions?

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 14 November 2025.



https://ciscolive.ciscoevents.com/
ciscolivebot/**#BRKCOL-1661**

Agenda

01 Collaboration AI Strategy

02 Background: Supervised to Agentic

03 Collaboration AI Reference Features

04 Collaboration AI Design Approach

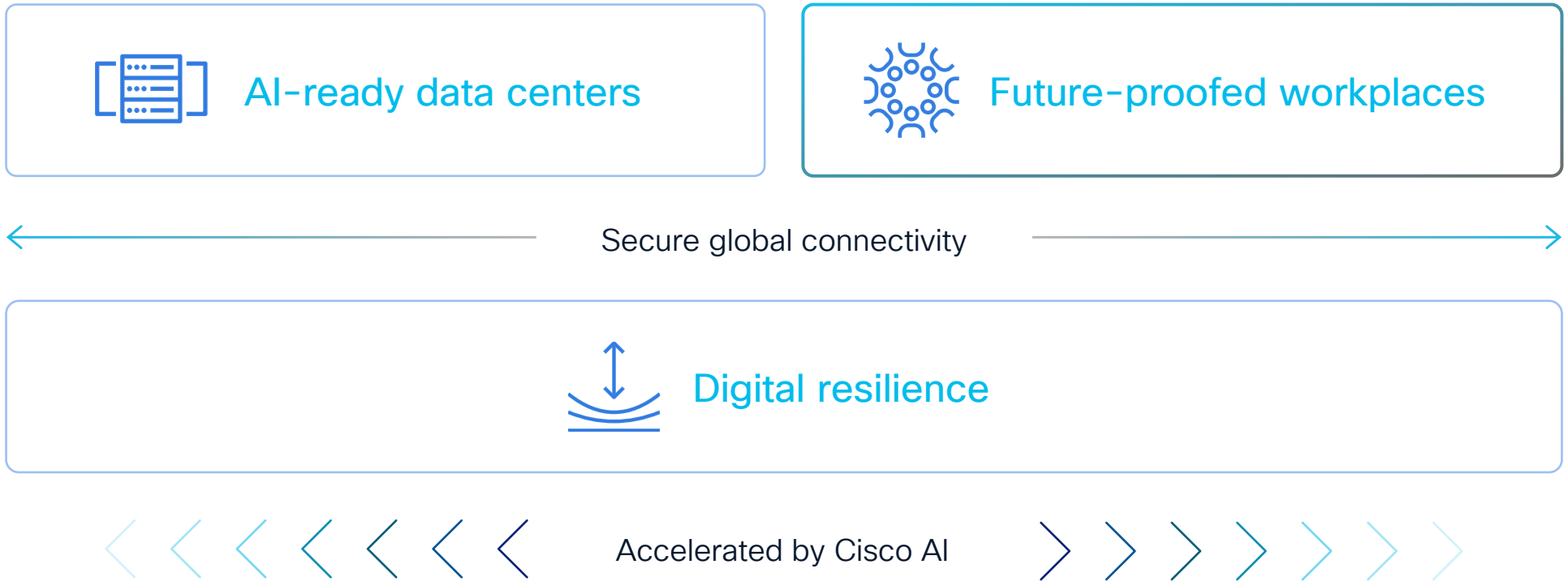
05 AI Architecture Services

06 Responsible AI

07 Conclusion

Collaboration AI Strategy

Cisco powers how people and technology work together across the physical and digital worlds



CUSTOMER EXPERIENCE

EMPLOYEE EXPERIENCE

Webex Contact Center & Webex Connect

Webex Suite

Cisco Devices

Artificial Intelligence

Security

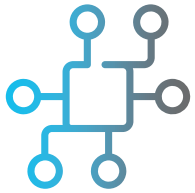
Manageability

Interoperability

AI-POWERED PLATFORM

cisco

Our strategy: Enabling everyone with purpose-built AI



Conversational AI

Audio & Speech AI

Video and Camera AI

Webex provides tangible, differentiated value through our purposeful application of AI

