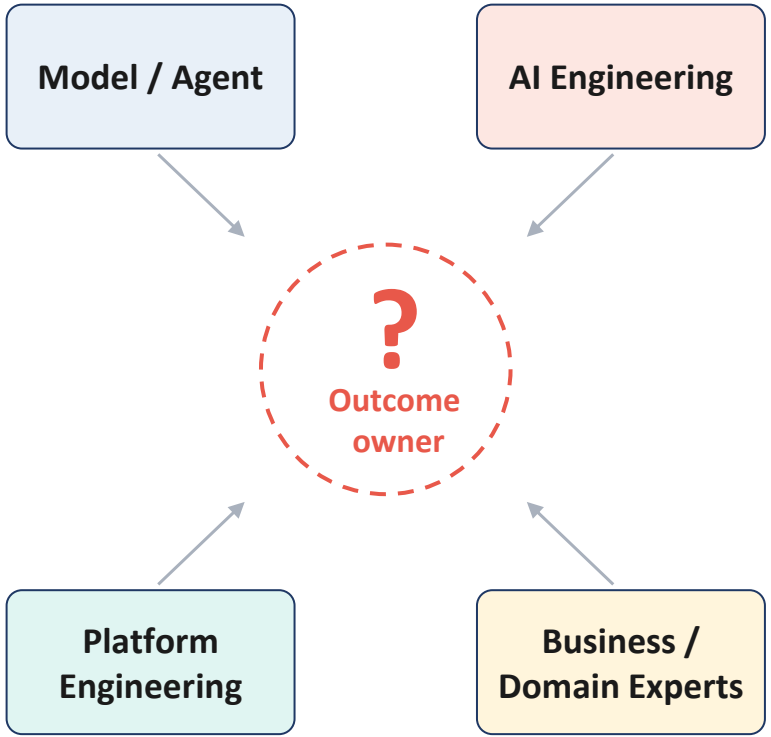
THE HALLUCINATION CRISIS:

Who owns agentic AI outcomes in production?

