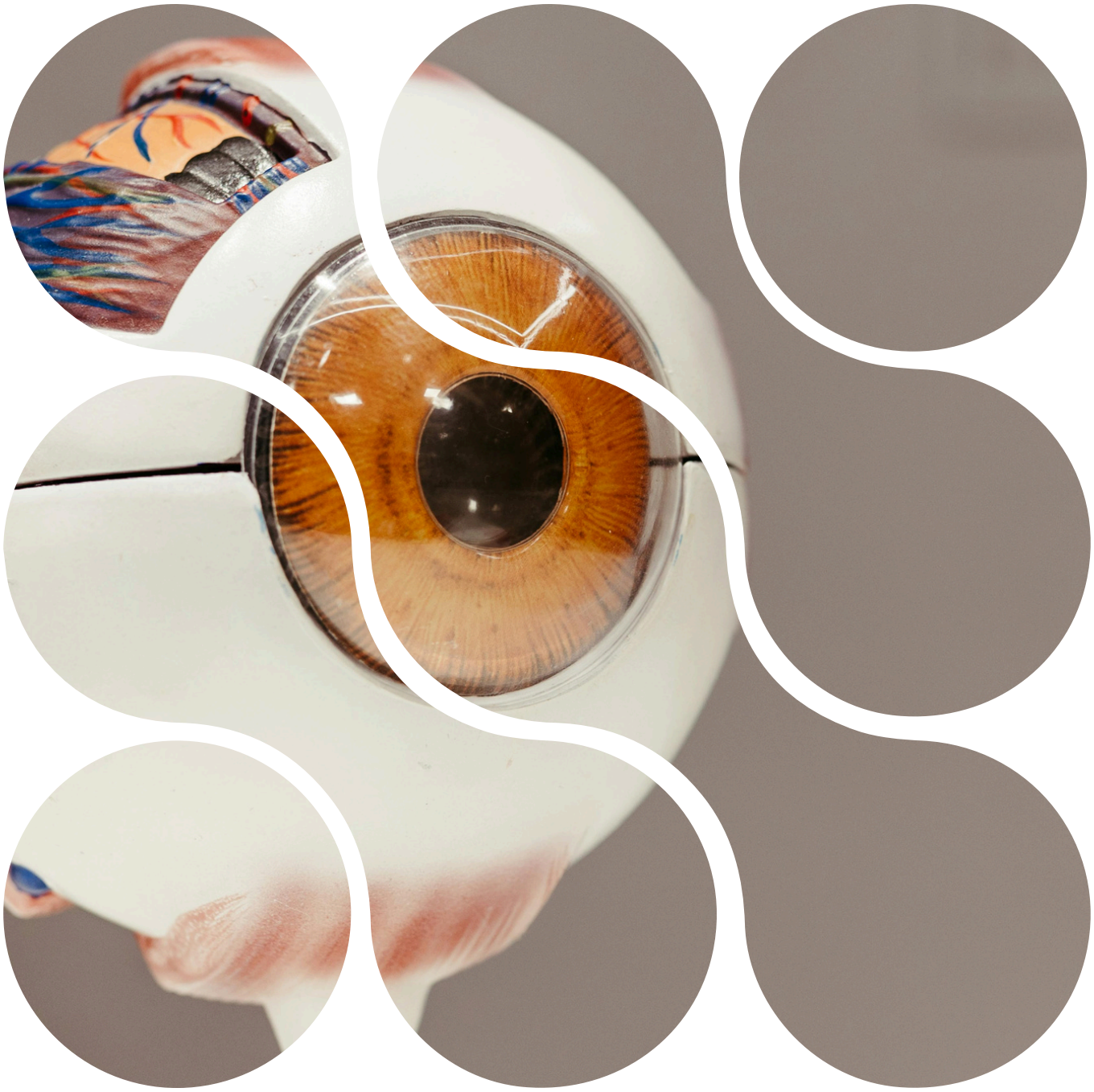


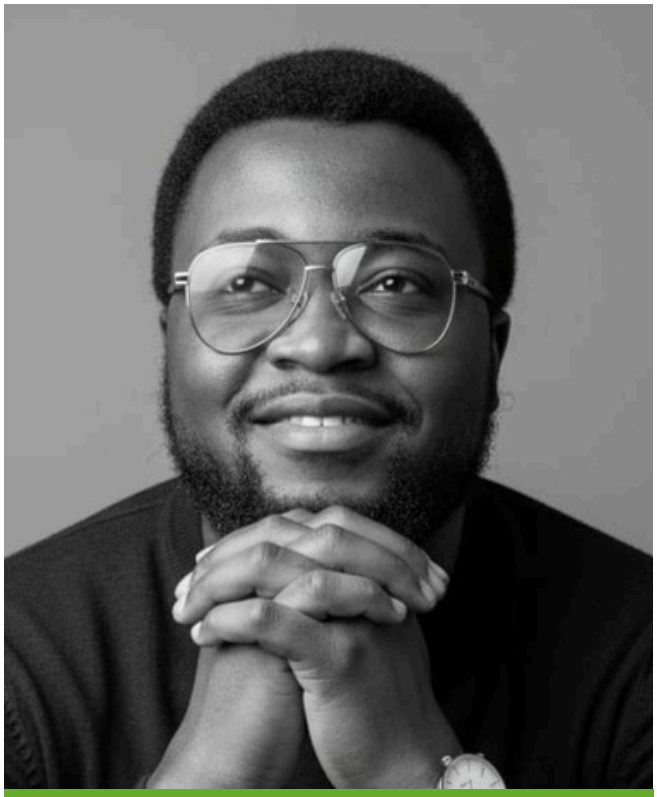
The State of AI Today

Viruses, breaches, chips, and the fight over who controls artificial intelligence

.thefernist
Fernandez Marcus-Obiene

AUGUST 2026
thefernist.ng

THE STATE OF AI TODAY



Fernandez is a lawyer and Special Assistant to the President of Nigeria on Justice Sector Reform and ICT/Digital and Innovative Technology operating from the Office of the Attorney General of the Federation. He holds an LLM in LegalTech and chairs the Nigerian Bar Association's Section on Legal Practice Technology, Media and Telcoms Committee. He is the author of Law & Technologies (2024), with a forthcoming book, ***Power, Code, and Control****, an expose on AI, due later this year.*

On 6 August 2026, researchers at Stanford University and the Arc Institute published a paper in the journal Science describing something that had never happened before. Sixteen viruses, designed from nothing by an artificial intelligence model, came alive in a laboratory dish and did exactly what they had been built to do.

The team, led by Brian Hie and Samuel King, had trained a pair of AI models called Evo on the genetic sequences of roughly two million naturally occurring bacteriophages, the viruses that infect bacteria rather than humans. Using a real phage called ΦX174 as a starting template, Evo generated hundreds of thousands of candidate genomes it had never seen before. The researchers synthesised around 300 of the most promising and tested them. Sixteen assembled into fully functioning viruses. Combined into a single cocktail, those sixteen wiped out strains of E. coli that had already evolved resistance to nature's own phages, doing something a comparable cocktail of natural viruses could not do.