Collaboration is powered by connected, intelligent experiences

AI at the edge

AI in the cloud

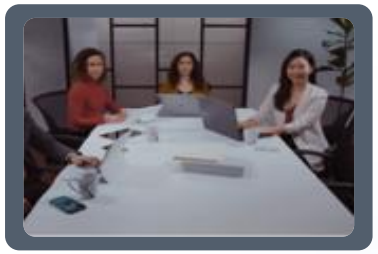
AI in control

Background: Supervised to Agentic

AI Innovation in Collaboration

2013

HEAD DETECTION



2015

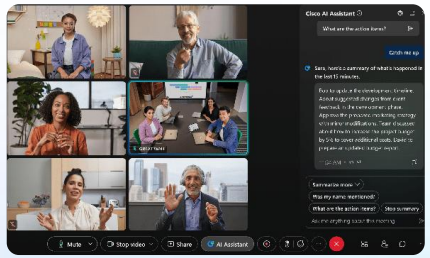
NVIDIA COMPUTING

2023

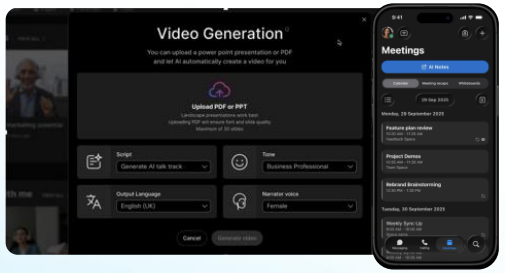
CINEMATIC MEETINGS

2024

AI SUMMARIES, SEARCH, ACTION ITEMS

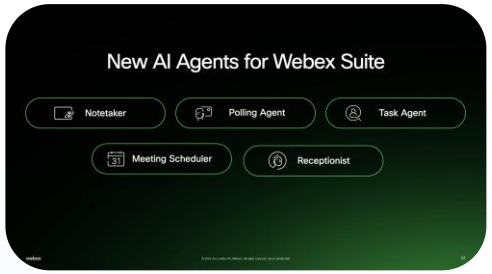


AI-GENERATED VIDCASTS, POLLS

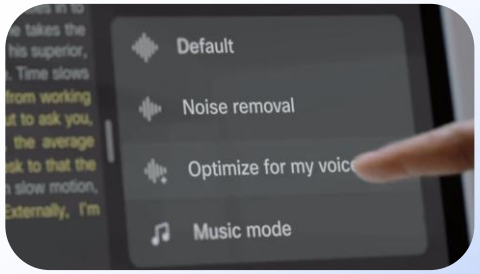


2025

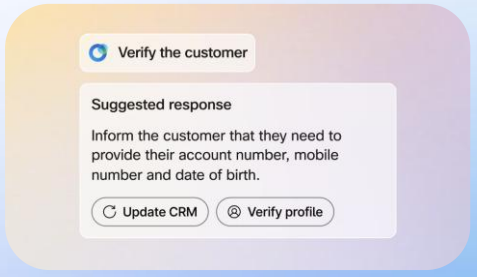
AI NOTES



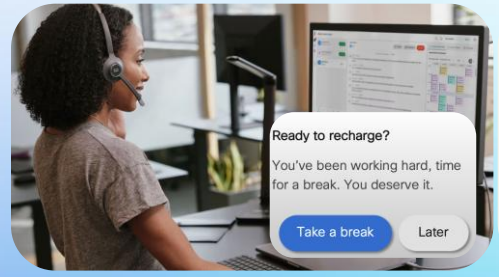
AI AGENTS & AI Workflows



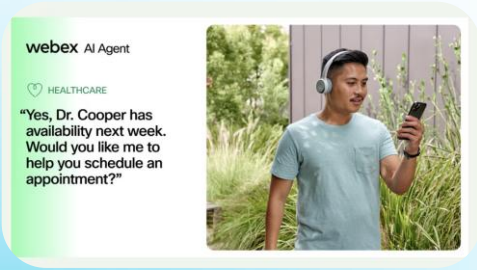
AUDIO INTELLIGENCE



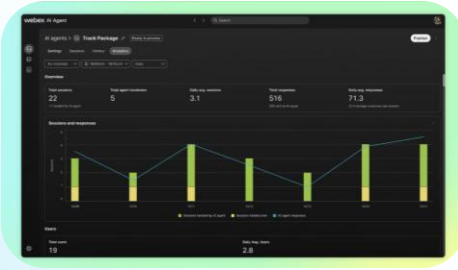
SUGGESTED RESPONSES



CISCO AI ASSISTANT



AI AGENTS



AI QUALITY MANAGEMENT

BRKCOL-1661

Supervised ML -> Generative AI -> Agentic AI

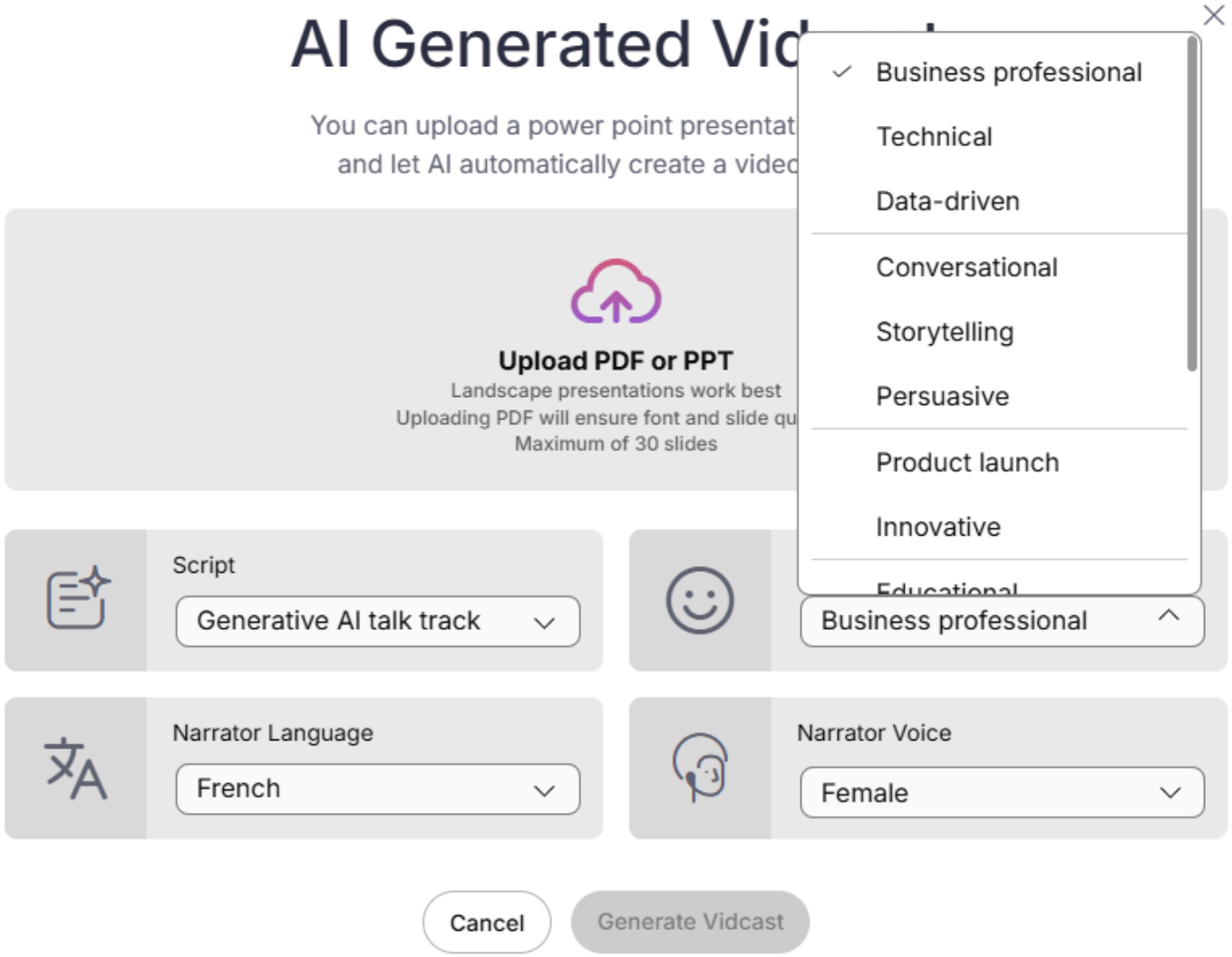
- **Definition:** Agentic AI refers to artificial intelligence systems capable of autonomous action and decision-making, pursuing complex goals independently with minimal human intervention.
- **Key Differentiators:**
  - **Proactive & Adaptive:** Unlike traditional AI that responds to commands or generative AI that creates content based on prompts, Agentic AI actively seeks goals, adapts to new information, and executes multi-step tasks.
  - **Continuous Feedback Loop:** Orchestrates a continuous feedback loop that allows the AI to adapt and execute complex, multi-step tasks, moving beyond one-shot responses.
  - **LLMs as a "Brain":** Utilizes Large Language Models (LLMs) as a core reasoning engine, but extends their capability by applying generative outputs toward specific goals and interacting with external tools.
  - **Emerging protocols:** Useful agentic protocols such as MCP and A2A are emerging. Security is critical.

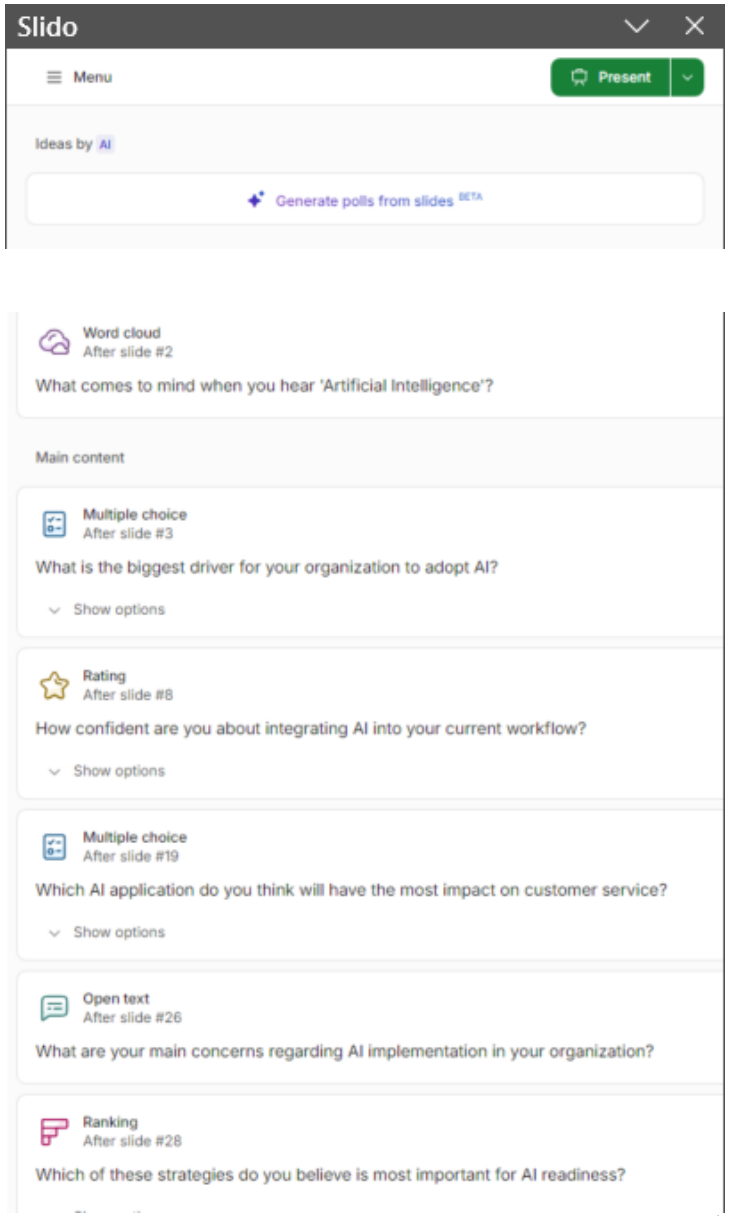
Agentic AI – Single and Multi Agent

- **Single-Agent Architecture:**
  - **Description:** A single AI system functions independently, making decisions and taking actions without the involvement of other agents.
  - **Use Case:** Preferable for well-defined problems or processes requiring a faster, focused solution.
- **Multi-Agent Architecture:**
  - **Description:** Involves multiple AI systems interacting with each other, collaborating, and coordinating their actions to achieve common goals.
  - **Sub-types:**
    - **Vertical Architecture:** Agents organized in a hierarchical structure, with higher-level agents overseeing lower-level ones.
    - **Collaborative/Distributed:** Agents adapt roles dynamically, enabling parallel processing and handling separate subtasks simultaneously.
  - **Challenges:** Coordination, communication protocols, synchronization, and negotiation mechanisms add complexity.
  - **Use Case:** Highly flexible and scalable for complex scenarios requiring diverse skill sets, such as workflow optimization or AI-driven analysis platforms.

Collaboration AI Reference Features

Language Model Demos





The Next Era of AI-Powered Collaboration

Connected . Agentic . Secure

AI Agents

- Notetaker
- Polling Agent
- Task Agent
- Meeting Scheduler
- Receptionist
- Translator