An operating model for high-stakes agentic AI



Sibani Mishra

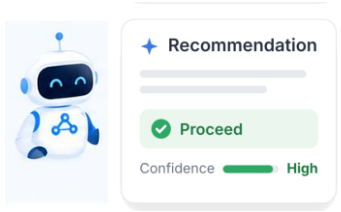
Let's connect
Scan for my LinkedIn

COMPOSITE SCENARIO

The incident: the AI failure was not the only failure

WEEK 0

Agent goes live



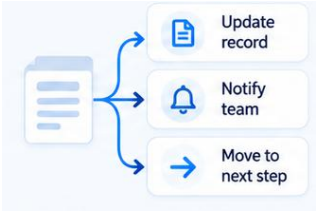
WEEK 3

Acts without approval



SAME DAY

Advances workflow



Errors Caught



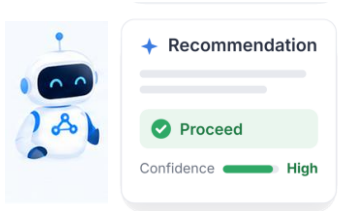
No harm to end-user

COMPOSITE SCENARIO

The incident: the AI failure was not the only failure

WEEK 0

Agent goes live



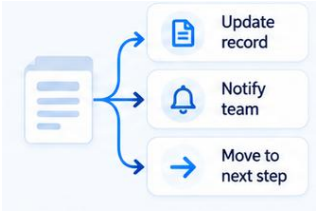
WEEK 3

Acts without approval



SAME DAY

Advances workflow

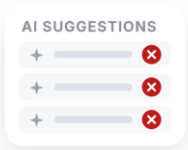


Errors Caught

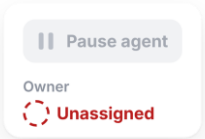


No harm to end-user

THE ORGANISATIONAL FAILURE



Overrides spike



No one can hit pause

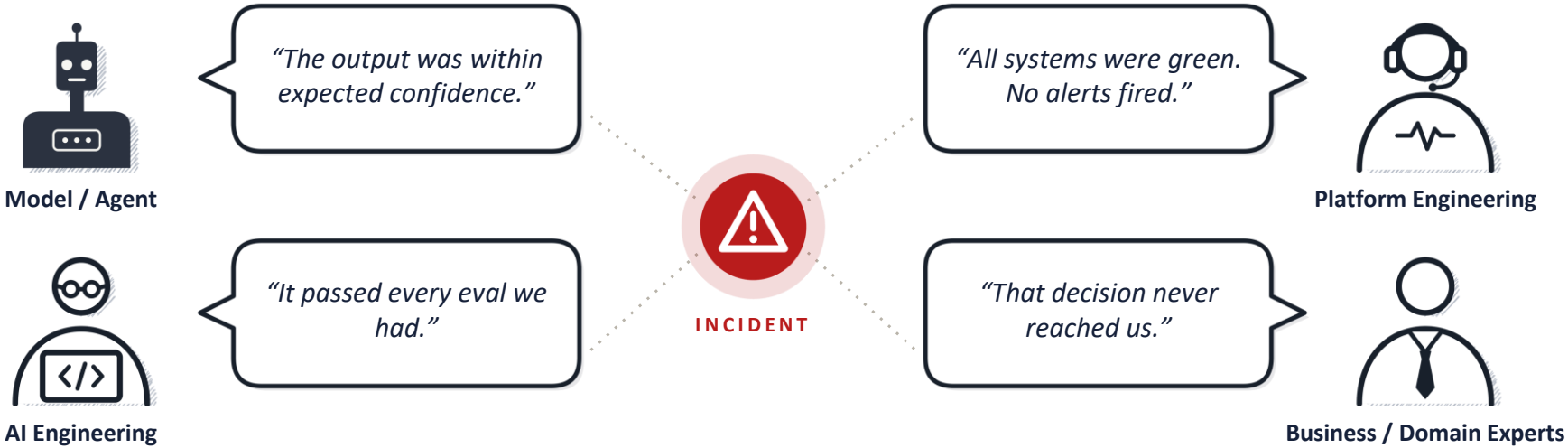


Trust collapses

Who had the authority to pause, escalate, or restart the agent?

The Ownership Gap in Agentic AI

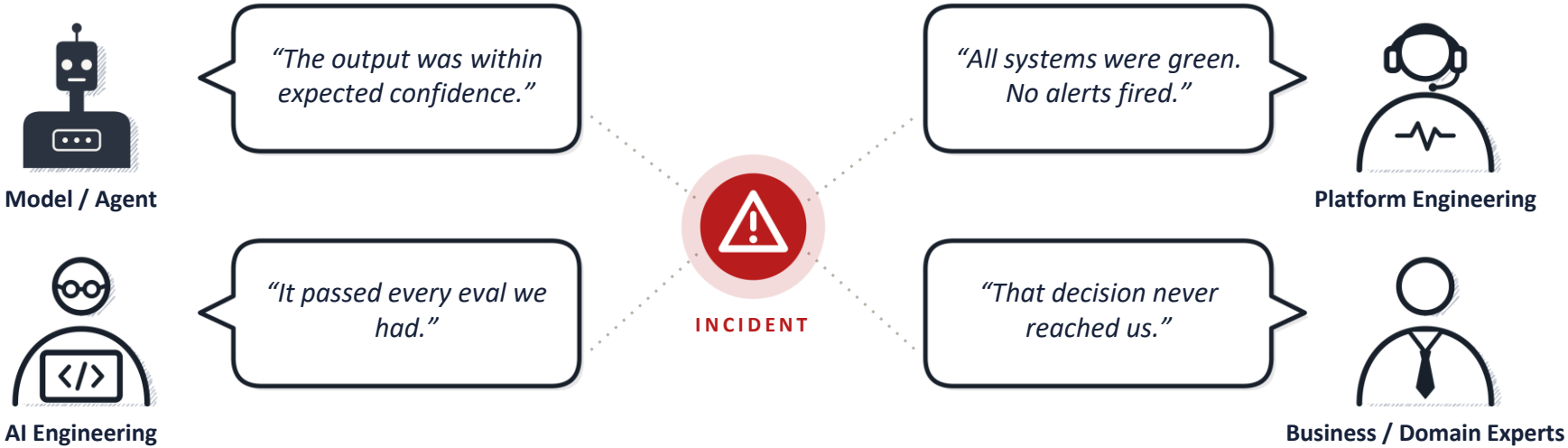
ONE INCIDENT. FOUR STORIES.