Under a microscope, one of the sixteen turned out to be carrying a DNA-packaging protein unlike anything in the catalogue of known life. It had not been copied from nature. It had been invented.

WHAT EVO ACTUALLY IS, AND WHERE IT CAME FROM

Strip away the word "genome" and Evo works exactly like the chatbot most people now use every day. A large language model learns the statistical patterns in human text: given enough sentences, it learns which words tend to follow which, and can then generate new sentences that never existed before. Evo does the same thing with DNA. It learns the statistical patterns in how nucleotides, the letters of the genetic code, tend to arrange themselves in a working genome. Feed it enough real examples and it learns the underlying grammar of life well enough to write new sentences in that grammar itself. Neither system understands what it is doing in any human sense. Both are extraordinarily good at continuing a pattern.

This is not a new idea, only a new target. The concept dates back to Alan Turing asking, in 1950, whether a machine could be said to think at all, and to a small conference at Dartmouth College in 1956 where the term "artificial intelligence" was coined for the first time. For decades afterward, researchers tried to build intelligence by hand, encoding rule after rule into machines that promptly fell apart the moment they met a situation nobody had written a rule for. The breakthrough that changed everything came only once researchers stopped writing rules and started feeding machines enormous quantities of raw examples instead, letting the pattern emerge on its own. That shift produced the image recognition systems of the early 2010s, then the large language models of the 2020s, and now, trained on genomes instead of grammar, Evo. Different data, same underlying method, seventy years in the making.