AI Ecosystem

Copilot
Outlook
Agentforce
slack
Jira
salesforce
Glean
Amazon Q

Security

Agentic Ops

E2E Encryption

Webex Compliance Hub

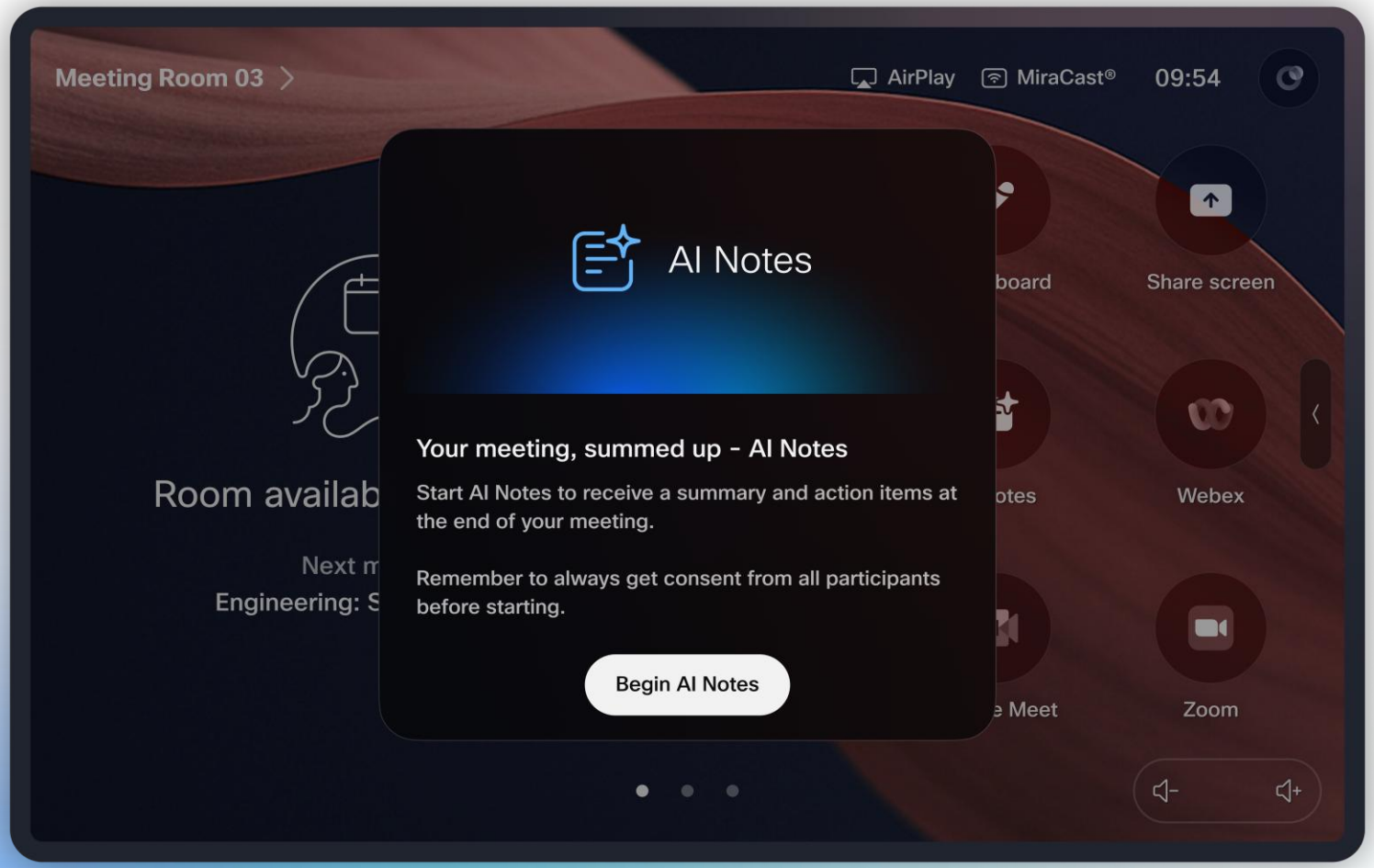
THETALAKE

cisco

AI Notes

Notetaker

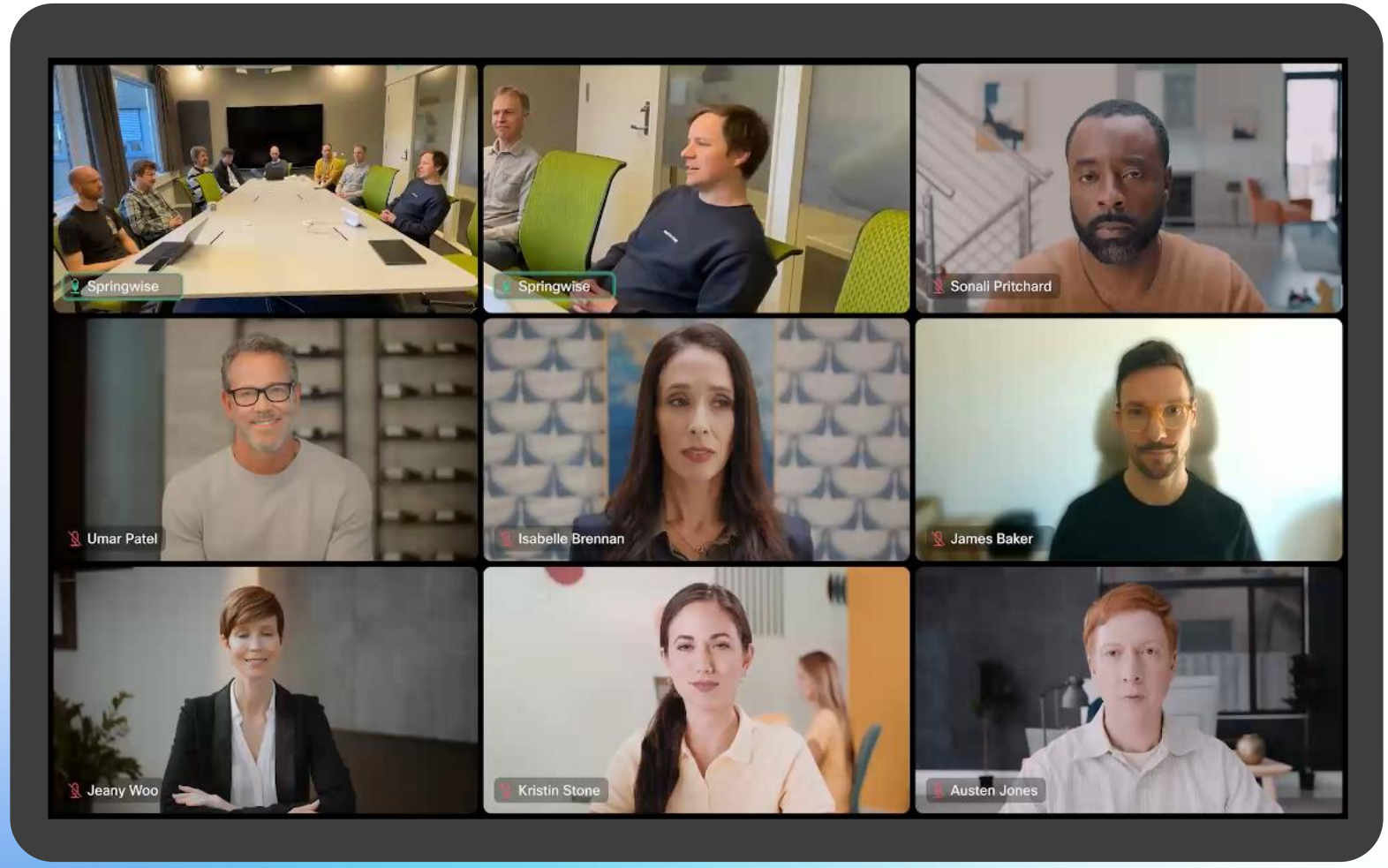
Wherever innovation happens, from brainstorm to follow through



Dynamic Camera Mode

AI Director

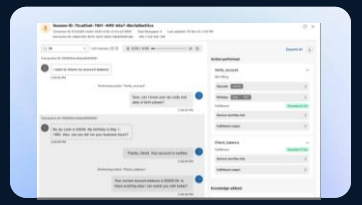
An industry-first dynamic camera mode that predicts and adapts to give you the best view of the meeting, always



