Trustible

Mass General Brigham

Watching the Watchers: Monitoring AI for Risk, Privacy, and Trust

CHAI 2026 Legal Summit
September 2026

About Trustible

Who We Are

Trustible is the leading AI governance platform enabling safe and compliant AI adoption. Our AI governance platform enables enterprises to identify, measure, and mitigate AI risk to accelerate AI adoption. The company is headquartered in the Washington D.C. area. For more information, visit trustible.ai.

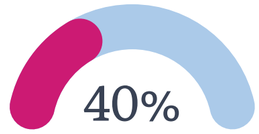
What We do

Responsible AI Governance Software & Solutions

Whether you are building internal AI models, leveraging generative AI or using AI-enabled SaaS applications, Trustible™ enables your organization to manage and mitigate AI risk, build trust, and accelerate responsible AI development. Plus, align your efforts to regulatory frameworks like the EU AI Act, NIST AI RMF, or ISO 42001.

Customer Snapshot

Percent of Trustible customers meeting these criteria



Fortune 500

Publicly Traded

Operate Globally

DEPLOY →

PRE-DEPLOYMENT

Governance by Design

- Intake & Discovery
- Inventory
- Risk Management & Triaging
- Documentation & Approvals

Trustible

FULL LIFECYCLE
AI GOVERNANCE PLATFORM

POST-DEPLOYMENT

Continuous Monitoring

- System Monitoring
- AI Incidents
- Vendor Changes & User Feedback
- Audit & Regulatory Compliance

< MONITOR & ITERATE

The Result

AI governance becomes the part of your AI strategy that helps you realize AI's potential.

4X
more AI use cases approved

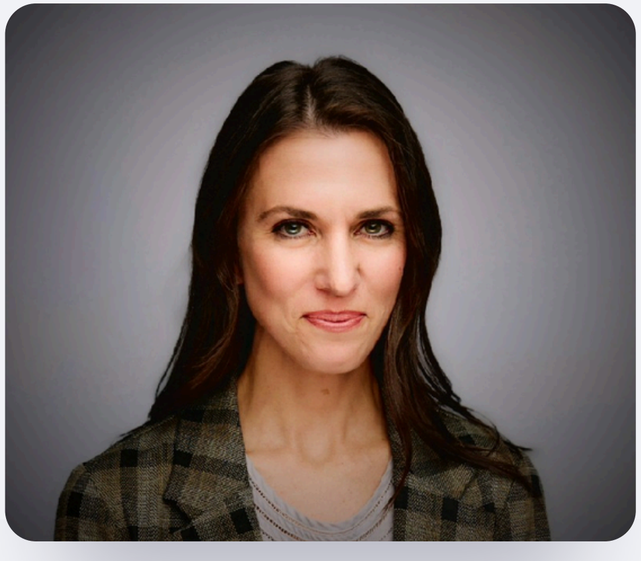
10X
faster AI intake

100%
Enterprise AI visbility

Today’s Speakers

Gerald Kierce
Co-Founder & CEO

Trustible



Amanda Centi, PhD
Director Digital AI Governance & Process

Mass General Brigham

Melissa Bateman Fitzgerald, Esq., CIPP
Chief Privacy Officer

Mass General Brigham

Why AI Monitoring is So Critical

AI Governance doesn’t stop at deployment

Regulatory requirements are specific, and already enforceable

EU AI Act (Article 72), ISO 42001, the NIST AI RMF, and ONC’s HTI-1 rule all say a version of the same thing: governance is not a one-time event.

Vendors change what you deployed, without asking

Vendors silently update models, turn on new capabilities, and update pricing with limited disclosure.

Risk compounds with every system you add

Hundreds of AI use cases running across internal models and vendor tools is hard to track and the risks evolve after they’ve been approved.

