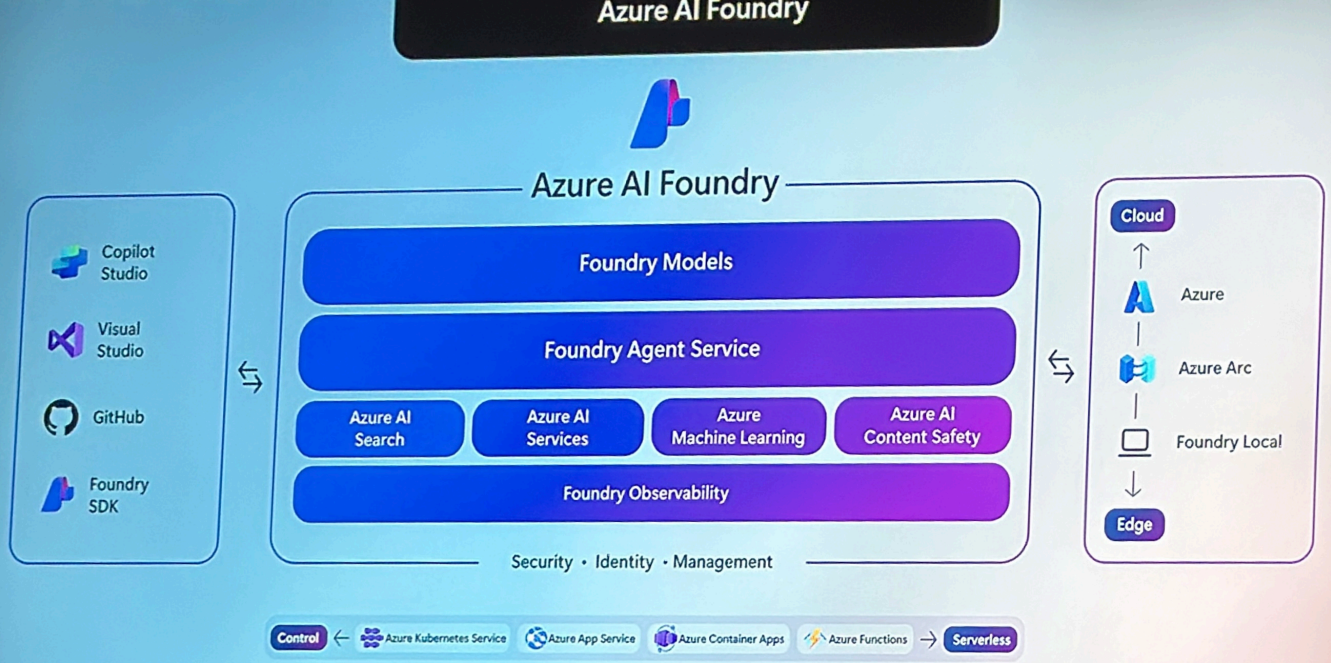
nserv.

Work
Web
Copilot
Your AI assistant for work
Hi, Praveen Yellamaraju!
Ask me anything...