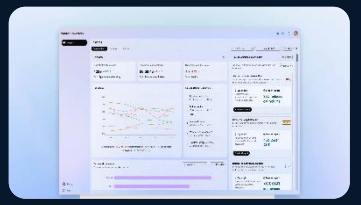
AI is core to the entire CX portfolio

Addressing all personas in CX with AI

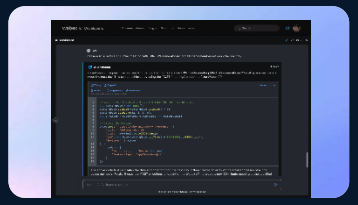
AI Knowledge Bases



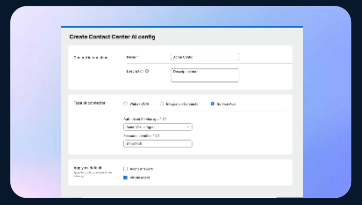
AI Agent
Generally Available



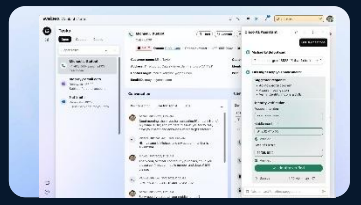
AI Real Time Intelligence



AI Assistant for Developers
Generally Available

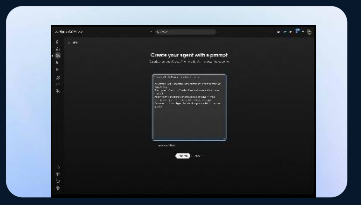


BYO VA
Generally Available



AI Assistant
Generally Available

AI QM



AI Agent Auto-creation

Near Real Time Topic Analytics

Build and Configure | **Deliver and Manage** | **Analyze and Improve**

Admins

Agents

Supervisors

Analysts, QA,
Operators

Collaboration AI Architecture Services

AI Design Patterns

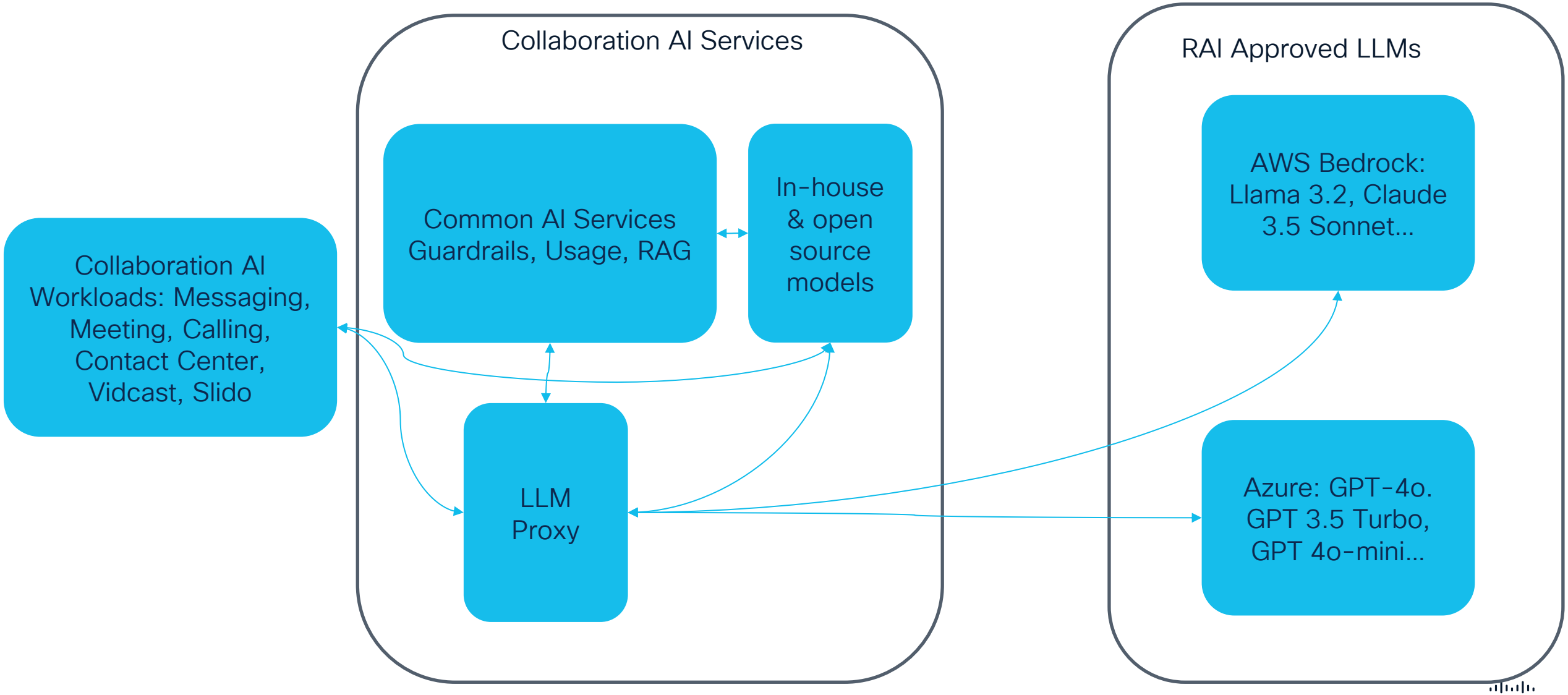
- Supervised Machine Learning
- Generative AI
- Small Language Models
- Large Language Models
- Realtime media models
- Multi Modal models
- Multi-model pipelines

AI Design Patterns – Multiple Model Use Case Examples

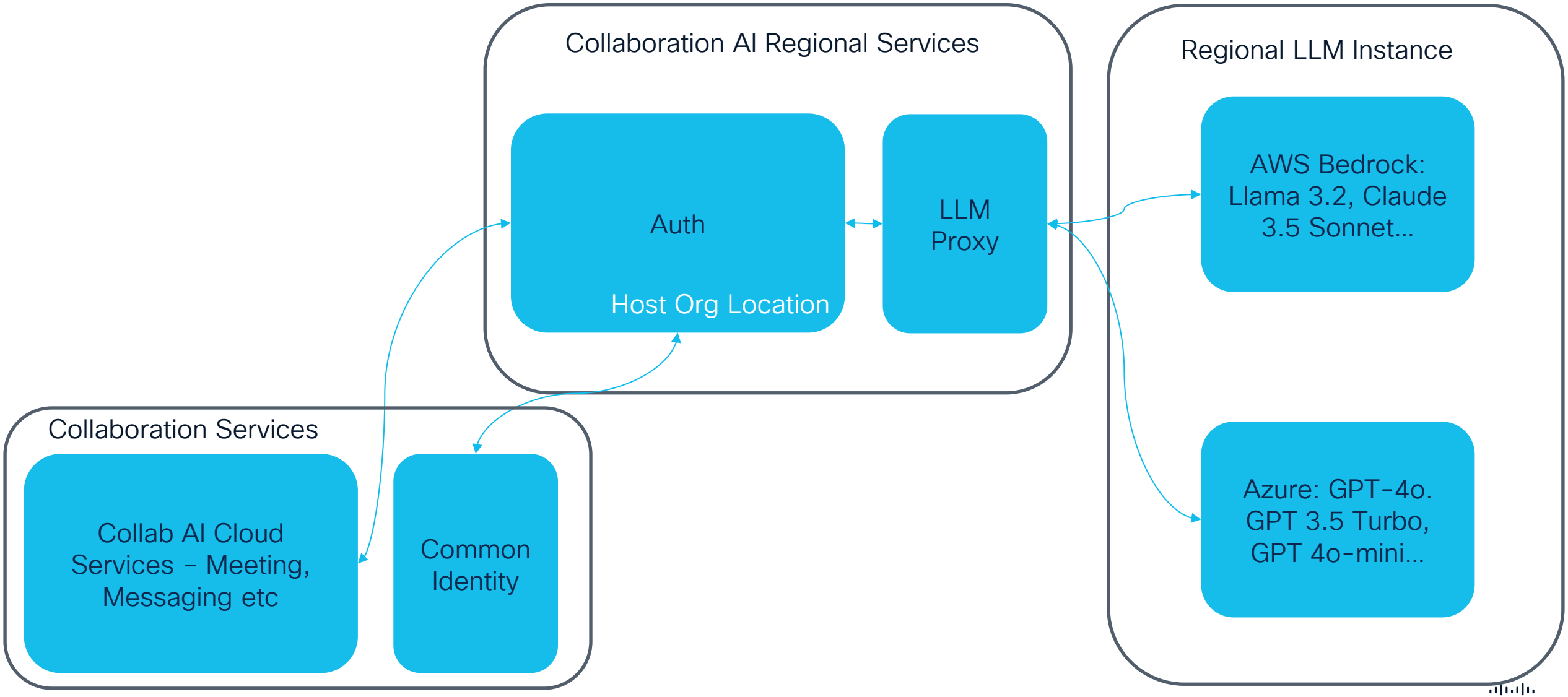
Multi-model: LLM + in-house models	Virtual Agents	AI Assistants	
Fine Tuned Large Language Models	Ask me anything	Question Answering	
Large Language Models	Content Generation	Summarization	Question Answering
Language Models	Transcription	Sentiment	Topic analysis
Media Models	Speech Enhancement	AI Codec	Computer Vision

Foundational AI Services

Collaboration AI Platform: Architecture Components



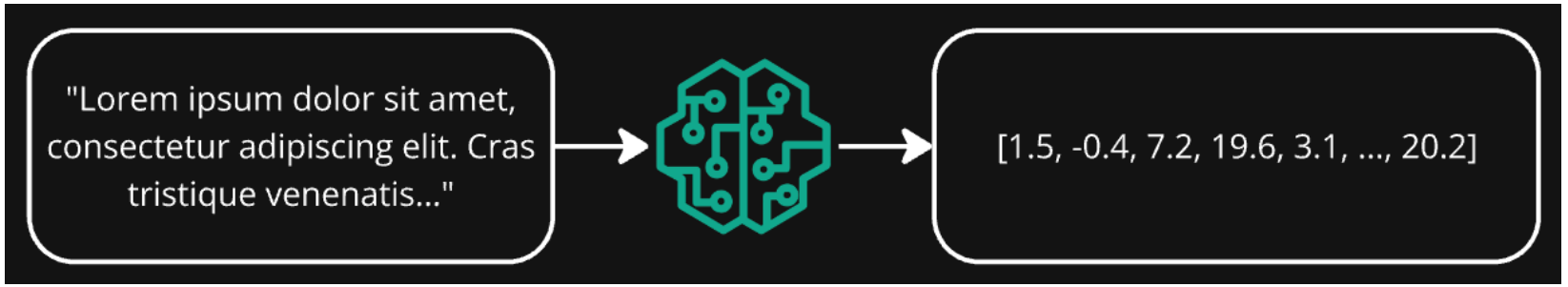
Collaboration AI Platform: Regional Deployment



cisco

Enterprise Context: RAG and Embeddings

Purpose: To generate embedding vectors for arbitrary inputs to enable RAG and context with LLMs



Guardrails

Purpose: To prevent LLMs from generating, mirroring, or consuming inappropriate, harmful, or non-compliant content.

Toxicity

- Obscenity
- Hate
- Threats
- Insults

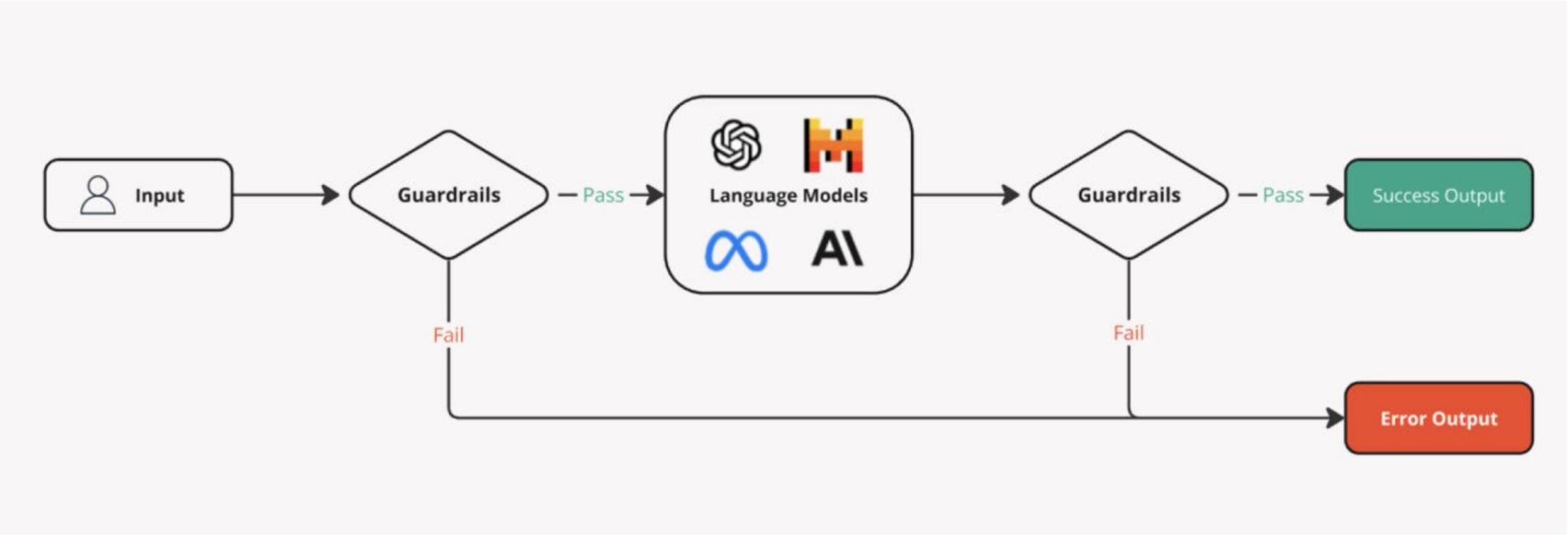
Harm

- Violence
- Crime
- Self-harm
- Exploitation
- Privacy

Security

- Jailbreaking

Guardrails & Prompt Engineering



- Prompt engineering: Creating effective prompts that instruct LLM's to perform specific tasks

