INSIGHTS DEEP DIVE

# AI is a five-layer cake. Marketers get the top tier

NVIDIA's Jensen Huang slices AI into energy, chips, infrastructure, models and applications. Here's what each layer means for your budget, tools, data and customers, and who's in each one.

By **Nicole Fung** · Tuesday, November 3, 2026 · 9 min read

Jensen Huang has a dessert theory of the AI economy. Speaking at CSIS, the Washington think tank, in December 2025, NVIDIA's CEO put it plainly: "let's simplify AI into a five-layer cake: energy, chips, infrastructure, models, and applications."[1]

Most marketers only ever taste the top tier: the chatbot, the copy tool, the AI button in the CRM. But the layers below decide what that button costs, where your customer data goes and how easily you can switch. In Huang's words, "Every successful application pulls on every layer beneath it, all the way down to the power plant that keeps it alive."[2]

> *"Most marketers only ever taste the top tier."*

So before the next budget round, here's a slice-by-slice tour.

## The cake, in his words

In March 2026, Huang set the model out on NVIDIA's blog: "When you look at AI industrially, it resolves into a five-layer stack."[2] From the bottom up:

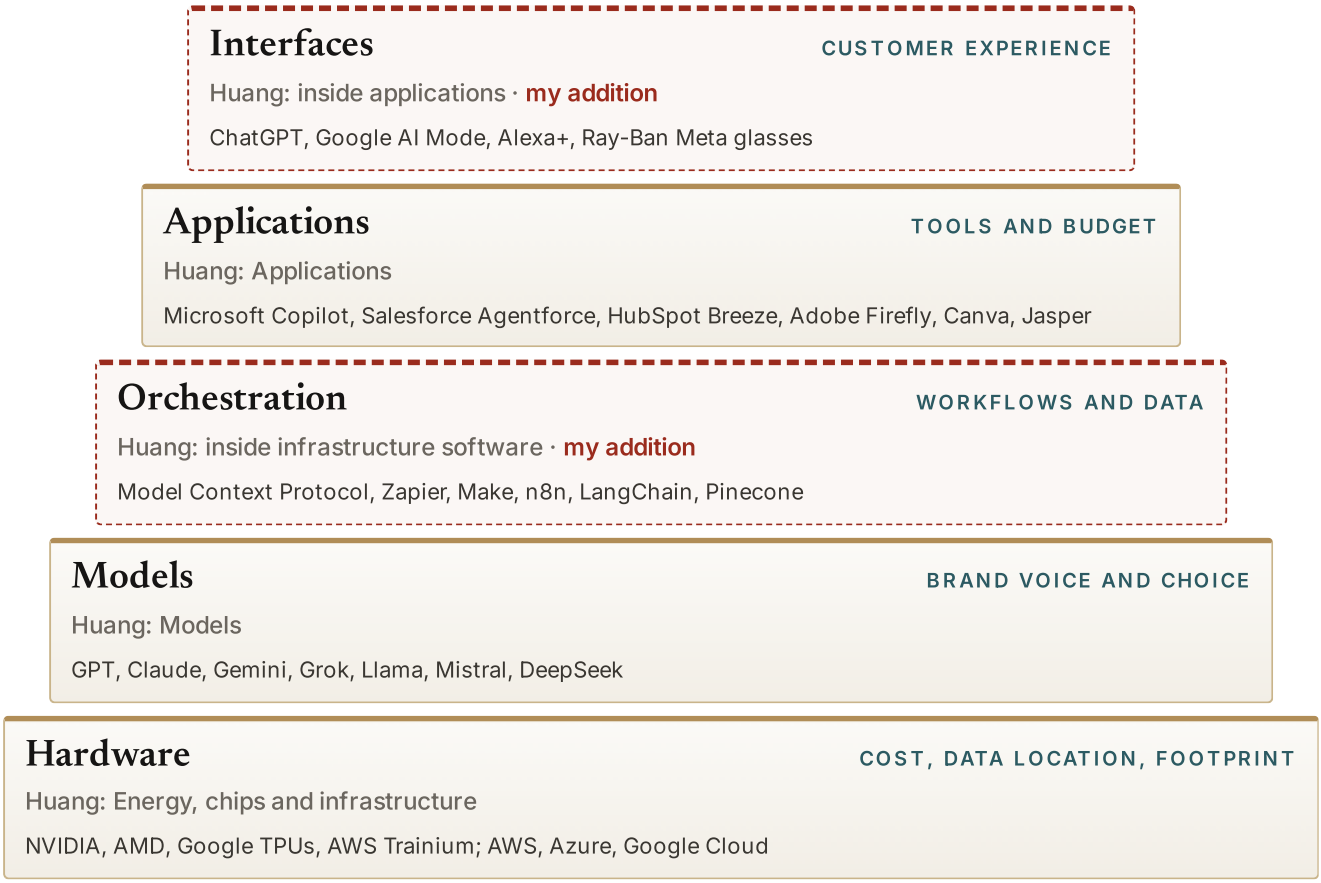
- **Energy.** "Energy is the first principle of AI infrastructure and the binding constraint on how much intelligence the system can produce."
- **Chips.** "Progress at the chip layer determines how fast AI can scale and how affordable intelligence becomes."
- **Infrastructure.** Land, power, cooling, networking and the systems that join tens of thousands of processors into one machine. "These systems are AI factories."
- **Models.** Not just chatbots. "Language models are only one category."
- **Applications.** "At the top are applications, where economic value is created."

