



## AI Unveiled (Nov 2025)

The pace of AI innovation continues to accelerate, unveiling new capabilities and challenges. Tech giants poured nearly **\$400 billion** into AI infrastructure this year <sup>1</sup>, driving unprecedented demand for compute, data centers and power. For example, major exchanges have already been disrupted by cooling failures in AI datacenters <sup>2</sup>. New investments and innovations – from optical processors to AI-specialized chips – are essential to support this growth. As one analyst warns, the biggest bottleneck now is **power and cooling**, not just chip availability <sup>1</sup> <sup>2</sup>.

*Illustration: The AI revolution requires massive hardware. "AI" letters on a motherboard underscore how compute-intensive modern models are. (Reuters)*

### Key Discoveries

- **Claude Opus 4.5 (Anthropic)** – Anthropic on Nov 24 unveiled Claude Opus 4.5, its latest flagship LLM, which achieves state-of-the-art performance on coding and reasoning benchmarks <sup>3</sup> <sup>4</sup>. Opus 4.5 “ranks among the most powerful models in the Claude family,” excelling at writing detailed code and even building agentic subroutines <sup>3</sup>. It is the first model to score above 80% on a respected software-engineering benchmark <sup>4</sup>. Anthropic also released companion tools (Chrome/Excel plugins) leveraging Opus for data analysis <sup>5</sup> and significantly improved its memory system for longer contexts <sup>5</sup>. In practice, Opus 4.5 can autonomously refine solutions and remember past insights to improve multi-step tasks <sup>6</sup>.
- **POMMM Optical Computing** – In hardware research, scientists reported a breakthrough *optical processor* architecture called **Parallel Optical Matrix-Matrix Multiplication (POMMM)** <sup>7</sup>. POMMM uses a single coherent light pulse to perform many tensor operations in parallel, addressing the traditional bottleneck of sequential optics. According to the report, this method “enables fully parallel tensor processing through a single coherent light propagation,” scaling performance with data dimensionality <sup>7</sup>. In effect, POMMM can execute multiple high-dimensional multiplications in one “shot,” offering orders-of-magnitude speedups. This paves the way for **light-speed AI computation** that could dramatically boost large model training and inference.
- **PopEVE for Genomic AI** – Researchers at Harvard Medical School unveiled *popEVE*, a new AI model for genomics. PopEVE scores genetic variants by their likelihood of causing serious disease, aiding rare-disease diagnosis. The team showed that popEVE can “predict which genetic mutations are lethal or severely damaging,” highlighting disease-causing variants that clinicians previously missed <sup>8</sup> <sup>9</sup>. By integrating protein-structure modeling with population genetics, popEVE flags harmful mutations quickly. In practice, it helped identify two new genes linked to rare diseases in a retrospective study, suggesting such AI tools could dramatically accelerate diagnosis and drug target discovery <sup>9</sup>.
- **BrainBody-LLM Robotics (NYU)** – A new algorithmic paradigm, **BrainBody-LLM**, was introduced for robotics: one LLM (“Brain”) plans high-level tasks, another (“Body”) refines execution. In this scheme,

the Brain LLM breaks down a command (“Eat chips on the sofa”) into subgoals, and the Body LLM chooses specific robot actions. In simulation and real robot tests, this two-step model improved task success by ~17% over prior methods <sup>10</sup> <sup>11</sup>. By using closed-loop feedback between the LLMs, the robot could adapt to errors in real time. This suggests LLMs can now not only generate instructions, but actively plan and correct embodied actions – a step toward more capable home and service robots <sup>10</sup> <sup>11</sup>.

## Emerging Technologies

- **AI Cloud Infrastructure** – The AI arms race is sparking huge cloud investments. Amazon Web Services announced a **\\$50 billion** program (adding ~1.3 GW of AI-optimized datacenter capacity) to serve U.S. government AI needs <sup>12</sup>. The project includes dedicated AWS Top Secret/GovCloud regions and will integrate advanced ML services (SageMaker, Bedrock, etc.). Reuters notes this move is part of “one of the largest cloud infrastructure commitments targeted at the public sector” <sup>12</sup>, underscoring how governments are rapidly expanding compute capacity to maintain technological leadership <sup>12</sup> <sup>13</sup>. Analysts warn the U.S. is in an AI “arms race” with China, driving ever-larger infrastructure builds <sup>13</sup>.
- **AI Memory and Chips** – Semiconductor advances are underway. For example, Micron announced plans to invest **¥1.5 trillion (~\\$9.6 billion)** to build a new high-bandwidth memory (HBM) chip plant in Japan <sup>14</sup>. These HBM chips are crucial for AI accelerators, and demand for them is “being driven by the growth of artificial intelligence and data centre investment” <sup>15</sup>. The new Japanese plant (due online in 2028) will help diversify supply away from Taiwan and meet surging AI demand <sup>15</sup>.
- **Data-Center Cooling & Power** – The AI boom is stressing data-center infrastructure, triggering a wave of hardware innovation. Traditional air cooling is often insufficient for modern servers, so companies are turning to liquid cooling and waste-heat reuse <sup>16</sup>. For example, news reports that **Eaton** is acquiring Boyd’s thermal-management business for \\$9.5 billion, aiming to expand liquid-cooling capacity in response to AI workloads <sup>17</sup>. At the same time, firms are building entirely new architectures (e.g. Microsoft’s water-free cooling data center) <sup>18</sup>. This trend – coupled with reported multibillion-dollar deals (Azure and AWS adding gigawatts of capacity) – shows that **AI hardware platforms** (from processors to cooling systems) are rapidly evolving to meet demand.
- **Specialized AI Frameworks** – Novel algorithms continue to emerge. Besides BrainBody-LLM, researchers showcased *Natural Language to 3D Control* and *captcha-breaking benchmarks* at NeurIPS, indicating a push towards LLMs that interact more directly with real-world systems. (For example, MBZUAI researchers proposed quantifying “AI security” by the bits leaked per query in adversarial tests <sup>19</sup>.) These innovations suggest next-generation AI will integrate deeper with specialized hardware and with real-time control loops.