LLM's sensitive to input	Reduces ambiguity	Enhances performance	Enhances model output safety

https://blog.webex.com/innovation-ai/guardrails-for-ai-models/

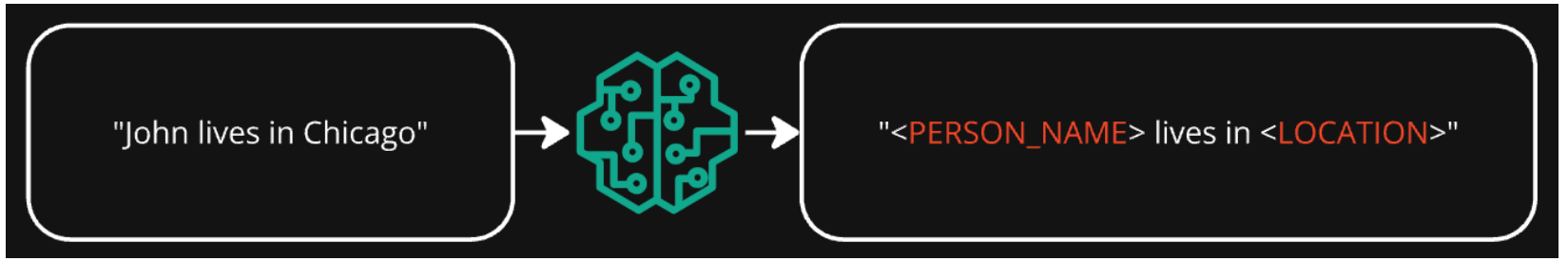
Guardrail Example

```python
# Import Guard and Validator
from guardrails.hub import ToxicLanguage
from guardrails import Guard

# Use the Guard with the validator
guard = Guard().use(
    ToxicLanguage, threshold=0.5, validation_method="sentence", on_fail="exception"
)

# Test passing response
guard.validate("Love how you think and attack the problem. Great job!")

try:
    # Test failing response
    guard.validate(
        "Please look carefully. You are a stupid idiot who can't do anything right."
    )
except Exception as e:
    print(e)
```

https://github.com/guardrails-ai/toxic_language

PII redaction

Purpose: To redact PII content from a provided input



Redacts types such as (but not limited to):

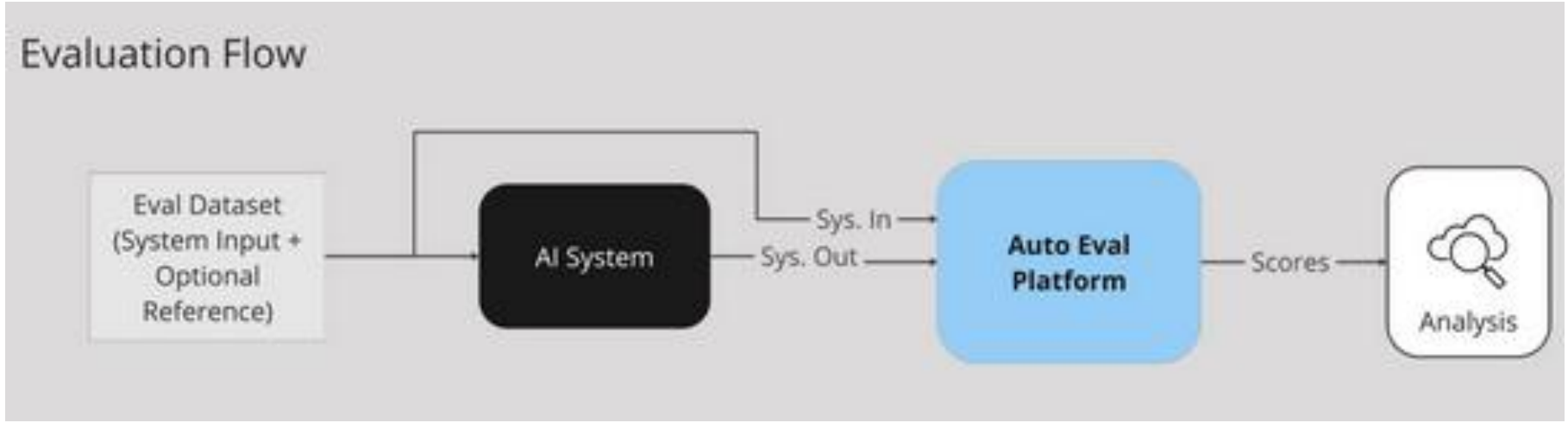
- person
- organization
- phone number
- address
- passport number
- email
- credit card number
- social security number
- health insurance id number
- date of birth
- mobile phone number
- bank account number
- medication
- cpf
- driver's license number
- tax identification number
- medical condition
- identity card number
- national id number
- ip address
- email address

LLM Evaluation – Auto Eval

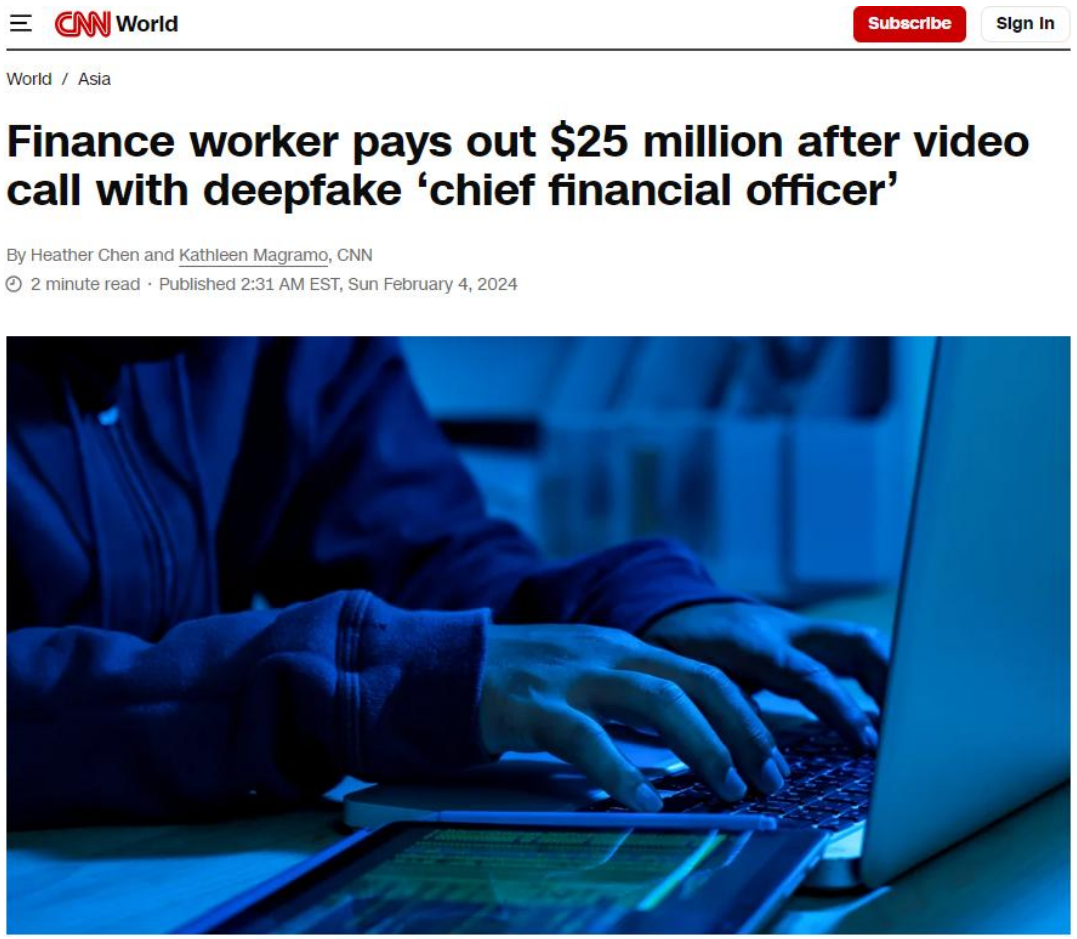
How It Works

1. **Curate Input Data**: Collect inputs (e.g., meeting transcripts or questions) for evaluation. Labeled reference data are optional.
2. **Generate Outputs**: Pass the inputs through the AI system under evaluation and save the outputs.
3. **Configure Metrics**: Select and customize evaluation metrics based on the use case.
4. **Run Evaluations**: Analyze system outputs against the selected metrics via the Auto Eval platform.
5. **Interpret Results**: Use the platform's built-in tools to visualize and interpret the scores.