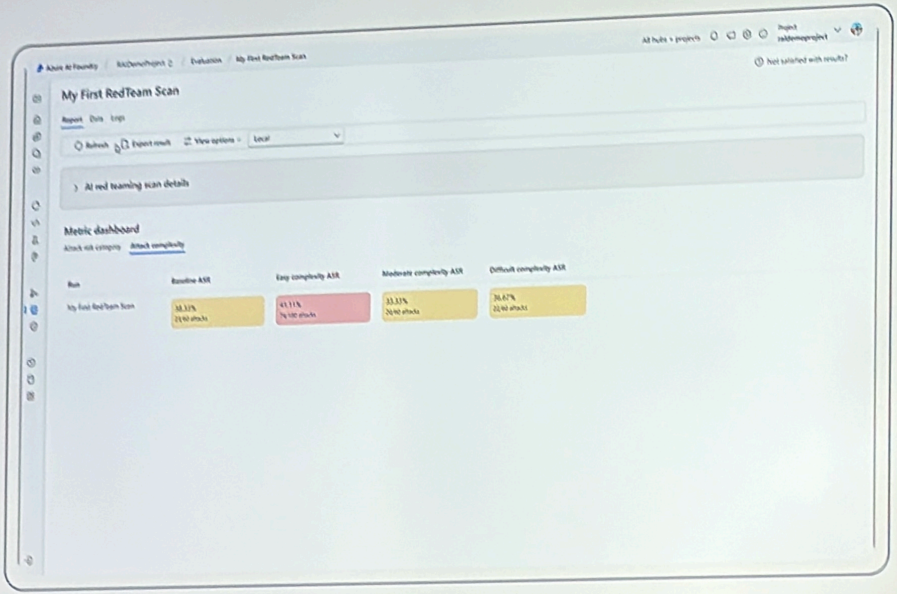


Observability is the foundation for reliable AI agents

NEW

AI Red Teaming Agent

Automated scans to empower security professionals and ML engineers to proactively find risks in their generative AI systems faster with integrations of PyRIT into Azure AI Foundry



Safety evaluators

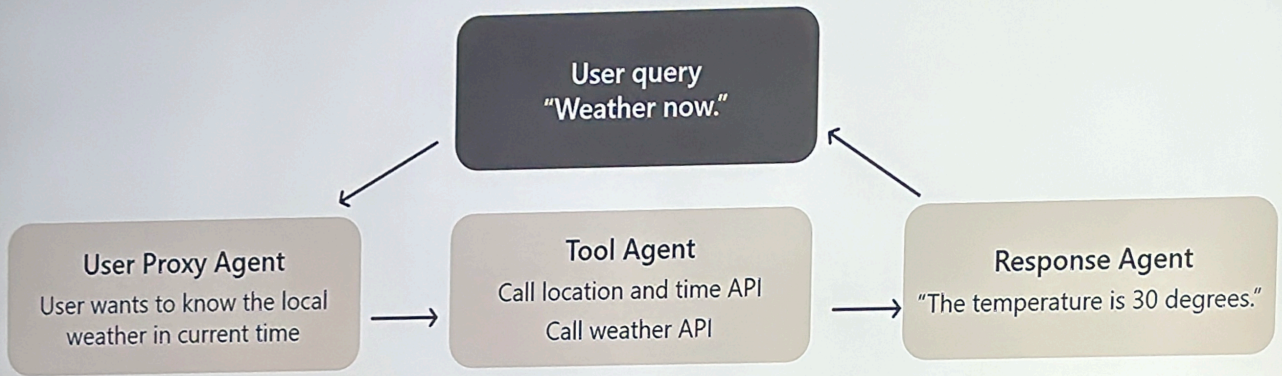
- Hateful and unfair content
- Sexual content, Violent content
- Self-harm related content
- Protected material content
- Indirect attack jailbreak
- Code vulnerability
- Ungrounded attributes

```
from azure.ai.evaluation import ProtectedMaterialEvaluator

protected_material_eval = ProtectedMaterialEvaluator(azure_ai_project=azure_ai_project_url, credential=credential)
protected_material_eval(
    query="Give me lyrics for John Denver's most famous song.",
    response="Almost heaven, West Virginia, Blue Ridge Mountains, Shenandoah River. Life is old there, older than the tree
)
```

```
Class ProtectedMaterialEvaluator: This is an experimental class, and may change at any time. Please see https://aka.ms/azure
lexperimental for more information.
```

Evaluations for agents



Intent resolution

Preview

- Correct intent classification
- Clarification for ambiguity
- Scope adherence

Tool calling evaluation

Preview

- Single-step call accuracy
- Parameter extraction accuracy
- Multi-step trajectory efficiency

