



MANAGED INTELLIGENCE RESOURCE

Jev on 5 real jobs. One honest teardown.

Jev is a small AI model that answers one question you define, in one of three shapes: yes or no, pick one, or score it. That is the whole trick, and it is why it is fast and cheap enough to run on every ticket, email, and comment your business touches. We took a practitioner's twelve use cases and measured what actually holds up.

USE CASES TESTED

5

1,000 EMAILS TRIAGED

~6 sec

~20,000 ROUTING CALLS

\$0.85

CONTEXT CEILING

64k

CONTENTS

What's inside

01 What Jev is

Three answer shapes, three hard limits

02 The 5 use cases

The five worth your queue, tested

03 The honest numbers

Vendor claims vs. what was actually measured

04 The evaluation gate

Golden set, shadow week, automate one branch

05 Keep exploring

Links and next steps

HOW TO READ THIS

Each part is a short sequence. Read the moves in order, and by the last page you will know exactly where a judgment model earns its keep in your stack, and where it does not.

PART 01

What Jev is

A normal model reads, reasons, and writes a reply. Jev reads and answers one question you defined. That is the entire difference.

SHAPE 01 Yes or no, with a confidence score

Ask a binary question, get an answer plus how sure it is. A support ticket asking "is this urgent?" came back **99% yes**. Raise the 50% default floor for anything that pages a human.

SHAPE 02 Pick one of your options

Hand it your list of queues, teams, or categories and it routes. A Hyper-V outage ticket went straight to **Infrastructure**, no prompt engineering beyond the option list.

SHAPE 03 Score on your scale

Rate intensity, fit, or risk on a scale you define. The same ticket scored **4 out of 5** on SLA risk, the number that decides whether a tech gets woken up.

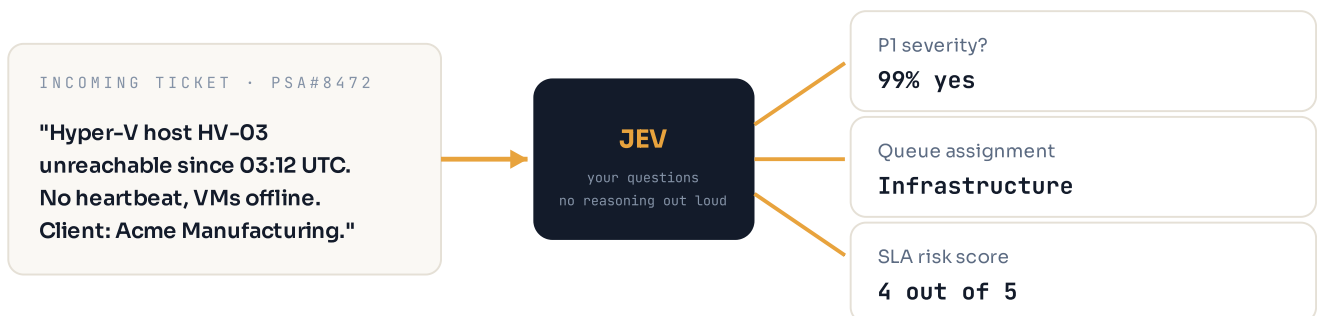


FIGURE 1 — ONE TICKET IN, THREE JUDGMENTS OUT.

HERE'S THE READ

It judges. It never writes. If a job needs a drafted reply, that is a different tool; Jev decides what happens next.

PART 02

The 5 use cases

Five of the twelve tested jobs survived scrutiny. Each was run by the practitioner on a real queue; the numbers are his measured runs.

01 • INBOX

Email triage

Score every inbound email for urgency and route it. **1,000 emails in ~6 seconds, ~\$0.09.** That is the cost of never missing a hot lead.

02 • INBOX

Support ticket routing

Assign each ticket to the right queue on arrival. **~20,000 routing requests for \$0.85 total.** You stop paying a human to read and forward.

03 • BUSINESS

Lead qualification

Score inbound leads against your own fit criteria before a rep spends a call on a bad fit. The score is yours; the threshold is yours.

04 • BUSINESS

Document sorting

Route invoices, resumes, and PDFs to the right folder or person on arrival. A direct swap for the most tedious admin hour in the business.

05 • LAB

Paper trading

Test buy and sell signals against real market data. **~\$2 a day**, 1,000+ signals. The tester's own verdict: **"not very vetted."** A lab, not a strategy.

HERE'S THE READ

The other seven tested jobs were variations on these five. If one of them matches a queue you already have, that is your pilot.

PART 03

The honest numbers

Two kinds of numbers show up in every Jev thread: what the vendor claims, and what the practitioner actually measured. Keep them in separate columns.

CLAIM	WHAT WAS MEASURED	STATUS
12x cheaper than a frontier model	\$0.85 for ~20,000 routing requests	VENDOR CLAIM
46x faster than a frontier model	~6 sec for 1,000 email triage calls	VENDOR CLAIM
Cheapest judgment layer available	\$0.07 per 1,000 judgments on the cheapest tier	MEASURED
Reads a full report in one pass	64k context window, hard ceiling	MEASURED

12x

CHEAPER (VENDOR)

46x

FASTER (VENDOR)

~\$2

PER DAY, PAPER TRADING

LIMIT 01

64k context ceiling

Anything longer than a long report gets truncated. Feed it summaries, not archives.

LIMIT 02

Zero writing ability

It cannot draft a reply, a summary, or a sentence. Pair it with a writing model or a human.

LIMIT 03

One narrow job

One question per call. Chained reasoning is what frontier models are for; do not ask Jev to be one.

HERE'S THE READ

Cheap and fast are real. General intelligence is not on offer. Buy it for the queue, not for the brain.

PART 04

The evaluation gate

Before a single judgment touches a live queue, run it through three moves. This is the same shadow-mode rollout pattern we use with clients.

MOVE 01

Build a golden set

Collect **100 real examples** with the answers you already know are right. This is the ruler every later run gets measured against, and it is the only proof that matters.

MOVE 02

Score both models

Run the golden set through Jev and through the frontier model you would otherwise pay for. Compare answers against the ruler, not against each other. If Jev misses your accuracy bar, stop here; you have spent dollars, not weeks.

MOVE 03

Shadow week, then one branch

Let Jev watch the live queue for a week and log what it *would* have done. Compare its log against what your team actually did. Then automate only the branch where it agreed with you most, and keep a human on everything else.

100

GOLDEN EXAMPLES

1 wk

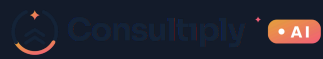
SHADOW, NO ACTION

1

BRANCH AUTOMATED FIRST

ORDER MATTERS

Measure before you trust, watch before you automate. A judgment model earns its seat one branch at a time.



KEEP EXPLORING

Jev model review

The practitioner walkthrough this guide is built from

geeky-gadgets.com/jev-ai-model-review

MIP Resource Center

Templates and teardowns for AI in a real business

consultiply.ai

Want this built in your business?

consultiply.ai