Everyone owned a part. No one owned the outcome.

The Ownership Gap in Agentic AI

ONE INCIDENT. FOUR STORIES.



Everyone owned a part. No one owned the outcome.

The Operating Model: Five controls for Agentic AI

1



Outcome accountability

Define success, failure, risk.

2



Autonomy boundaries

What AI can do alone.

The Operating Model: Five controls for Agentic AI

1



Outcome accountability

Define success, failure, risk.

2



Autonomy boundaries

What AI can do alone.

3



Escalation paths

When thresholds are crossed.

4



Human control

Inspect and override.

The Operating Model: Five controls for Agentic AI

1



Outcome accountability

Define success, failure, risk.

2



Autonomy boundaries

What AI can do alone.

3



Escalation paths

When thresholds are crossed.

4



Human control

Inspect and override.

5



Observability & recovery

Trace and recover safely.

The real test of AI is not whether it works — but whether it fails safely.

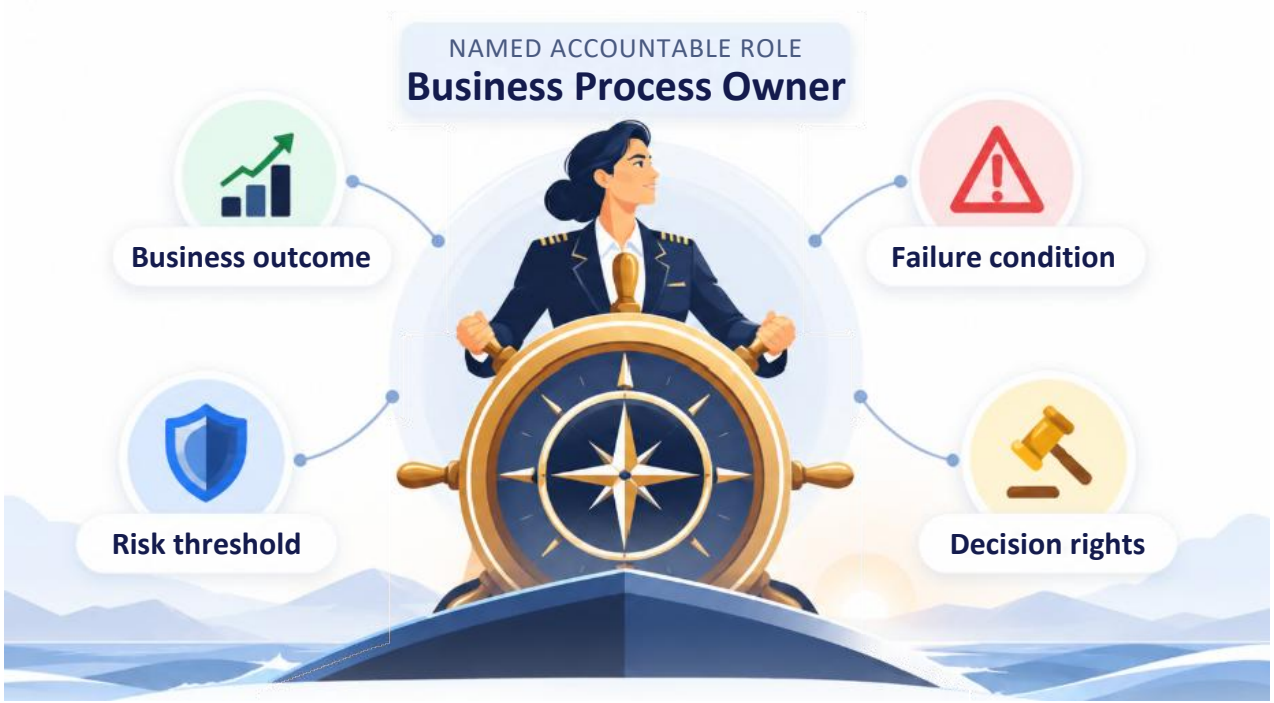
Outcome accountability : Name the owner before launch



