WEKA

Peer Intelligence Report

You answered four questions on the floor at RAISE. Here's what we heard: what your peers are prioritising, where the gaps are, and the one thing everyone is thinking but no one is saying out loud

80	84%	1/3
Completed Responses	European Respondents	C-Suite, Founder, or VP

The gap between instinct and budget.

Ask what's holding AI back and the answer is familiar: not enough GPUs. Ask what they measure, and where the next dollar actually goes, and the answers stop matching.

36%

Say compute availability is the bottleneck

40%

Measure cost per token above all else

28%

Would put the next dollar into more GPUs

49% of the room would spend their next dollar making existing infrastructure work harder.

The instinct says buy. The budget says optimise. Even among those who named GPU scarcity as their constraint, half measure cost per token, an economics number. This room stopped believing more hardware fixes the problem; it just hasn't admitted it yet.

Where the next dollar goes

Sovereignty stopped being a burden.

For years the industry framed sovereignty as Europe's tax on ambition. Not this crowd.

Europe has stopped apologising for sovereignty. The majority position: owning where your data lives and who governs it is how you win, not a constraint to route around. The minority still treating it as a compliance line item are now negotiating against competitors who treat it as strategy.

Call sovereignty a competitive advantage, rising to 55% among European respondents.

See it as compliance cost.

Say it’s overstated

#1 Data Residency is the most-cited decision factor.

The metric of the moment is cost, not speed.

40% Measure cost per token or per inference — two and a half times the next measure on the list.

What the room measures

Inference is now the economic centre of AI, and this room's dashboards prove it. Nobody's asking how fast anymore. They're asking what every token costs. Two more signals from the data:

Power is the second constraint.

20% named energy the primary bottleneck, ahead of memory, talent, and capital. Tokens per watt is already on the dashboard for 1 in 5 of your peers.

The blocker is bandwidth, not belief.

Asked what's holding back the efficiency play, the top answer wasn't ROI doubt or immature tooling. It was team bandwidth. Nobody needs convincing. They need it to run without consuming their people's time.

Which one are you?

- The Economics Constrained: You measure cost per token, and it's scaling faster than usage.
- The Data Starved: Your GPUs are idle, waiting on data.
- The Sovereignty-Committed: Data residency shapes your architecture.
- The Bandwidth Blocked: You believe in efficiency but can't staff it.

Whichever one you are, the fix is the same. Make the infrastructure you own do more, and let your team do less.

WEKA's read

You and your peers described the problem WEKA® NeuralMesh™ was built for. When 40% of leaders measure cost per token, when half the budget goes to optimising over buying, and when the constraint on efficiency is your team's time, the move isn't buying more hardware.

- It's making the hardware you already own earn its keep.
- GPUs stay fed instead of idling.
- Repeated tokens come from memory instead of a fresh recompute.
- And it runs sovereign, in your territory, without needing a team you don't have.

TESTED BY COREWEAVE IN PRODUCTION

Let's Talk