## Industry Applications

- **E-Commerce & Retail** – AI is embedding into shopping. In late Nov, OpenAI and partners launched “**shopping research**” assistants that let ChatGPT search products and compare options for users <sup>20</sup> <sup>21</sup>. For instance, simply asking ChatGPT to find the “quietest cordless stick vacuum for a small apartment” yields a personalized buyer’s guide, as ChatGPT “asks clarifying questions, researches

deeply, and delivers a custom recommendation”<sup>22</sup> <sup>20</sup>. This ChatGPT shopping agent (rolling out on mobile/web) was confirmed by Reuters, which noted ChatGPT introduced “a personal shopping assistant... to pave the way for monetization through ads or commission-based sales”<sup>20</sup> <sup>21</sup>. Retailers and startups are already integrating such AI tools to enhance online shopping experiences.

- **Healthcare & Life Sciences** – AI is impacting diagnostics and research. As noted above, Harvard’s *popEVE* model for genomics marks a breakthrough in rare-disease diagnosis<sup>8</sup> <sup>9</sup>. Similarly, pharmaceutical companies are beginning pilot projects using generative models to design molecules and predict patient outcomes faster than before (ongoing news suggest collaborations between big pharma and AI labs). The trend is that AI base models (like new LLMs and diffusion models) are enabling new applications in medicine, from interpreting X-rays to simulating biochemical pathways, albeit still in early stages.
- **Other Sectors** – AI is also seeping into many fields. For example, some public reports (beyond last week’s window) describe AI-powered surveillance in agriculture (monitoring crops via drones) or finance (AI agents recommending portfolio allocations). Energy companies are piloting AI forecasting for grid management, and logistics firms use AI planning for warehouse automation. These experiments illustrate how the latest models (LLMs, vision models, etc.) are quickly being tested in real workflows, foreshadowing broader adoption next year.

## Challenges & Considerations

- **Infrastructure & Energy** – The infrastructure strain is a major concern. Reuters details how an AWS datacenter cooling failure briefly halted a major exchange’s trading<sup>2</sup>. Experts note that traditional cooling is inadequate for supercharged AI servers, prompting shifts to liquid cooling and heat reuse<sup>16</sup>. Power usage is rising sharply: an AFP report warns hyperscale data centers could consume as much as 7–12% of U.S. electricity by 2030<sup>23</sup>. In fact, analysts predict a potential **45 GW power shortfall by 2028** if AI growth continues unchecked<sup>24</sup>. These pressures underscore the need for new cooling technologies and greener designs.
- **Security & Reliability** – New vulnerabilities have emerged as AI becomes ubiquitous. Security researchers recently exposed critical remote-code-execution (RCE) flaws in popular AI inference engines (affecting Meta’s Llama, NVIDIA’s TensorRT-LLM, etc.)<sup>25</sup>. These bugs – stemming from unsafe model-loading code – could allow attackers to hijack AI servers or steal proprietary models. Ensuring the safety and robustness of AI systems (against both cyberattacks and simple misuse) remains an urgent challenge.
- **Ethical and Legal Issues** – As always, AI’s rapid advance raises policy questions. This week the EU signaled a more relaxed approach to the AI Act, simplifying regulations to avoid stifling innovation (a so-called “digital omnibus” proposal<sup>26</sup>). Meanwhile, courts and legislators continue to grapple with copyright, bias, and liability in AI outputs (for example, a German court recently ruled that AI-generated lyrics infringed copyright). These legal debates – alongside concerns about hallucinations, data privacy, and fairness – underscore that **governance must evolve** in tandem with the technology.

## Outlook

In the near term, the AI “arms race” shows no signs of slowing. Experts expect sustained **building of compute infrastructure** (more custom chips, memory, and data centers) and **deployment of ever-smarter models**. For instance, Reuters notes that AWS’s \$50B project will equip agencies with the latest AI services (Amazon’s Nova, Claude, etc.), emphasizing how governments are upping the ante <sup>27</sup>. At the same time, breakthrough research like optical processors (POMMM) could eventually ease the hardware bottleneck <sup>7</sup>.