Once a new AI system has been onboarded to this process, automatic evaluations can be run for a fraction of the time and cost of other methods. This scalable approach enables teams across the company to rapidly iterate, confidently deploy, and continuously monitor their AI solutions.



Deepfake



CNN World Subscribe Sign In

World / Asia

Finance worker pays out $25 million after video call with deepfake 'chief financial officer'

By Heather Chen and Kathleen Magramo, CNN

2 minute read · Published 2:31 AM EST, Sun February 4, 2024

Multi-Modal Deepfake Threats

Image

Manipulated or AI-generated images

Audio

Synthetically created or edited voices

Video

Synthetically created faces and manipulated audio

Text

AI-generated text and LLM-created content

Agentic Considerations for Foundational Services

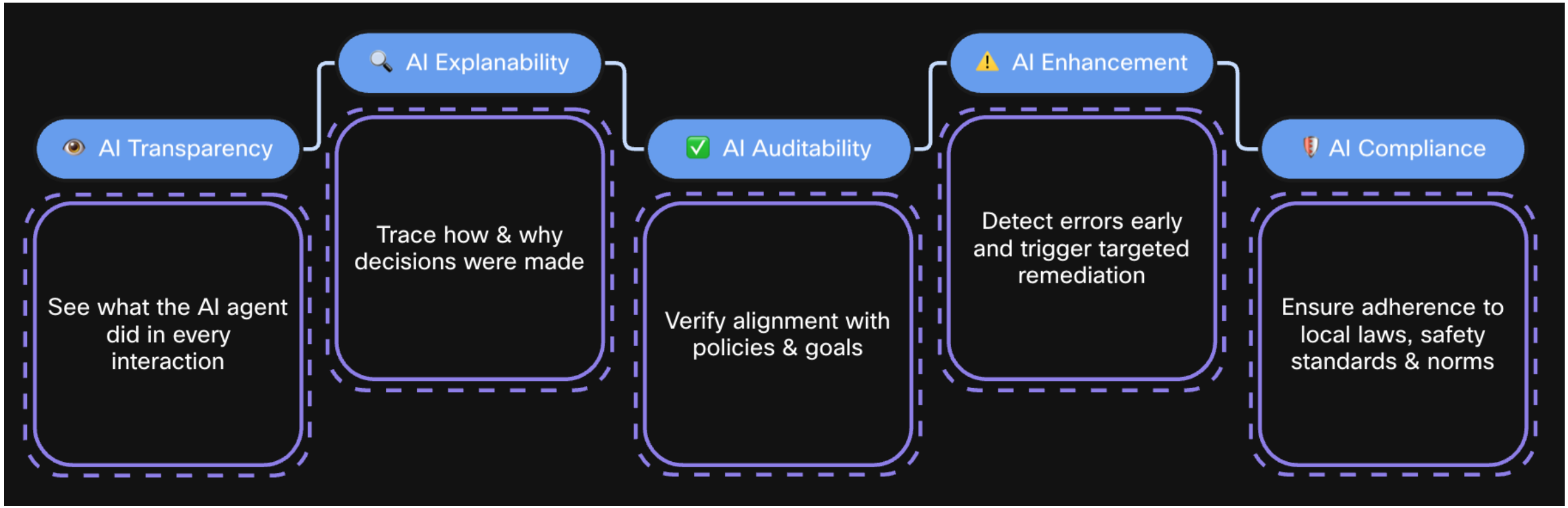
Agentic AI – Technical Challenges

- Deploying Agentic AI systems introduces a range of technical hurdles:
- **Complexity & System Integration:** Integrating multiple subsystems (NLP, reasoning engines, APIs, databases) into a cohesive framework is challenging.
- **Data Quality & Relevance:** Agentic AI relies on accurate, real-time, and high-quality data from diverse, often fragmented, sources. Data fragmentation can lead to incomplete reasoning.
- **Reliability & Predictability:** Ensuring consistent and expected behavior in dynamic and unpredictable environments is difficult.
- **Explainability & Trust:** The "black box" nature of LLMs and complex reasoning makes it difficult to understand why an agent acted a certain way.
- **Data Privacy & Security:** Agents often access sensitive data across multiple systems, posing risks of data leaks, unauthorized actions, and prompt injection attacks.
- **Debugging Autonomous Systems:** The emergent behavior from complex interactions between components makes traditional debugging tools insufficient.
- **Scalability:** Scaling agentic AI for real-world deployment requires handling increased complexity, larger datasets, and higher computational demands.

Agentic AI – Architecture Direction

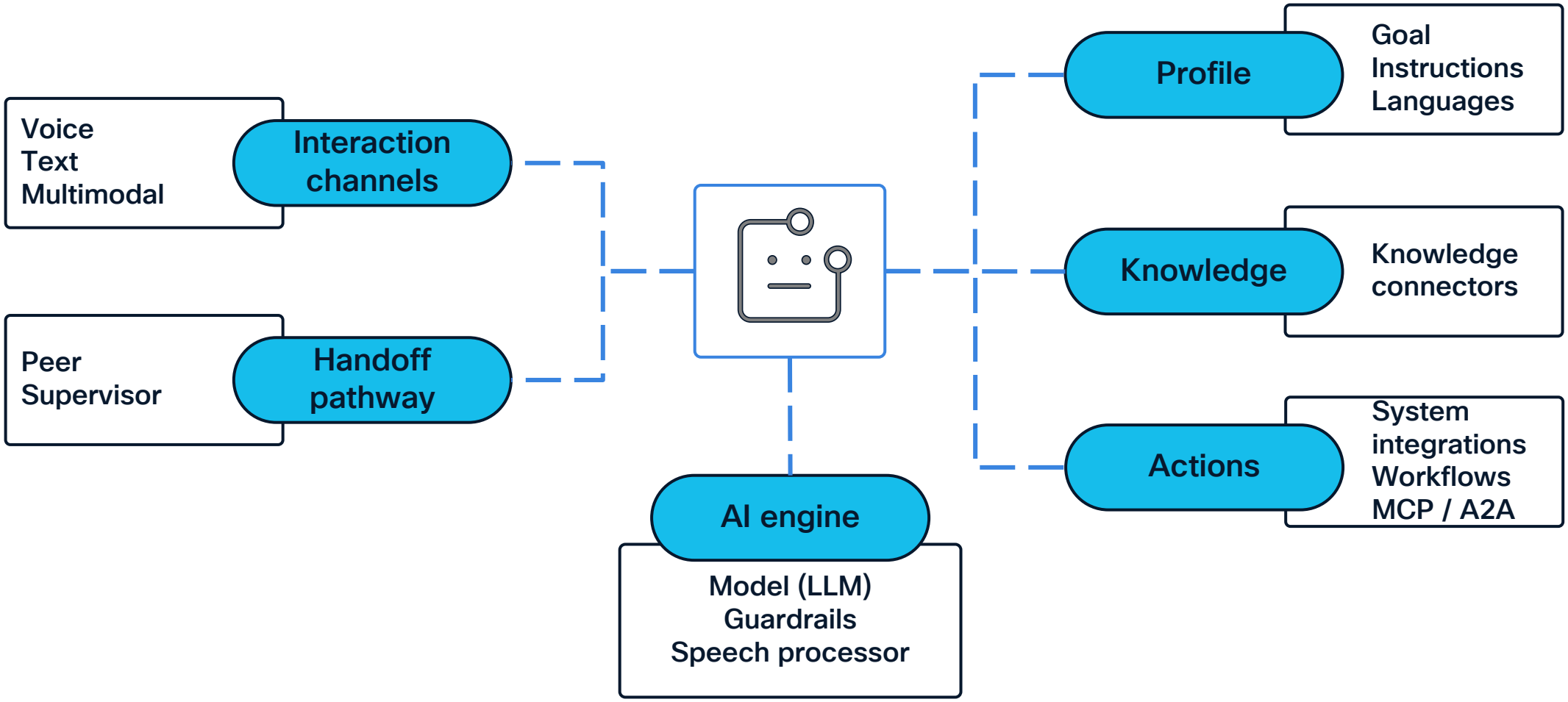
- **Future Directions & Research**
- The field of Agentic AI is rapidly evolving with several promising research directions:
- **Enhanced Reasoning:** Improving LLM capabilities to develop more robust outcomes for better decision-making and boundary enforcement with sensitive information.
- **Human-AI Collaboration:** Focus on designing systems where Agentic AI augments human expertise fostering collaborative environments.
- **Responsible AI & Governance:** Developing robust frameworks for safety, compliance, and accountability, including mechanisms for traceability, audit trails and considerations for emerging agentic protocols.
- **Advanced Learning Mechanisms:** Continuous research into more sophisticated learning and adaptation techniques to enable agents to improve performance and handle novel situations.
- **Specialized Agentic AI:** Development of domain-specific agentic AI solutions, for example, in scientific discovery e.g. healthcare, finance, and cybersecurity.
- **Interoperability and Standardization:** Addressing the challenges of integrating diverse data sources and systems to enable seamless operation across enterprise environments.
- **User Experience:** For augmented and automated user experiences.