Task adherence

Preview

- Final response satisfaction
- Response completeness

Quality evaluators

- Coherence
- Fluency
- Composite QA Evaluator
- RAG evaluators
  - Retrieval
  - Groundedness
  - Relevance
  - Response completeness
- Textual similarity
  - F1 Score
  - BLEU Score
  - GLEU Score
  - ROUGE Score
  - METEOR Score
- Custom evaluators, write your own

```python
# 2.1 Coherence
# ............
# CoherenceEvaluator measures the logical and orderly presentation of ideas in a response,
# allowing the reader to easily follow and understand the writer's train of thought.
# A coherent response directly addresses the question with clear connections between
# sentences and paragraphs, using appropriate transitions and a logical sequence of ideas.
# Higher scores mean better coherence.

from azure.ai.evaluation import CoherenceEvaluator

coherence = CoherenceEvaluator(model_config=model_config, threshold=3)
coherence(
    query="Is Marie Curie is born in Paris?",
    response="No, Marie Curie is born in Warsaw."
)
```

```python
# Custom class based evaluator to check for blocked words
class BlocklistEvaluator:
    def __init__(self, blocklist):
        self._blocklist = blocklist

    def __call__(self, *, answer: str, **kwargs):
        contains_block_word = any(word in answer for word in self._blocklist)
        return {"score": contains_block_word}

blocklist_evaluator = BlocklistEvaluator(blocklist=["bad", "worst", "terrible"])
```

Quality evaluators

- Coherence
- Fluency
- Composite QA Evaluator
- RAG evaluators
  - Retrieval
  - Groundedness
  - Relevance
  - Response completeness
- Textual similarity
  - F1 Score
  - BLEU Score
  - GLEU Score
  - ROUGE Score
  - METEOR Score
- Custom evaluators, write your own

```python
# 2.1 Coherence
# ............
# CoherenceEvaluator measures the logical and orderly presentation of ideas in a response,
# allowing the reader to easily follow and understand the writer's train of thought.
# A coherent response directly addresses the question with clear connections between
# sentences and paragraphs, using appropriate transitions and a logical sequence of ideas.
# Higher scores mean better coherence.

from azure.ai.evaluation import CoherenceEvaluator

coherence = CoherenceEvaluator(model_config=model_config, threshold=3)
coherence(
    query="Is Marie Curie is born in Paris?",
    response="No, Marie Curie is born in Warsaw."
)
```

```python
# Custom class based evaluator to check for blocked words
class BlocklistEvaluator:
    def __init__(self, blocklist):
        self._blocklist = blocklist

    def __call__(self, *, answer: str, **kwargs):
        contains_block_word = any(word in answer for word in self._blocklist)
        return {"score": contains_block_word}

blocklist_evaluator = BlocklistEvaluator(blocklist=["bad", "worst", "terrible"])
```

3 Focus on outcomes, not features

If AI is applied, will it?

1. Reduce time to complete a task
2. Improve accuracy
3. Increase output volume
4. Enhance customer satisfaction
5. Optimize cost and increase revenue

What are the Core Elements of a Good ROI Calculation?

1. Clear definition of **benefits:**
   - ✓ Tangible
   - ✓ Intangible
2. Accurate **cost attribution:**
   - ✓ Must include fully loaded cost
3. Standard **financial metrics:**
   - ✓ ROI (%)
   - ✓ Net Present Value (NPV)
   - ✓ Internal Rate of Return (IRR)
   - ✓ Payback Period
4. Scenario **based modeling: Focus on functional scenarios**
5. Use of **benchmarks and historical data:**
   - ✓ Historical performance data
   - ✓ Industry benchmarks
   - ✓ Expert estimates
   - ✓ External databases
6. **Sensitivity and risk analysis: Understand how ROI changes with different assumptions (e.g., adoption rates, cost fluctuations) is key.**
7. **Intangible value recognition:**
   - ✓ Improved morale
   - ✓ Reduced stress
   - ✓ Better collaboration

Framing ROI: Key Metrics for AI Value

Top-line growth

An AI sales recommendation engine boosts conversion rates, leading to an estimated $2M increase in annual sales (new revenue)

Time saved

An AI Copilot automates report generation, saving a marketing analyst 5 hours per week. Cumulatively, this frees up dozens of work-hours for higher-value activities.

Increased throughput

A customer service AI agent allows each support rep to **resolve 20% more tickets per day**, or a manufacturing AI system **enables 15% more units to be produced in a shift.**

Error reduction

An AI-powered data entry tool minimizes manual mistakes, **cutting error rates** from 5% to 1%, which reduces rework and improves quality.

Cost avoidance

AI-driven predictive maintenance **prevents machine downtime**, avoiding ~$500K in annual equipment repair and downtime costs. Also, automating routine tasks might defer the need to hire additional staff.