Trends to watch include the shift to *multimodal* and *agentic* AI (integrating language, vision, robotics), and further specialization (tiny models for edge devices, huge models for science). Cooling and power will drive innovation in energy-efficient AI (e.g. liquid cooling, waste-heat reuse) <sup>16</sup>. The **broader business impact** is also looming: analysts predict hundreds of millions of paying AI users by 2030 <sup>20</sup>, and corporations are eyeing AI for productivity gains across sectors.

Overall, last week’s developments reveal a dual narrative: astounding technical progress (new model capabilities, optical compute) alongside growing infrastructure demands. In the coming months, we can expect more breakthroughs (especially in AI–hardware co-design) and a sharpening focus on sustainable, secure AI deployment. Stakeholders will need to balance innovation with safety – for example, by ensuring models are well-tested before widely releasing them. As one Reuters analyst puts it, the U.S. is in an “AI arms race” where staying competitive may require unprecedented compute expansion <sup>13</sup>, but success will also hinge on overcoming the **power and cooling** constraints that come with it <sup>1</sup> <sup>2</sup>.

**Sources:** Reports from OpenAI, Anthropic, AWS, Harvard Medical School, and news outlets (Reuters, TechCrunch, Nature Photonics, etc.) over the past week were used to compile these findings <sup>3</sup> <sup>7</sup> <sup>20</sup> <sup>8</sup> <sup>25</sup> <sup>12</sup> <sup>14</sup>. Each item above is confirmed by multiple credible sources.

---

- 1 23 24 **AI revolution running out of power**  
<https://www.thenews.com.pk/latest/1356937-ai-revolution-running-out-of-power>
- 2 16 17 18 **Keeping cool: heat a key challenge for data centers and AI | Reuters**  
<https://www.reuters.com/business/keeping-cool-heat-key-challenge-data-centers-ai-2025-11-28/>
- 3 6 **Anthropic bolsters AI model Claude's coding, agentic abilities with Opus 4.5 | Reuters**  
<https://www.reuters.com/business/retail-consumer/anthropic-bolsters-ai-model-claude-coding-agentic-abilities-with-opus-45-2025-11-24/>
- 4 5 **Anthropic releases Opus 4.5 with new Chrome and Excel integrations | TechCrunch**  
<https://techcrunch.com/2025/11/24/anthropic-releases-opus-4-5-with-new-chrome-and-excel-integrations/>
- 7 **Direct tensor processing with coherent light | Nature Photonics**  
[https://www.nature.com/articles/s41566-025-01799-7?error=cookies\\_not\\_supported&code=9fa4e54c-1c0e-4436-ad5d-17c4314886db](https://www.nature.com/articles/s41566-025-01799-7?error=cookies_not_supported&code=9fa4e54c-1c0e-4436-ad5d-17c4314886db)
- 8 **New Artificial Intelligence Model Could Speed Rare Disease Diagnosis | Harvard Medical School**  
<https://hms.harvard.edu/news/new-artificial-intelligence-model-could-speed-rare-disease-diagnosis>
- 9 **New Artificial Intelligence Model Could Speed Rare Disease Diagnosis - Global Genes**  
<https://globalgenes.org/raredaily/new-artificial-intelligence-model-could-speed-rare-disease-diagnosis/>
- 10 11 **BrainBody-LLM algorithm helps robots mimic human-like planning and movement**  
<https://techxplore.com/news/2025-11-brainbody-llm-algorithm-robots-mimic.html>
- 12 13 27 **Amazon pledges up to \$50 billion to expand AI, supercomputing for US government | Reuters**  
<https://www.reuters.com/business/retail-consumer/amazon-invest-up-50-billion-ai-supercomputing-us-government-customers-2025-11-24/>
- 14 15 **Micron to invest \$9.6 billion in Japan to build AI memory chip plant, Nikkei reports | Reuters**  
<https://www.reuters.com/world/asia-pacific/micron-invest-96-billion-japan-build-ai-memory-chip-plant-nikkei-reports-2025-11-29/>
- 19 **How many queries does it take to break an AI? We put a number on it. - MBZUAI**  
<https://mbzuai.ac.ae/news/how-many-queries-does-it-take-to-break-an-ai-we-put-a-number-on-it/>
- 20 **OpenAI projects 220 million paying ChatGPT users by 2030, The Information Reports | Reuters**  
<https://www.reuters.com/technology/openai-projected-least-220-million-people-will-pay-chatgpt-by-2030-information-2025-11-26/>
- 21 **OpenAI and Perplexity are launching AI shopping assistants, but competing startups aren't sweating it | TechCrunch**  
<https://techcrunch.com/2025/11/25/openai-and-perplexity-are-launching-ai-shopping-assistants-but-competing-startups-arent-sweating-it/>
- 22 **Introducing shopping research in ChatGPT | OpenAI**  
<https://openai.com/index/chatgpt-shopping-research/>
- 25 **Researchers Find Serious AI Bugs Exposing Meta, Nvidia, and Microsoft Inference Frameworks**  
<https://thehackernews.com/2025/11/researchers-find-serious-ai-bugs.html>
- 26 **Digital Omnibus on AI Regulation Proposal | Shaping Europe's digital future**  
<https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal>