

AI infrastructure translated

The top 10 AI infrastructure terms every leader needs to know—minus the jargon.

AI factory

An operational setup that runs your AI infrastructure like a digital assembly line. Raw data goes in, and ready-to-use AI models and instant responses come out the other side automatically.

Agent sandboxing

Isolating autonomous AI code inside a secure digital vault. This stops experimental agents from accessing sensitive systems or networks, ensuring your core enterprise data remains protected.

Amdahl's law

Adding extra processors won't speed up AI tasks that must run step-by-step. To prevent overspending on redundant hardware, engineering teams must instead focus on removing sequential workflow bottlenecks.

Evaluating for co-design

Investing in technology built to work together from day one instead of asking engineers to stitch together disparate parts. This optimizes software for specific hardware, significantly reducing deployment timelines.

Fallback lists

An automatic backup plan for chip and storage selection. If your primary choice is unavailable, the software instantly provisions pre-approved secondary choices so jobs keep moving without human intervention.

Goodput

The percentage of your system's power, hardware, and network that does useful AI work, subtracting the time lost to system resets, hardware stalls, and background maintenance.

Jevons paradox

Making AI cheaper and more efficient drives widespread adoption. This accessibility helps business units discover new high-value applications, expanding AI's competitive impact across your organization.

Power usage effectiveness

A report card for data center energy efficiency, measuring how much power goes to computing versus overhead like cooling. A 1.0 is a perfect score, and lower is always better.

Sim-to-real

An engineering methodology that trains AI models in virtual worlds before deploying them onto physical hardware. By simulating millions of edge cases, teams can accelerate deployment readiness while safeguarding real-world assets.

Time-to-first-token (TTFT)

The delay between a user hitting enter and the AI beginning its response. The wait time is the single biggest driver of perceived AI speed and user engagement.

Access The Encyclopedia of AI Infrastructure today to learn key terms and concepts (no jargon).

[Learn more](#)