Insight AI GovOps

The Operating Model for Enterprise AI Engineering

Solution Area

Cloud and Infrastructure

The Shift: From Per-seat to Metered AI

AI engineering tooling has moved from predictable per-seat licensing to metered, usage-based billing (UBB). GitHub Copilot now layers premium model calls billed per request; Agentic SDLC capabilities - coding agents that autonomously generate code, run tests, raise pull requests and trigger deployments - consume compute per task.

Spend now scales with **developer behaviour, AI business users and agent autonomy,** not headcount. Organisations without the right operating model face uncapped, unattributed AI spend - and no line of sight from cost to value.

Benefits

- Clear current state and evidence-based next steps
- Predictable AI spend, clear accountability, reduced risk, continuous optimisation
- Controlled models, reduced shadow AI, auditable compliance, regulator-ready
- Faster governed delivery, scalable agentic AI within guardrails, without Insight

The Governance Gap

Managing this shift requires more than cost controls. It demands an operating model that aligns three disciplines:

- Financial accountability - attribute AI spend to the project, team, and outcome it serves
- Policy and compliance - govern which AI services are approved, how data is handled, and how agents are authorised to act
- Engineering discipline - embed guardrails directly into how agentic software is designed, shipped, and maintained

Insight AI GovOps gives enterprises a single operating model to fund, govern, and build with agentic AI - turning unpredictable consumption into predictable economics, provable compliance, and faster, safer delivery. It spans every AI surface your users touch: developer tools, AI models, autonomous agents, and AI-infused applications.

The objective is not to restrict which model an engineer uses – but to ensure model selection, agent behaviour and AI consumption occur within defined financial, security, compliance and engineering guardrails.

Insight AI GovOps - Four offerings, one operating model

Designed to be mobilised individually or as an end-to-end programme - starting wherever your pain is greatest.

- AI GovOps Diagnostics
- AI FinOps
- AI Trust & Compliance
- Agentic SDLC

1. AI GovOps Diagnostic

Duration: 1 – 2 weeks

The right starting point for any organisation experiencing cost shock or governance uncertainty. Insight conducts a rapid current-state assessment of your AI tool landscape, billing structure, identity architecture, and compliance posture - then delivers a prioritised remediation roadmap.

What Insight Delivers

- AI tool and agent landscape inventory - what is running, what is billing, what is unmanaged.
- Cost attribution gap analysis - where spend cannot be tied to team, project, or outcome.
- Compliance and shadow AI exposure review - approved vs unapproved AI surfaces and data-handling risks.
- Developer goal alignment baseline - how AI productivity is currently measured (or not).
- Prioritised findings and recommended programme - phased, appropriate to your specific situation.
- Executive briefing: plain-language diagnosis with investment/ROI case for each recommended next step.

2. AI FinOps

Duration: 2 – 4 weeks

While GitHub and Azure provide budget controls and usage reports, Insight adds the operating discipline that organisations need to attribute AI spend to project outcomes.

What Insight Delivers

- Workshop-driven attribution model design explore cost allocation patterns within GitHub and AI platforms
- Identity and attribution foundation -Entra ID and GitHub Enterprise mapped to cost centre, team, repository, and project ownership.
- Consumption metering and token tracking across Copilot, premium models, and agents.
- Budget, forecast, and anomaly detection -spend gates and alerts aligned to business outcomes.
- ROI measurement methodology -pre-AI productivity baseline, measured delivery uplift normalised against AI cost to express value in dollars.
- Embedded operating cadence - a repeating governance rhythm your team owns after Insight leaves.

3. AI Trust & Compliance

Duration: 4 – 6 weeks

Governance lets you move from blocking AI tools to allowing them by default -with controls that satisfy your board, your auditors, and your regulators. Operationalise compliance across every AI surface your organisation touches -Developer Tools (GitHub Copilot, Copilot Studio), AI Foundry models, AI-infused applications, and agentic workflows

What Insight Delivers

- Approved model registry and services catalogue, with risk classification by model and use case.
- AI tool and agent landscape inventory - what is running, what is billing, what is unmanaged.
- Data-handling and classification policies covering all AI surfaces, not just developer tooling.
- Agent permissions framework and prompt governance standards.
  Shadow AI reduction programme-identification and remediation of unapproved AI adoption.
- Consistent identity, data, and threat controls applied across every agent.
- Defensible evidence of governance, ownership, and policy enforcement when regulators or auditors ask

4. Agentic SDLC

Duration: 8 – 10 weeks

Embed guardrails directly into how agentic software is designed, shipped, and maintained -lifting developer productivity while keeping autonomous AI auditable, reliable, and sustainable as it scales. A working architecture, embedded practice, and an enabled team—not just documentation.

What Insight Delivers

- Engineering leadership workshops -co-design agent patterns, orchestration boundaries, and human-in-the-loop control points with your senior engineers.
- Working reference architectures -not documents; actual implementation patterns integrated into your repository and pipeline standards.
- Model routing policies and agent orchestration with enforced permission boundaries.
- Repository policies, branch protection, and AI code-review controls integrated into CI/CD.
- Human approval gates, testing requirements, and release/deployment governance aligned to agent
- risk tier.