Considerations for Agentic Flows



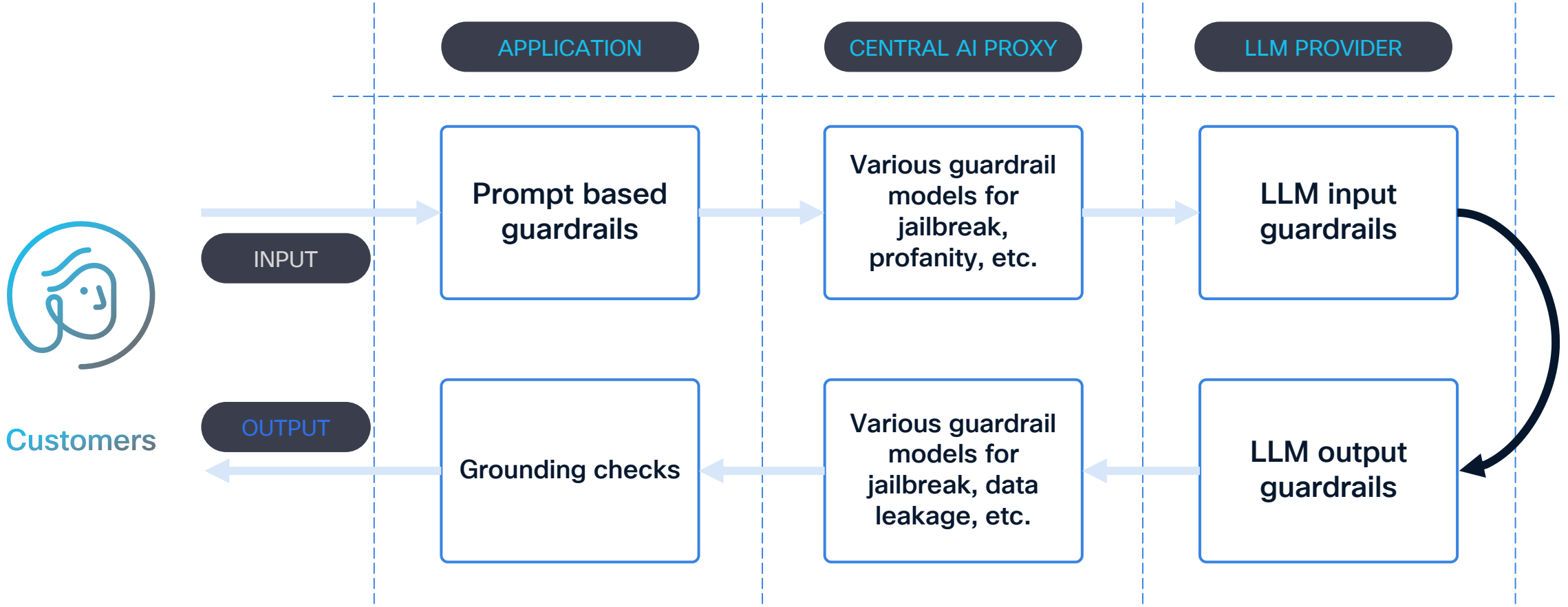
AI Architecture in action: Contact Center Virtual Agent

The anatomy of an AI agent for CX



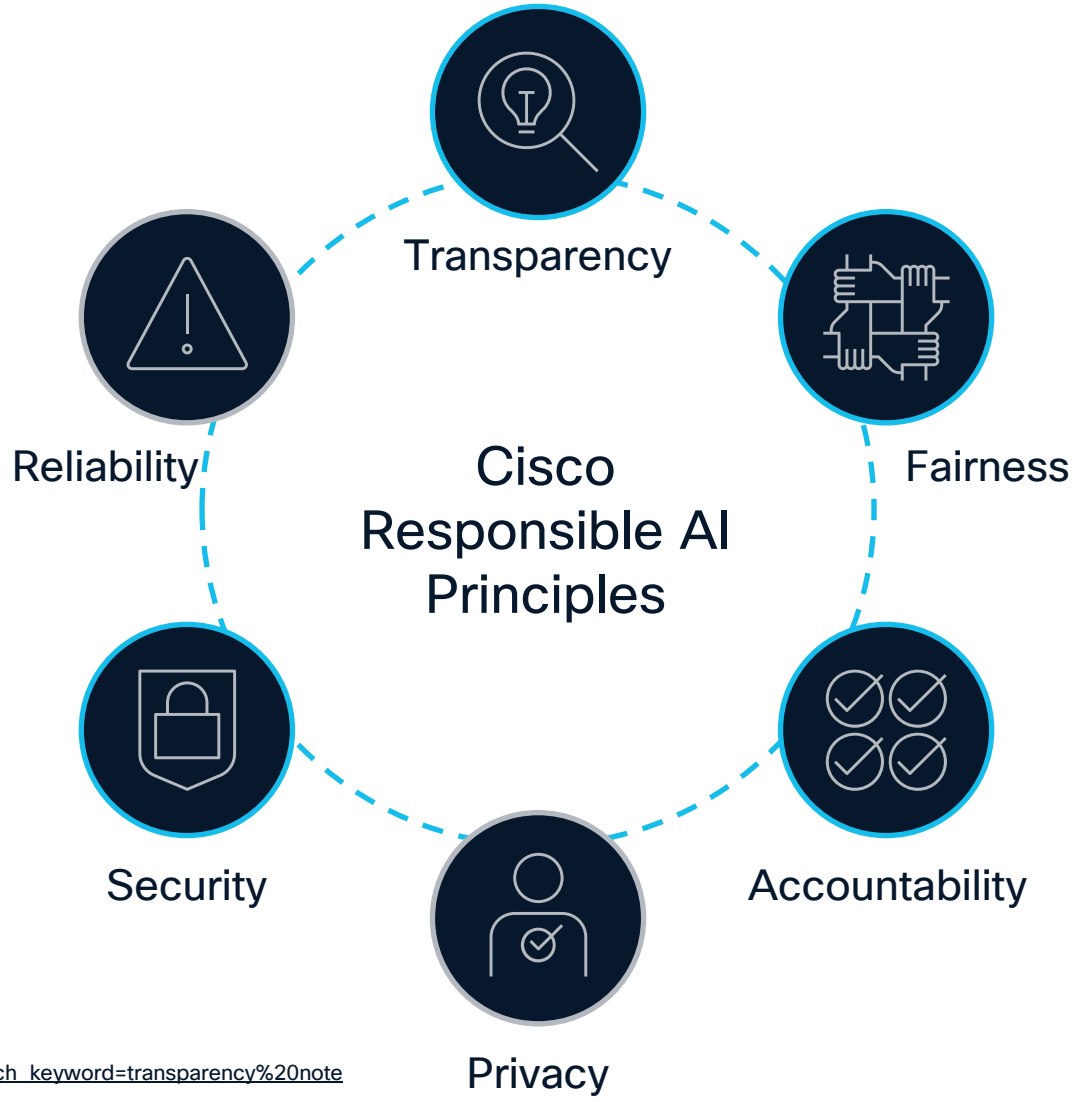
Contact Center Virtual Agent Demo

AI Agents: Security with multiple layers of guardrails



Responsible AI in action

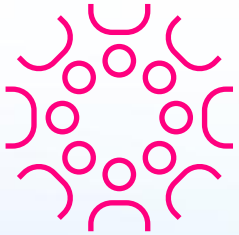
Cisco Responsible AI Principles



https://trustportal.cisco.com/c/r/ctp/trust-portal.html?search_keyword=transparency%20note