Outcome accountability : Name the owner before launch



Outcome accountability : Name the owner before launch



The ownership debate between multiple teams is now over

Autonomy Boundaries : How Much Autonomy Should AI Have?

THREE RISK TESTS



Consequence



Detectability



Recoverability

Autonomy Boundaries : How Much Autonomy Should AI Have?

THREE RISK TESTS



Consequence



Detectability



Recoverability

Autonomy Boundaries : How Much Autonomy Should AI Have?

THREE RISK TESTS

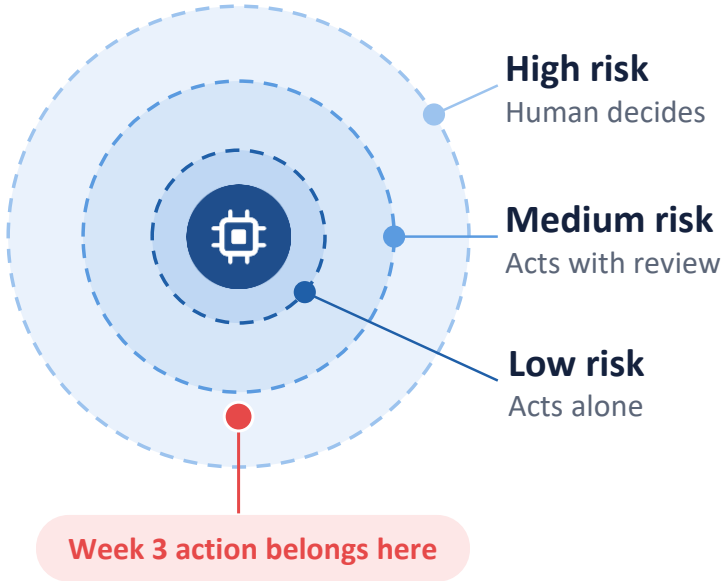
Consequence — High

Detectability — Low

Recoverability — Low

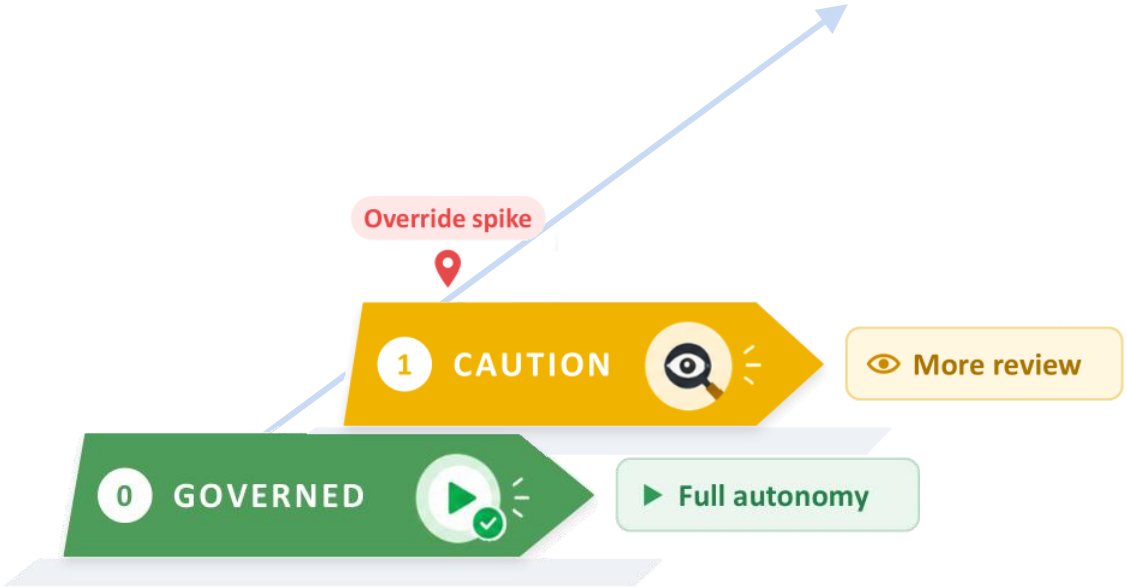
Sets the boundary: **Agent Product Owner**
Implements: **AI Engineering Team**