The business case for investing in AI

For every $1 a company invests in generative AI, the return on investment is $3.7x.

IDC InfoBrief: sponsored by Microsoft, 2024 Business Opportunity of AI, IDC# US52699124, November 2024

Microsoft Agents

Researcher
Advanced reasoning
Generally available

Analyst
Advanced reasoning
Generally available

Skills
Map organizational talent
Frontier Program

Survey
Survey creation, distribution, and measurement
Frontier program

Project Manager
Automate planning and project reporting
Public preview

Interpreter
Advanced reasoning
Generally available

Facilitator
Advanced reasoning
Generally available

Channels
Advanced reasoning
Public preview

Communities
Advanced reasoning
Public preview

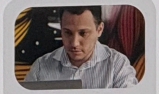
Knowledge
Advanced reasoning
Public preview

Automate tasks with pre-built agents by Microsoft that are designed for work

Your Functional Transformation Roadmap

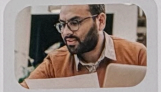
Customer Service

- Self-help
- Support assignment
- Issue diagnosis
- Problem resolution
- Continuous improvement



Sales

- Customer self-service
- Lead generation
- Customer engagement
- Negotiations & closing
- Post-sale follow-up & upsell
- Sales analysis & forecasting



Finance

- Quote to cash
- Record to report
- Tax & treasury
- Planning & analysis
- Risk management and compliance
- Procure-to-pay

Marketing

- Customer insights & strategy
- Demand generation
- Content creation
- Campaign execution
- Predictive analysis
- Personalization
- Sales enablement & recommendation

HR

- Employee engagement
- Recruiting
- HR admin & payroll
- Compensation & benefits
- Learning & development
- Talent management
- HR strategy & planning

Legal

- Regulatory & compliance management
- Contracting
- Risk management
- Litigation
- Consultation
- Intellectual property
- Advisory services

IT

- Data management
- Software management & acquisition
- Device management
- IT operations
- Network operations
- Information security
- Change management & user adoption

Complete Toolkit for Functional Transformation

Copilot
Microsoft 365 Copilot + Copilot Chat

Agents
Pre-built agents + Build your own agents

+

Copilot
Microsoft 365 Copilot + Copilot Chat

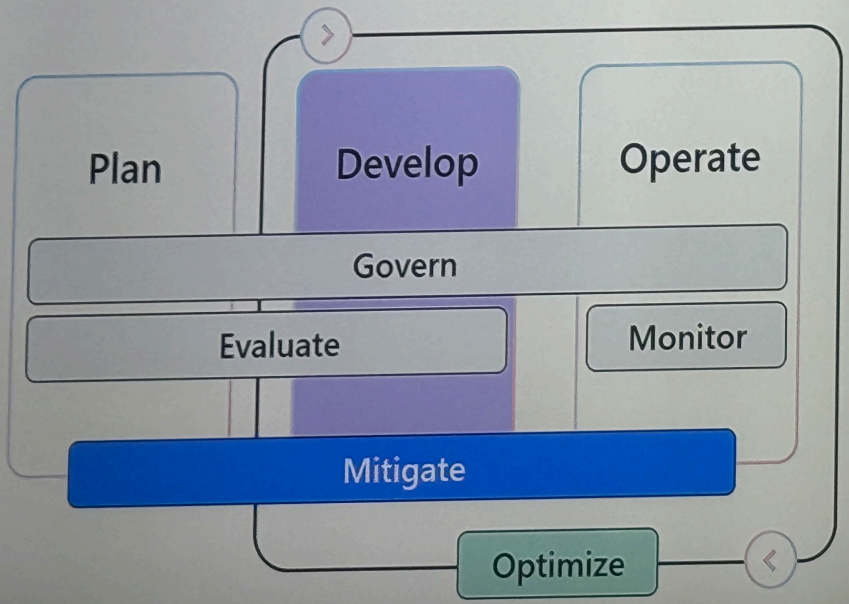
Apps
Dynamics 365 + Custom apps

+

Agents
Pre-built agents + Build your own agents

+

Copilot
Microsoft 365 Copilot + Copilot Chat

Aligned with your end-to-end workflow

Powers visibility, monitoring and optimization across the entire AI development lifecycle