## My version, for marketers

Huang's cake is built for people building frontier labs, data centers, and generally closer to the wholesale landscape. Marketers need a different stack, so here's my take. I fold his bottom three layers that marketers almost never touch into one, keep models and applications, and add two layers where AI-forward marketing actually happens: orchestration, which connects models to your data and tools, and interfaces, where customers and your marketing team meet AI.

THE MARKETER'S AI STACK

My adaptation of Jensen Huang's five-layer cake, top tier first. Dashed red tiers are my additions; the bottom tier folds his energy, chips and infrastructure into one.




Huang's order, bottom to top: energy → chips → infrastructure → models → applications

Product names are examples, not endorsements. Now, layer by layer, from the bottom.

## Hardware: cost, data location, footprint

The International Energy Agency expects data center electricity demand to "more than double by 2030 to around 945 terawatt-hours (TWh), slightly more than the entire electricity consumption of Japan today."[3] Meanwhile, chips keep getting better at turning that power into answers. Stanford's AI Index found that "the inference cost for a system performing at the level of GPT-3.5 dropped over 280-fold between November 2022 and October 2024," while hardware costs fell 30% a year and energy efficiency improved 40% a year.[4]

The names here include NVIDIA's Vera Rubin platform, AMD's Instinct GPUs, Google's TPUs and Amazon's Trainium chips. You will never buy one. You pay for them by the token.

This is the data centers and clouds: Amazon Web Services, Microsoft Azure, Google Cloud and a growing list of national projects. Huang's CSIS description of this layer was "infrastructure, basically software," plus land, power and capital.[1]

Denmark has its own slice. Gefion, an NVIDIA DGX SuperPOD with 1,528 H100 GPUs, went live in October 2024, hosted in a Danish data center that "runs on 100% renewable energy."[5] At GTC Berlin in October 2026, Huang made the case to Europe directly: "you don't want to just be consuming the AI; you want to be producing the AI."[6]

**What it means for marketers (strategy).** Treat AI prices as a moving target. The same task tends to get cheaper, so be wary of locking today's rates into multi-year deals, and don't pay for the biggest model when a smaller one will do. If your brand makes sustainability claims, AI use belongs in that conversation too, not in a footnote.

**What it means for marketers (data).** This is where your customer data physically sits when AI touches it. Ask vendors which region processes your prompts and files, which country's law applies and whether your inputs train their models. Our glossary explains data sovereignty and vendor lock-in in plain English.

## Models: the engine, not the car

Models are the famous names: OpenAI's GPT models, Anthropic's Claude, Google's Gemini, xAI's Grok, Meta's Llama, Mistral and DeepSeek. Huang called the first four just a handful of the more than a million AI models in the world.[1] Plenty of capable models are open source, as Huang attested, "Most of the world's models are free."[2]

Many apps already mix and match. Adobe Firefly offers Adobe's own models "plus partner models from Google, OpenAI, ElevenLabs, Luma AI, Runway, and others."[7]

**What it means for marketers (tools and brand voice).** Models are becoming interchangeable parts, and the leader changes often. So don't let your process and documents live inside only one. For example, keep your brand guidelines, positive examples, and approved content structured in a secured cloud that different team members and AI models can easily access. Whether you choose an open, closed, or hybrid model, choose the combination based on cost, privacy, and control, not on hype. This is also where data quality matters. So if your data is not good enough, consider using AI to do a data clean-up project for a good foundational start and spend the time to do a knowledge graph mapping to guide your AI's reasoning.

## Orchestration: the plumbing Huang doesn't name

Here's my first layer addition. Between a model and a useful result sits plumbing: connectors, workflows, memory and the rules that decide which model gets which job. Huang folds this kind of software into infrastructure. I pull it out, because this is where marketing tools get wired together.

The Model Context Protocol (MCP), which Anthropic launched as an open standard for "connecting AI assistants to the systems where data lives," is now hosted by the Linux Foundation.[8] [9] Workflow tools such as Zapier, Make and n8n chain steps together; LangChain does the same for developers; vector databases such as Pinecone store your content so AI can search it. Even Microsoft's Copilot now has an Auto setting that "weighs accuracy, speed, and cost for each request, helping determine the right model and effort for the task at hand."[10]

**What it means for marketers (workflows and data).** This layer decides whether AI can reach your approved assets, product data and brand guidelines, or just keep politely making good-sounding things up. Set up a workflow that mirrors the best lean process you wish your team had. Find out which parts you want AI to work on, what the iteration or approval loop looks like, and which parts are with or without a human. For the sections where AI can add speed and value, grounding its reasoning with a knowledge graph and using other agents for a quality check before the output reaches a human will save time and frustration. Once you have a good working process, automate the chain with clear roles and responsibilities for which nodes require human sign-offs.

## Applications: where the value lands

Huang is unambiguous about where the money is. At Davos he said, "This layer on top, ultimately, is where economic benefit will happen."[11] His blog lists customer service among the applications "already showing strong product-market fit."[2]

For marketers, this is the layer with invoices: Microsoft Copilot in Word, PowerPoint and Outlook; Salesforce Agentforce; HubSpot's Breeze; Adobe Firefly; Canva's AI tools; and Jasper, whose site invites you to "Put AI agents to work for marketing."[12]

**What it means for marketers (tools and budget).** Most of your AI arrives inside software you already pay for, often switched off or underused. Before buying a new tool, audit what your current ones can do, and check which model and which data each one uses underneath.

## Interfaces: where customers meet it

My second layer addition. Huang keeps this inside applications, but for marketers it deserves its own tier, because it is where your team interacts with the AI and where your customers do too. Interfaces include chat assistants such as ChatGPT and Gemini, Google's AI Overviews and AI Mode, voice assistants such as Amazon's Alexa+, and smart glasses such as Ray-Ban Meta.

The shift is measurable. Pew found that Google users who saw an AI summary clicked a traditional result in 8% of visits, versus 15% for those who didn't.[13] When the answer becomes the shop window, fewer people step inside. We covered what that means in Your next customer may never visit your website.

**What it means for marketers (customer experience).** Your brand now gets described by AI before it gets visited. Check how assistants in multiple AI platforms describe you, make sure your content is sharp, consistent, and differentiated among your competitors, and plan for the questions your prospects already have, because they will ask the AI and you would want the AI to quote you instead of stringing the answer up through probabilities.

## Why the whole cake matters

The cake is a useful corrective to thinking of AI as one product you either have or don't. Every tool you buy sits on a model, which sits on someone's data center, which sits on someone's power bill. Costs, risks and opportunities move up and down the stack. For the longer story of how we got here, see our AI History.

Huang's own advice at Davos was that "it is essential to learn how to use AI – how to direct it, manage it, guardrail it, evaluate it."[11] In the EU, teaching staff about AI is also a legal duty: since Feb 2, 2025, Article 4 of the AI Act has required organizations that build or use AI systems to take measures to support their staff's AI literacy, a rule the EU eased in July 2026 but kept.[14] And for anyone worried that the machines will take over the creative work, he told CSIS that "original thought, original writing, your taste made with human hands are still always going to be valuable."[1] Good news for marketers: that is the frosting.