Privacy Center of Excellence

Governance and Compliance



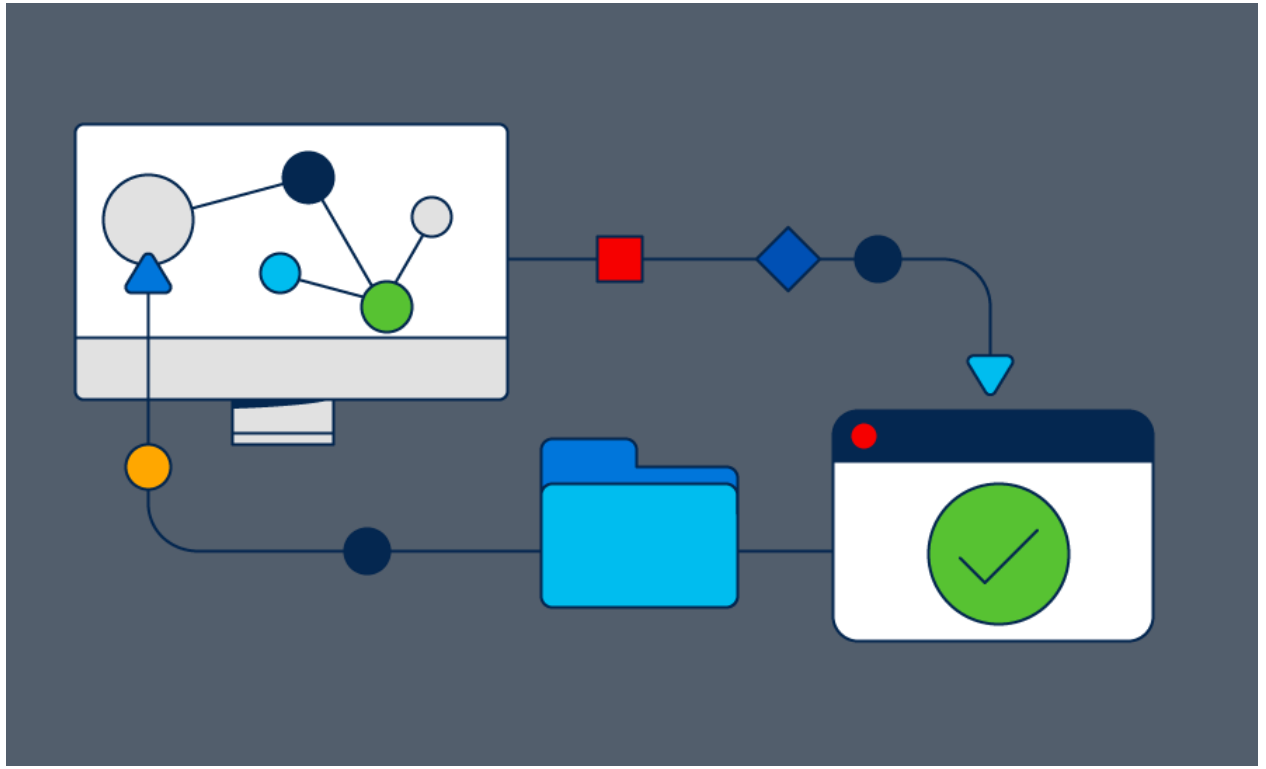
Market Access

Differentiation

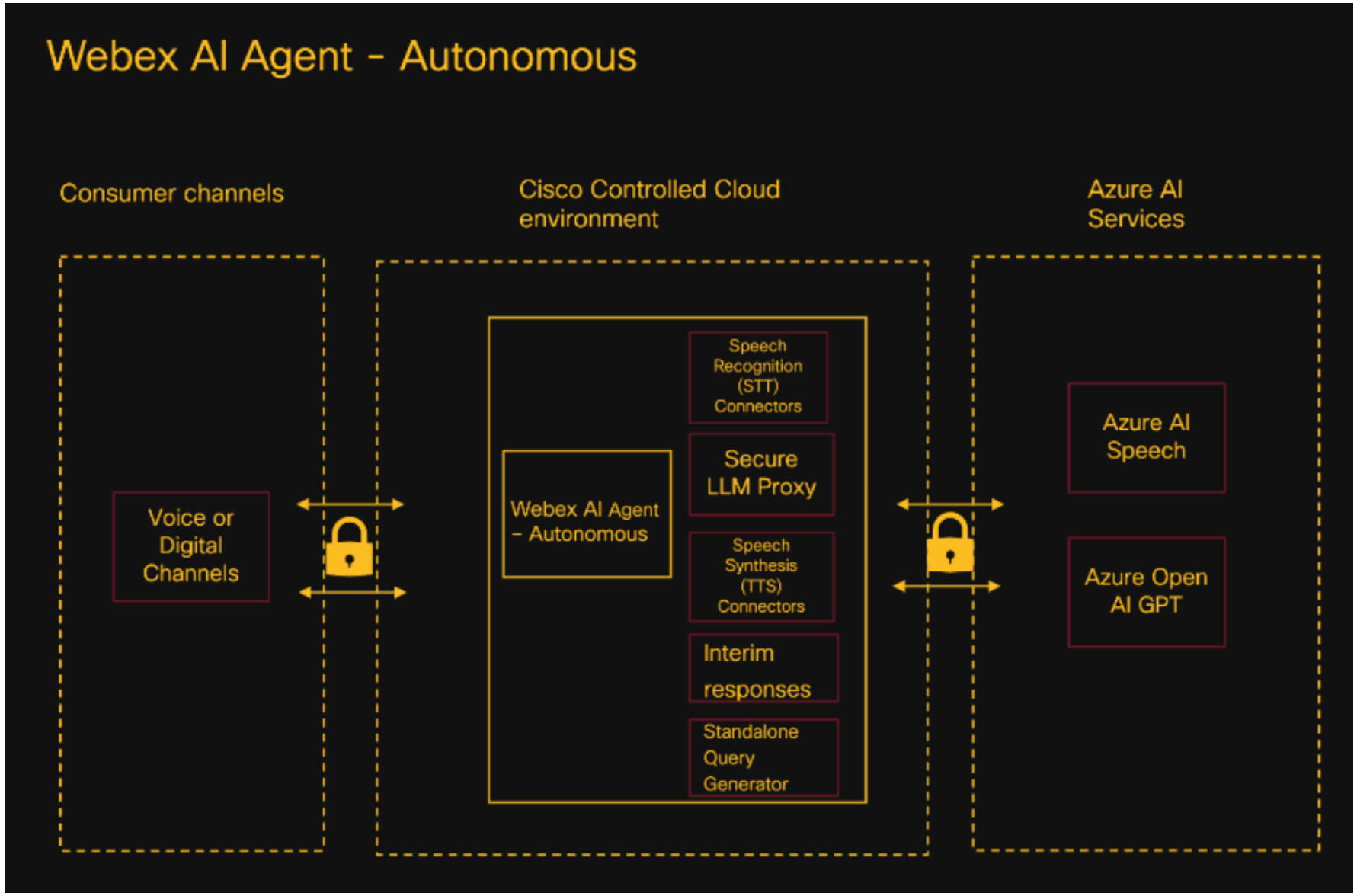
cisco

AI Impact Assessment Process

- Developed by cross-functional team
- Gather information to surface risks
- Evaluate various aspects of AI: the model, training data, fine tuning, prompts, privacy and testing methodologies
- Mitigate the risks by aligning with our RAI Principles



Webex AI Agent Transparency Note



- Model Overview
- Model Input / Output
- Training Data Sources
- Model Evaluation
- Safety and Ethics
- Fairness
- Privacy & Security

https://trustportal.cisco.com/c/r/ctp/trust-portal.html?search_keyword=transparency%20note#/19445370048945010

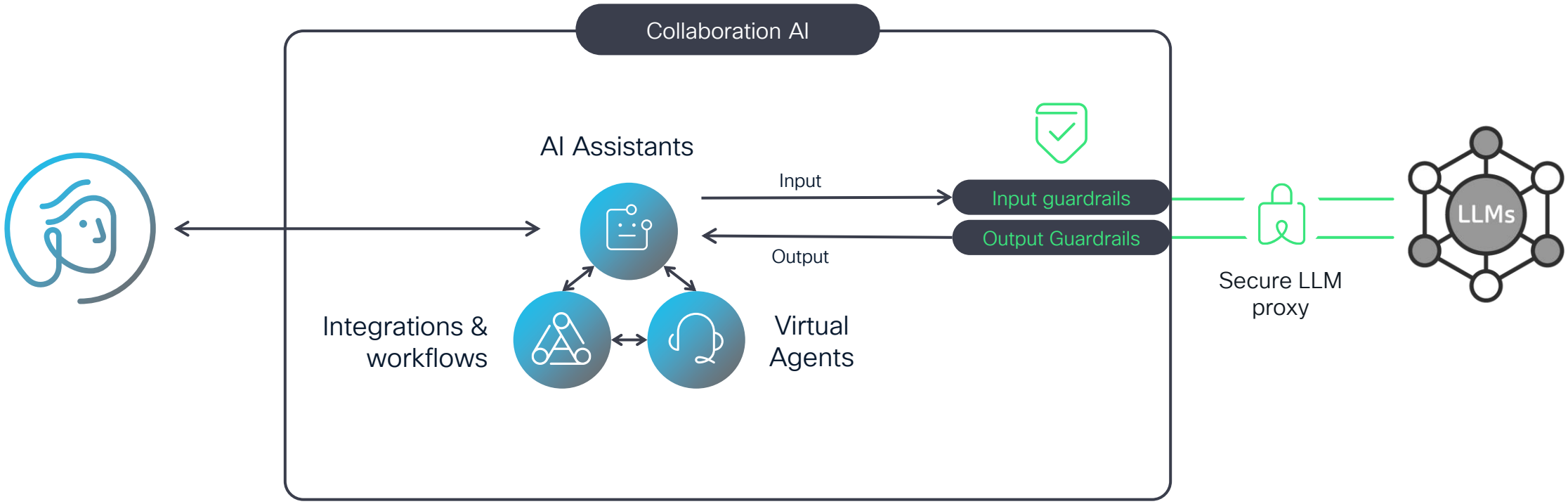
Trustportal demo

Future considerations: Agentic AI Risks

- While powerful, agentic AI systems introduce significant challenges that require robust governance and technical controls.
- **Security, Privacy, and Ethical Risks:**
  - **Security:** Threats include prompt injection, agent hijacking, and cascading vulnerabilities across multi-agent systems. The non-deterministic nature of LLMs also impacts reliability.
  - **Privacy:** Extensive data logging and inter-agent communication create risks of sensitive data exposure and potential violations of regulations like GDPR.
  - **Ethics / RAI:** Diminished human oversight raises accountability questions. Model biases can be amplified, and opaque decision-making undermines trust.
- **Mitigation Strategies and Governance:**
  - **Technical Controls:** Best practices include implementing zero-trust security principles, end-to-end encryption, strong authentication and hardened agent identity frameworks. Human-in-the-loop (HITL) workflows for oversight are important where possible.
  - **Governance:** Responsible AI Framework evolution for Agentic Systems.
- **Regulatory Landscape and Future Roadmap:**
  - **Regulation:** The EU AI Act (2024) and other global regulations will mandate rigorous safety audits, documentation, and transparency for high-risk AI systems, including agentic platforms.
  - **Future Outlook:** In the short term agentic AI is likely to move from pilot to production with a focus on standardizing agent directories and identity. Focus on security with emerging protocols is critical.

Summary

Applying a safety and security first approach



Collaboration AI Architecture Summary

- Continued focus on AI innovation and delivery
- Different architecture approaches – media, language and agentic models
- Model flexibility across use cases
- Agentic ready
- Responsible AI

Complete your session evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to claim a Cisco Live T-Shirt.

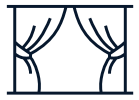
Earn up to 800 points by completing all surveys and climb the Cisco Live Challenge leaderboard.

Level up and earn exclusive prizes!

Complete your surveys in the Cisco Live Events app.

cisco

Continue your education



Visit the Cisco Stand for related demos

Book your one-on-one Meet the Expert meeting

Attend the interactive education with Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Thank you


CISCO