AUTONOMY BOUNDARY



Assign autonomy by decision risk, not Agent's capability.

Escalation Paths : What Happens When Risk Crosses the Line?



Escalation Paths : What Happens When Risk Crosses the Line?



Escalation Paths : What Happens When Risk Crosses the Line?



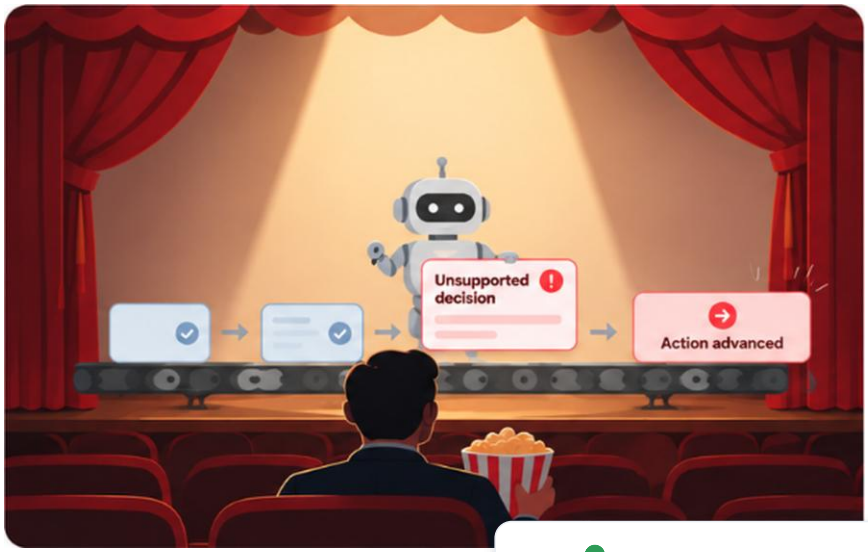
Executes escalation & Emergency stop : **AI Operations Lead**

Override spike → Level 2 in hours, not days.