**WHAT IT MEANS FOR YOU**

- You only buy from the top of the cake, but every layer shows up on your bill, in your data policy or in front of your customers.
- Map your own AI tools layer by layer: hardware, models, orchestration, applications and interfaces. Gaps and duplicates show up fast.
- Keep your brand voice, guidelines and data portable, so you can switch models without starting over.
- Ask vendors where your data is processed, which model they use and whether your inputs train it.
- Check how AI assistants describe your brand today. That's your newest shop window.

---

**SOURCES**


1. NVIDIA's Jensen Huang on Securing American Leadership on AI (transcript) (https://www.csis.org/analysis/nvidias-jensen-huang-securing-american-leadership-ai), Center for Strategic and International Studies, Dec 3, 2025
2. AI Is a 5-Layer Cake (https://blogs.nvidia.com/blog/ai-5-layer-cake/), Jensen Huang, NVIDIA Blog, Mar 10, 2026
3. AI is set to drive surging electricity demand from data centres while offering the potential to transform how the energy sector works (https://www.iea.org/news/ai-is-set-to-drive-surging-electricity-demand-from-data-centres-while-offering-the-potential-to-transform-how-the-energy-sector-works), International Energy Agency, Apr 10, 2025
4. AI Index Report 2025 (https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf), Stanford HAI, Apr 2025

5. Denmark's first AI supercomputer is now operational (https://novonordiskfonden.dk/en/news/denmarks-first-ai-supercomputer-is-now-operational/), Novo Nordisk Foundation, Oct 23, 2024
6. GTC Berlin 2026 keynote live updates (https://blogs.nvidia.com/blog/gtc-berlin-2026-keynote/), Brian Caulfield, NVIDIA Blog, Oct 21, 2026
7. Adobe Firefly (product page) (https://www.adobe.com/products/firefly.html), Adobe, accessed Oct 23, 2026
8. Introducing the Model Context Protocol (https://www.anthropic.com/news/model-context-protocol), Anthropic, Nov 25, 2024
9. Linux Foundation announces the formation of the Agentic AI Foundation (https://www.linuxfoundation.org/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation), Linux Foundation, Dec 9, 2025
10. Microsoft Copilot: AI built for work (product page) (https://www.microsoft.com/en-us/microsoft-365/copilot), Microsoft, accessed Oct 23, 2026
11. Jensen Huang and Larry Fink at the World Economic Forum in Davos (report) (https://blogs.nvidia.com/blog/davos-wef-blackrock-ceo-larry-fink-jensen-huang/), Brian Caulfield, NVIDIA Blog, Jan 21, 2026
12. Put AI agents to work for marketing (home page) (https://www.jasper.ai/), Jasper, accessed Oct 23, 2026
13. Google users are less likely to click on links when an AI summary appears in the results (https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/), Pew Research Center, Jul 22, 2025
14. Article 4: AI literacy, AI Act (Regulation (EU) 2024/1689), consolidated text as amended by the Digital Omnibus on AI (https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-4), European Commission AI Act Service Desk, accessed Oct 26, 2026; dates of application in Article 113 (https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-113)

---

## Frequently asked questions

### What are the five layers of the AI cake?

Energy, chips, infrastructure, models and applications, from the bottom up. That is how NVIDIA's CEO, Jensen Huang, describes the AI economy. My marketer's stack folds his bottom three into hardware, keeps models and applications, and adds orchestration and interfaces.

### Which layer do marketers buy from?

Mostly the top layers: applications, orchestration, and interfaces. But every layer shows up on your bill, in your data policy or in front of your customers.

### Is AI getting cheaper?

Yes, fast. Stanford's AI Index found that the inference cost for a system performing at the level of GPT-3.5 dropped over 280-fold between November 2022 and October 2024. Be wary of locking today's rates into multi-year deals.

### What should you ask an AI vendor?

Which region processes your prompts and files, which country's law applies, which model they use and whether your inputs train it.

The five-layer cake is Jensen Huang's model; the marketer's stack is my adaptation. Product names are examples, not endorsements.

© 2026 Nicole Fung · nicolefung.pro