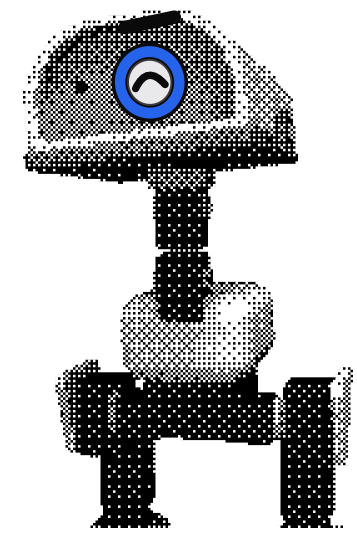
// how much better is it?

Same size. *Smarter* decisions.



We put our AI next to other AI models of the same size on a public test of everyday decisions, and show the jumps it makes on hard problems. Every score here is out of 100: **higher is better.**

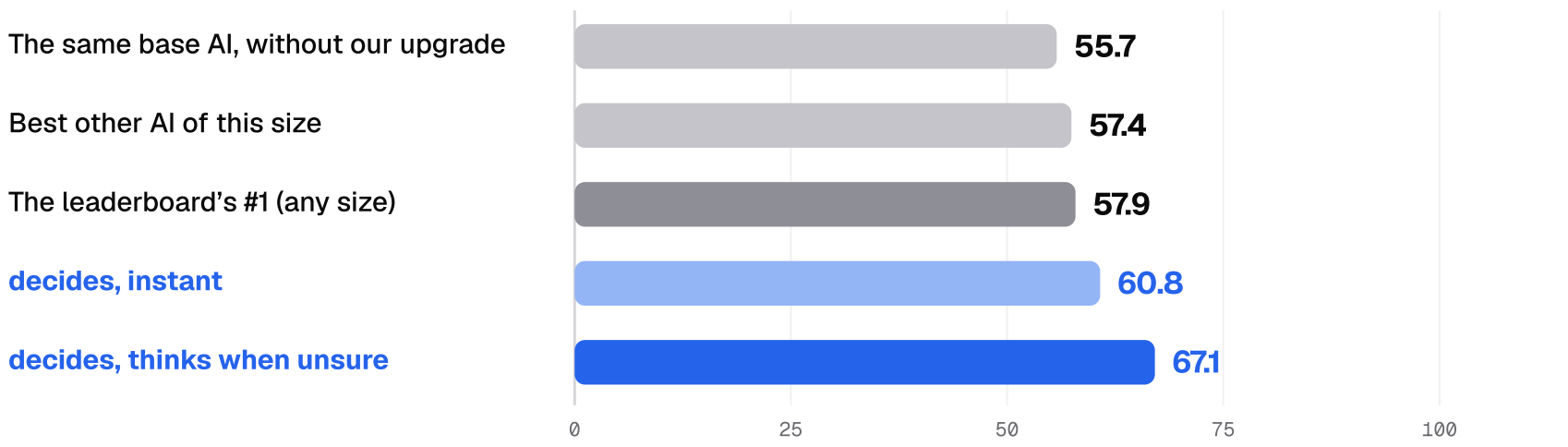
// 1 — against other AIs of the same size

Ahead of every model its *size*

The Decision Index is a public test of 38 everyday decision tasks, like sorting support tickets or picking the right tool.

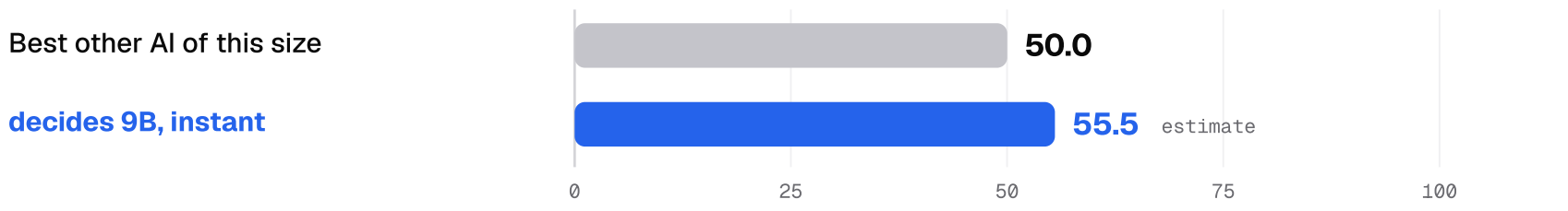
Large models (about 27B)

B = billions: a rough measure of an AI's size



Mid-size models (9B to 12B)

our smaller version



// 2 — the big jumps

Where it gets *much* better

Before: the same AI answering in one go. After: with our upgrades.

Hardest logic puzzles

24% → **70%**

checks and fixes its work +46

Medium logic puzzles

27% → **82%**

checks and fixes its work +55

Hard science questions

29 → **69**

thinks when unsure +40

Spotting phishing emails

26 → **70**

thinks when unsure +44

A tough public decision test (JevBench, hard set)

73% → **91%**

thinks when unsure +18

Reading science diagrams

79 → **86**

its built-in decision skill +7

// 3 — what this means for you

One AI that's quick, careful and *checks* itself

Quick

Answers in about 0.17 seconds when it's sure, and tells you how sure it is.

Careful

Stops to think only when it isn't sure, so hard questions get the attention they need.

Self-checking

Checks its own answer against the rules and fixes what's wrong.

// decides.io

One model, *not two.*

Open source · coming soon

decides.io

Scores are out of 100. Ours are our own runs, not yet submitted to the leaderboard. Other models' scores were measured by the leaderboard (Decision Index, September 2026). The 9B score is estimated on a large sample of the test. Built on open models: Qwen3.8-27B and Cloudflare's Clef, both Apache-2.0.

