

Seed - Edinburgh, United Kingdom

# Kittiwake

A runtime guard for AI agents that checks every tool call against policy and keeps evidence for audits

The problem

# What's broken today

Companies are giving AI agents access to email, payments and internal databases, but the controls are mostly prompts asking the model to behave. A single poisoned web page or support ticket can instruct an agent to forward customer data or approve a refund, and afterwards security teams often cannot reconstruct which instruction led to which action.

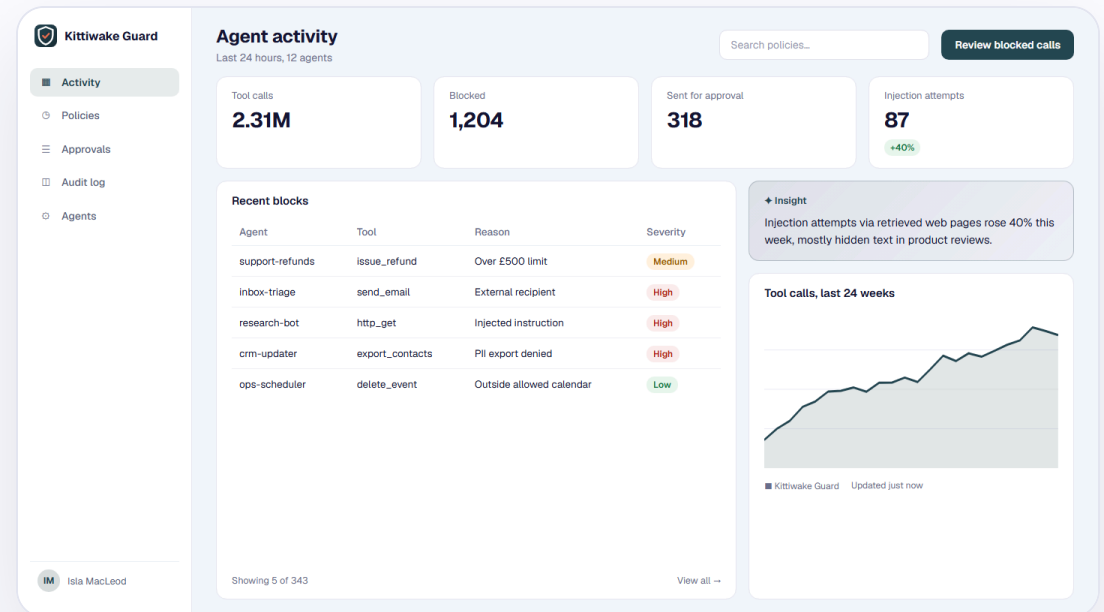


Kittiwake

Our solution

# Kittiwake Guard

Kittiwake sits between the agent and its tools as a proxy. Each tool call is checked against written policies, such as spend limits, data classes and allowed recipients, and against a classifier that spots injected instructions in retrieved content. Risky calls are blocked or sent for human approval, and every decision is logged with the inputs behind it.



# Blocked call BLK-90412

 Kittiwake Guard

Activity

Policies

Approvals

Audit log

Agents

Blocked call BLK-90412

inbox-triage agent, send\_email

Search policies...

Add sender to watchlist

Injection score

0.97

Policy hits

2

Data class

Customer PII

Decision

Blocked

Decision trace

| Step | Input              | Check                | Result                |
|------|--------------------|----------------------|-----------------------|
| 1    | Ticket #77120 body | Injection classifier | Hidden instruction    |
| 2    | Recipient domain   | Allowed recipients   | Not on allowlist      |
| 3    | Attachment         | Data class scan      | 42 customer emails    |
| 4    | Agent rationale    | Policy match         | Followed ticket text  |
| 5    | Final              | Combined decision    | Blocked, team alerted |

Showing 5 of 64View all →

◆ Insight

The instruction was white text inside the ticket's HTML signature, invisible to a human reader.

 Kittiwake

IM

Isla MacLeod

Why now

# The timing

Agents moved from answering questions to taking actions in 2025, and auditors now ask companies to show who approved what an agent did.

Market

Banks, insurers and software companies deploying agents with write access to real systems; agent security budgets are forming inside application security spend, estimated at US\$8B a year.



Kittiwake

# How we make money

Subscription priced by screened tool calls, starting at US\$2,000 a month for 5M calls, with an enterprise tier for self-hosting and audit exports.

Traction

## Where we are

**7**

Paying customers

**US\$34K**

Monthly revenue

**2.3M**

Tool calls screened per day

**0.4%**

False-block rate

- 7 paying customers, including two regional banks and an insurer
- 2.3M tool calls screened per day with a 0.4% false-block rate
- Classifier catches 96% of attacks in our 12,000-case injection test set



Kittiwake

Team

# Who's building it

IM

**Isla MacLeod**

CEO and Co-founder

KA

**Kwame Asante**

Chief Scientist and Co-founder

FL

**Fergus Lindsay**

Head of Engineering



Kittiwake

The ask

# What we're raising

**Raising US\$3.5M seed to reach 30 enterprise customers by 2027**

- Build connectors for the main open agent frameworks
- Hire two security engineers and a solutions lead
- Complete an independent security certification