Human Control : Oversight Without Authority Is an Illusion

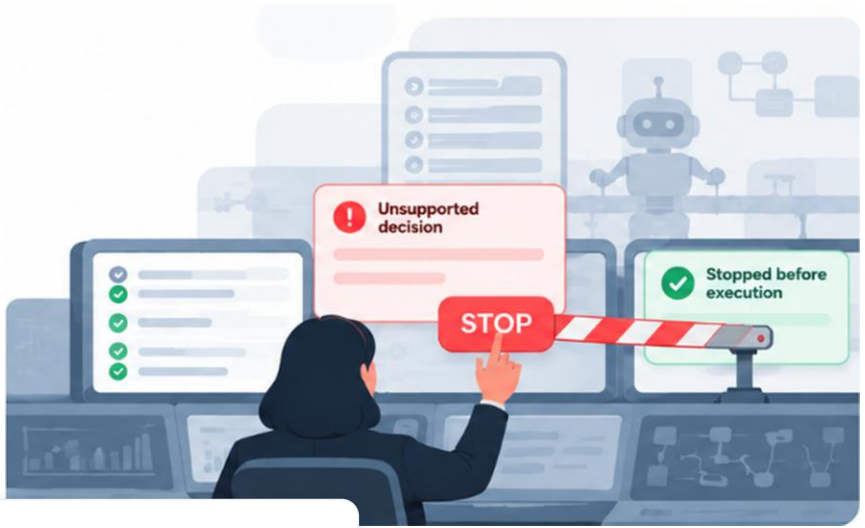
Visibility

OVERSIGHT THEATRE



Authority

REAL CONTROL



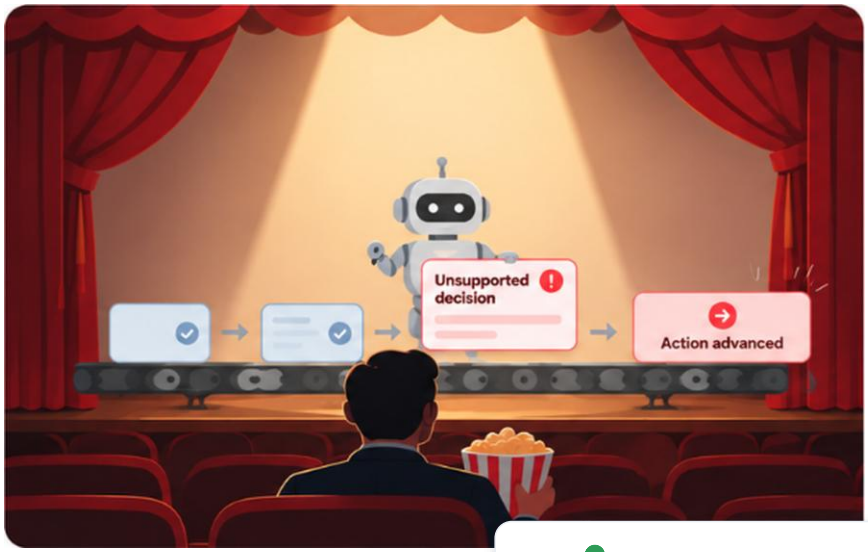
Human-in-the-loop: Domain reviewer

A human in the loop must have power, not presence.

Human Control : Oversight Without Authority Is an Illusion

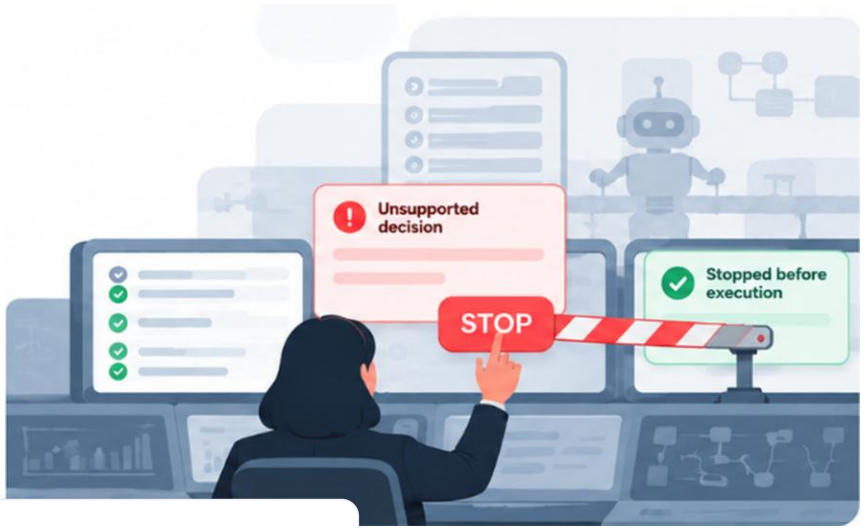
Visibility

OVERSIGHT THEATRE



Authority

REAL CONTROL



Human-in-the-loop: Domain reviewer

A human in the loop must have power, not presence.