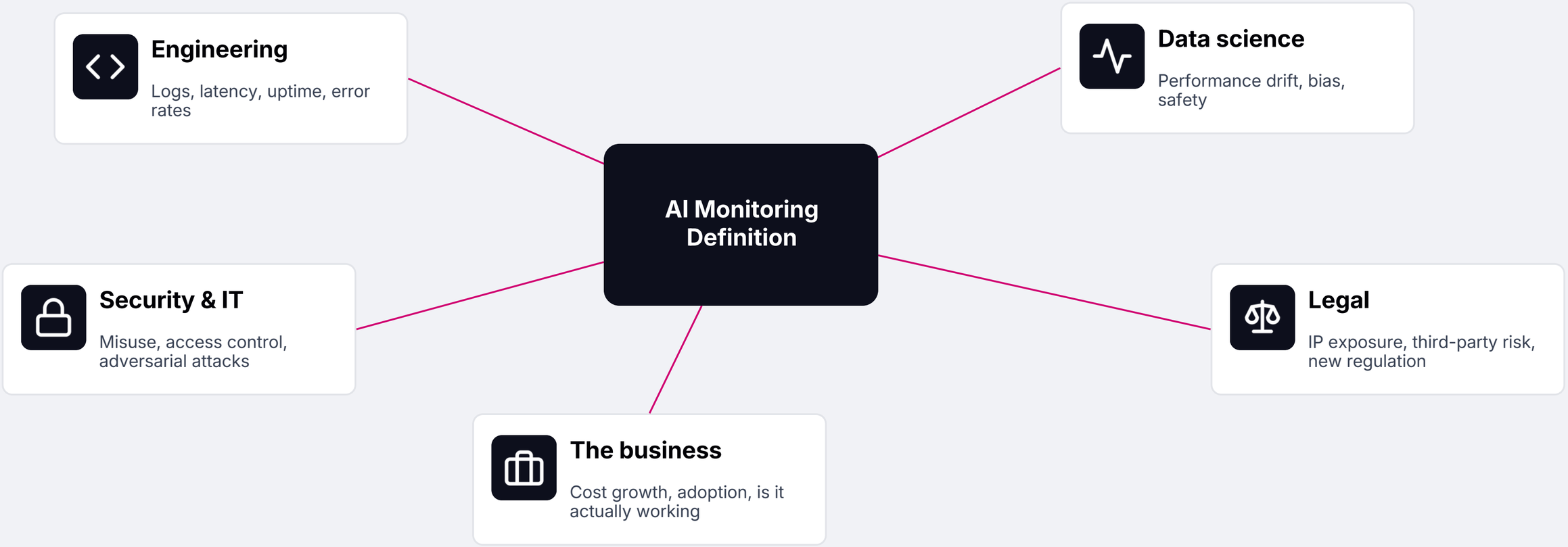
Cost climbs before value is proven

Token spend, licensing, and infrastructure cost climb with adoption, faster than anyone can show it's paying off.

WHAT WE'LL ANSWER TODAY

What exactly do we monitor?
And how do we do it?

Ask five teams what “AI monitoring” means, and you’ll get five answers



General categories of AI monitoring should be holistic

There are no universal standards for how to monitor, and it is often unclear what to even monitor in the first place. In March 2026, NIST published AI 800-4, the first federal mapping of post-deployment monitoring. **It organized the field into six categories:**

Functionality

Is the system still working as intended?

Operational

Is it maintaining consistent service?

Human Factors

Is it transparent, and are the interactions any good?

Security

Is it holding up against attacks and misuse?

Compliance

Does it still meet the regulations that apply?

Large-scale impacts

What is it doing out in the world?

Tracking those monitoring categories is a real challenge

Metrics are use-case-specific

There is no universal metrics you need to track for all types of use cases. What matters for a claims assistant is not what matters for a code reviewer or an ambient listener.

The most important signals need a human

You often can’t know an AI system is behaving without a person confirming its outputs are right or wrong. Much of that judgment can’t be pulled from an API. Someone has to say so.

Watching one model misses the picture

A single use case may run on several models, so model-level telemetry alone never tells the whole story.

Deployment type changes everything

Monitoring a model you host yourself is a different problem from monitoring a vendor-hosted tool you can only observe from outside.

Govern the use case, not (just) the model

A model is a component. A use case is the thing you deploy, own, and answer for. It's the level where monitoring actually makes sense.

MONITOR THE MODEL

The component view

- One endpoint at a time, blind to the rest
- Breaks the moment you swap or add a model
- No owner, no risk profile, no obligation
- Tells you nothing about vendor-hosted tools

MONITOR THE USE CASE

One governed system

- Runs on many models or agents, tracked as one thing
- Survives vendor swaps and added agents
- Carries one owner, risk profile, and story
- Same surface whether self- or vendor-hosted

Anchor monitoring to the use case and the hard problems dissolve: swapped models, mixed deployments, and multi-model systems all roll up to one place you can govern.

In order to govern the use case, teams needs to have an infrastructure for both **internal** and **external monitoring**

Internal monitoring

What goes into and comes out of the system.

Metrics you sample over time: cost, quality, drift, safety, usage. The numbers that show the system is still doing its job.

External monitoring

Signals about the system from outside its boundary.

Vendor changes, public incidents, and shifting regulation. None of it shows up in your own logs, and it’s often the highest-impact trigger.

Most organizations do some of the first and almost none of the second, but both matter to understand risks.

Let’s start with internal monitoring

Internal monitoring

What goes into and comes out of the system.

Metrics you sample over time: cost, quality, drift, safety, usage. The numbers that show the system is still doing its job.

External monitoring

Signals about the system from outside its boundary.

Vendor changes, public incidents, and shifting regulation. None of it shows up in your own logs, and it’s often the highest-impact trigger.