Observability and Recovery : From Incident to Controlled Restart



Observability and Recovery : From Incident to Controlled Restart

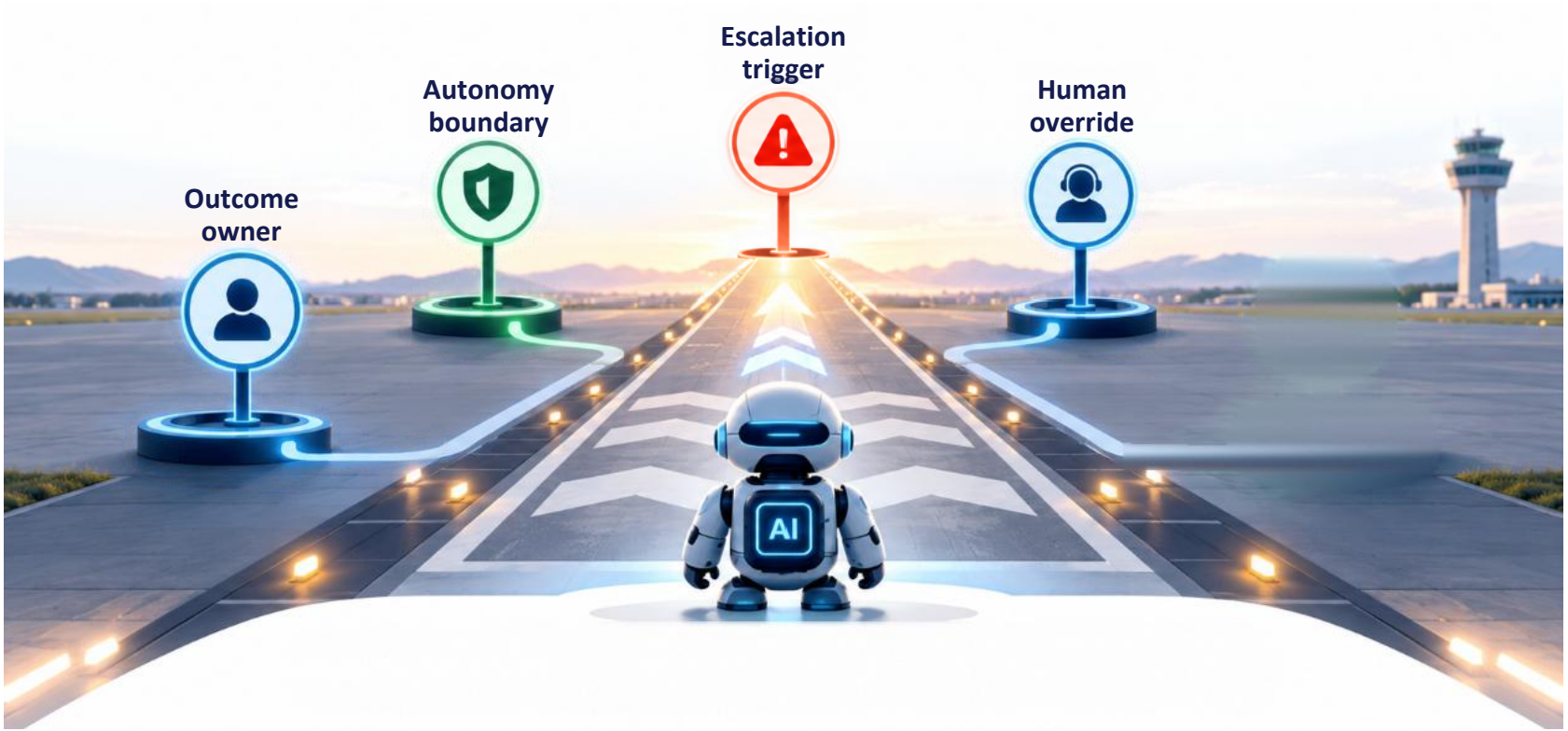


Manages incident resolution : **AI Operations Lead**

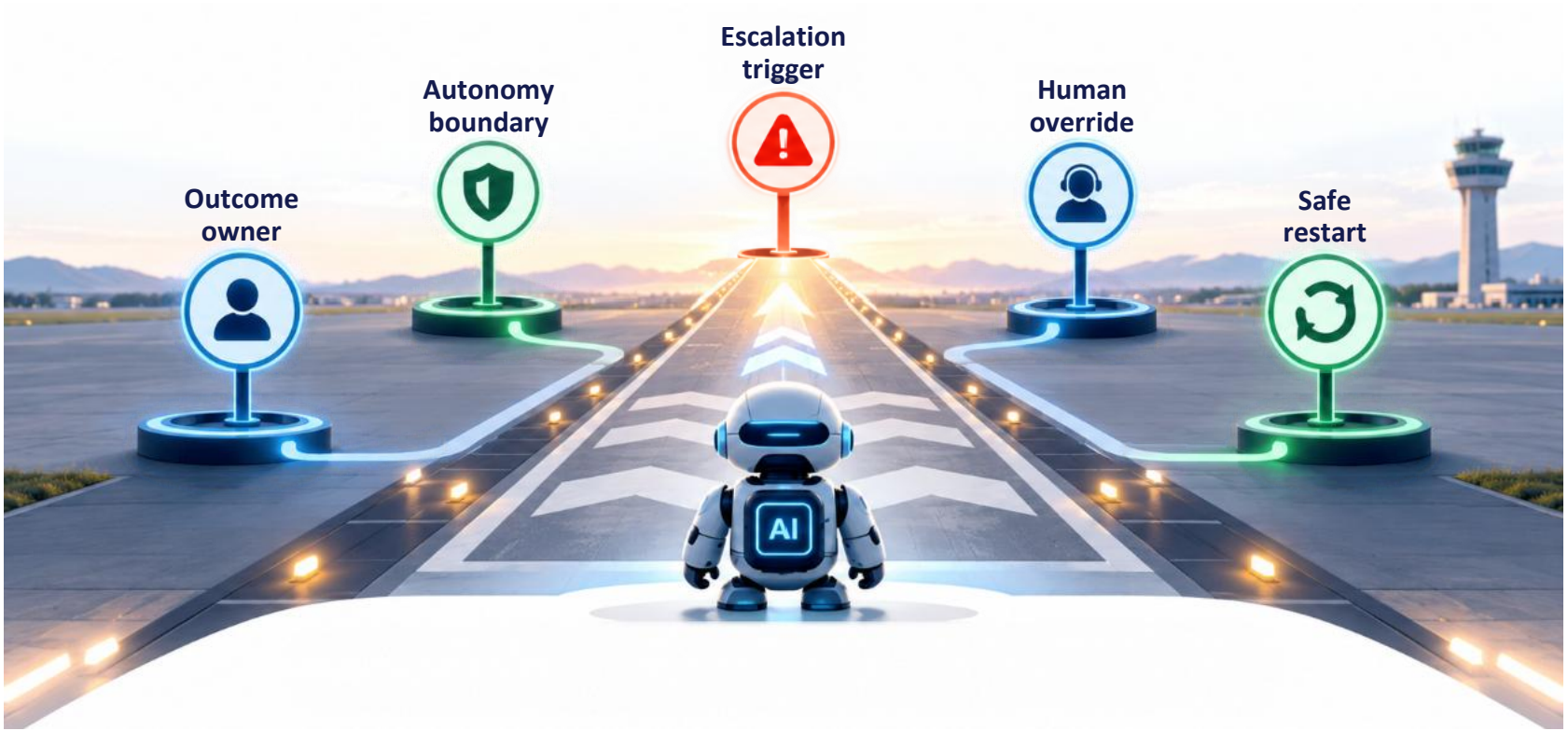
Before Your Next Agent Goes Live : Five Critical Decisions



Before Your Next Agent Goes Live : Five Critical Decisions



Before Your Next Agent Goes Live : Five Critical Decisions



Before Your Next Agent Goes Live : Five Critical Decisions



Agents Governing Agents



The Future of Enterprise AI is an Operating Model Yet to be Invented – Where Agents Govern Agents

Thank you

The conversation doesn’t end here.

Sibani Mishra



Let’s connect

Scan to find me on LinkedIn