AI Monitoring requires visibility across all enterprise AI

Trustible
AI Monitoring Hub

- Recommended metrics per use case
- Dynamic risk scores and remediations
- Real time notifications based on triggers
- Governance controls and compliance

USE CASE TELEMETRY / METRICS TO COLLECT

Cost & Usage | Performance | Model Drift | Agent Behavior | Security | Ethics | User Feedback | Compliance

WHERE YOUR AI RUNS

- AI Cloud Providers: aws, Azure, Google Cloud
- EHR Platforms: Epic, Oracle Cerner
- AI Chatbot Applications: Claude, ChatGPT, Copilot
- AI Powered Software: salesforce, workday., servicenow

A snapshot of key internal metrics to collect

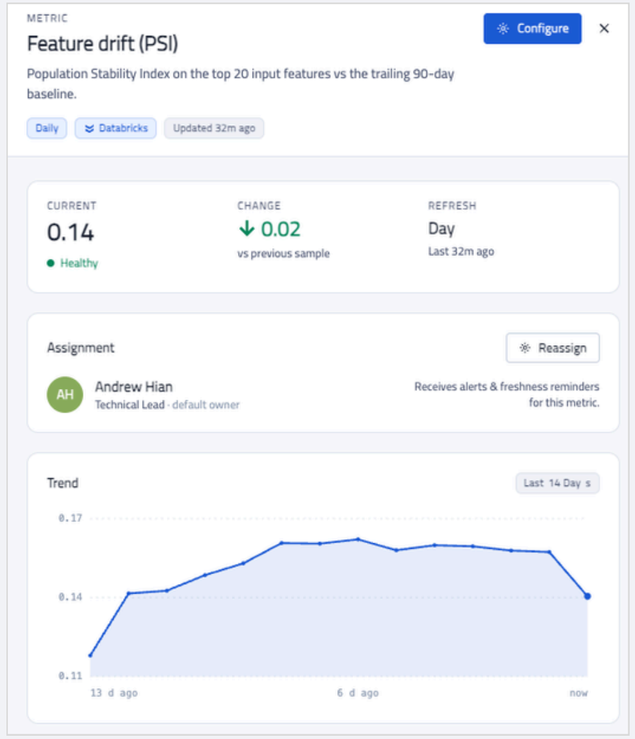
Numeric or pass/fail, sampled on a cadence, trended over time. The hard part was never storing a number. It’s knowing which numbers matter. We organize the internal surface into nine categories.

- **Performance & Reliability**
  Availability, latency, error rates
- **Cost & Usage**
  Token consumption, total spend
- **AI Quality**
  Precision, hallucination, error rate
- **Agent Behavior**
  Tool call error, runaway rate
- **Model & Data Drift**
  Concept, input/output drift
- **Responsible AI**
  Bias, toxit output, PII leak
- **Security**
  Jailbreak rates, unauthorized access
- **Compliance**
  Human review, residency, disclosure
- **User Feedback**
  Complaints, escalations, satisfaction

A simplified process for internal AI monitoring

1 Gather Metrics Samples

Values recorded over time from your AI platforms



2 Get Alerted

Configure a centralized alerting process

Determine absolute threshold

A value crosses a fixed line you set — a cost ceiling, an accuracy floor, a guardrail tripping.

Percent change from your base

A value moves sharply from before, even when no fixed limit was crossed. The swing itself is the signal.

2 Take Action

Determine whether that metric requires action

Update the alert

Tune the threshold if it was wrong.

Acknowledge & document

Record that a human saw it, and why it's fine.

Ignore

Dismiss it, with a justification on the record.

Begin a governance workflow

Escalate into a real, tracked process.

A quick view at external monitoring

Internal monitoring

What goes into and comes out of the system.

Metrics you sample over time: cost, quality, drift, safety, usage. The numbers that show the system is still doing its job.

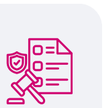
External monitoring

Signals about the system from outside its boundary.

Vendor changes, public incidents, and shifting regulation. None of it shows up in your own logs, and it’s often the highest-impact trigger.

What happens in the outside world matters

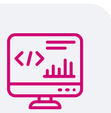
Contextualizing risk signals happening outside of your organization is critical to understanding the potential and experienced risks of your AI systems



Regulatory & legal

New rules, guidance, litigation, and enforcement that change your exposure.

Know which obligations just shifted for which systems, before an auditor does.



Vendor & model updates

Model swaps, deprecations, pricing changes, and provider advisories.

Catch a silent model change under a live use case before it moves your metrics.

Incidents

Public failures tied to the models, vendors, or domains you depend on.

Learn from someone else's failure while it's still someone else's failure.

Thank you!

Website	www.trustible.ai
E-mail	contact@trustible.ai
LinkedIn	https://www.linkedin.com/company/trustible/
HQ address	1201 Wilson Blvd, Floor 9, Arlington, VA 22209

Trustible