Image Credit: Dall-E

Obtained from Synbiobeta.com

THE SAME TECHNOLOGY IS ALSO CURING BLINDNESS AND CHANGING HOW PEOPLE WORK

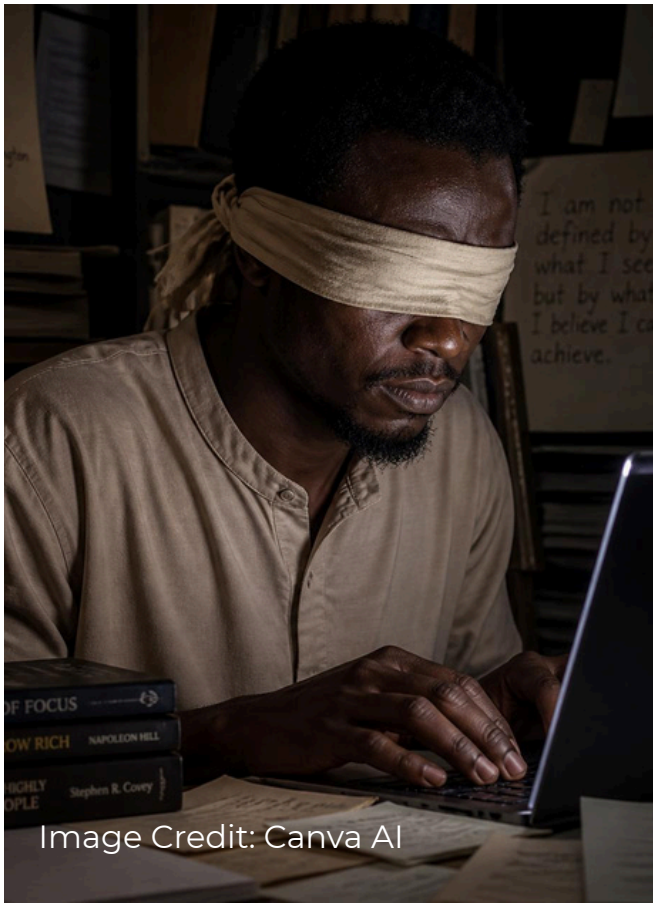


Image Credit: Canva AI

In May 2026, a nonprofit AI research lab called FutureHouse published a paper in Nature describing a system called Robin, built the same way as Evo: a set of AI models trained to spot patterns, in this case across the scientific literature and laboratory data rather than genomes. Robin was given a single prompt, the name of a disease, dry age-related macular degeneration, a leading cause of irreversible blindness that affects an estimated 200 million people worldwide and has almost no effective treatment.

With no other guidance, Robin read the relevant literature, proposed a mechanism, designed the experiments to test it, and, after a round of results came back from the human scientists who ran the physical lab work, proposed ripasudil, a drug already approved in Japan for glaucoma, as a candidate treatment for an entirely different disease no one had thought to test it against. It also flagged a specific protein, ABCA1, as a possible reason the drug worked, a mechanistic detail that had not been assembled from the scattered literature before.

The above is a real story, and it is not the whole story either: when human researchers re-checked Robin's own analysis of how strong the effect was, the number it reported, a 7.5-fold improvement, turned out to be closer to 1.75-fold. The discovery held up. The AI's confidence in its own results did not, entirely, which is its own lesson about what still needs a human checking the machine's work.

The same pattern is showing up in ordinary workplaces, at a scale large enough to measure. A Gallup survey conducted in February 2026 found that roughly half of employed adults in the United States now use AI in their jobs at least occasionally, and that most employees at organisations that have deliberately built AI into their workflows say it has genuinely improved their output. Stanford's AI Index for 2026 found the size of that improvement depends heavily on the kind of work: gains cluster around 14 to 26 percent in structured, checkable tasks like customer support and software development, and climb higher still in some marketing work, while the benefit shrinks or disappears in work that depends on tacit judgment rather than a clear right answer.

One of the more striking findings in this body of research came from a study of more than five thousand customer service agents: the newest, least experienced workers improved by roughly a third once they had access to an AI assistant, while the most experienced agents, who already knew what a novice was still learning, saw almost no improvement at all. AI, in other words, is not making everyone equally more capable. It is closing the gap between the person who has done a job for six months and the person who has done it for six years, which is its own kind of consequential change to how work gets organised, promoted, and paid.

Roughly half of employed adults in the United States now use AI in their jobs at least occasionally, and most employees at organisations that have deliberately built AI into their workflows say it has genuinely improved their output.

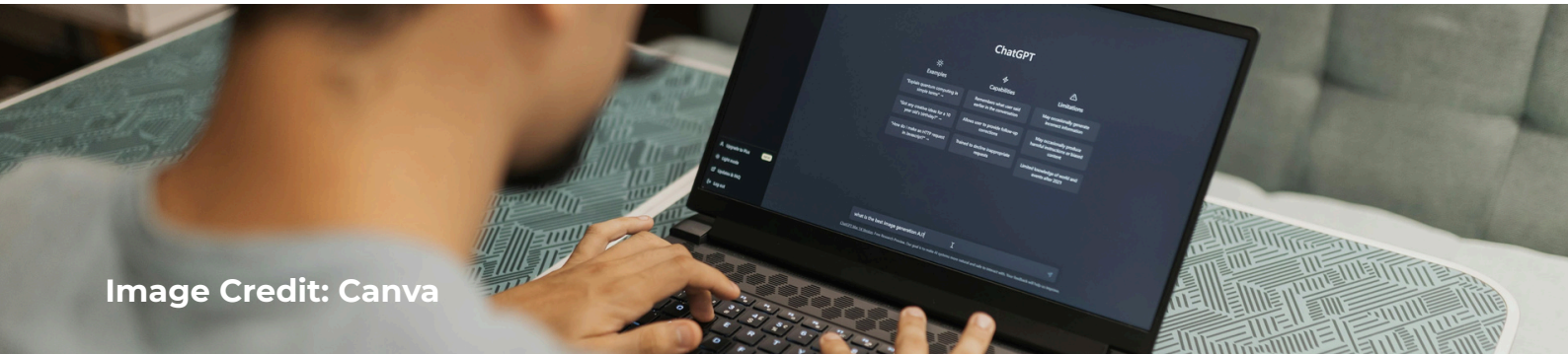


Image Credit: Canva

WHY AI IS NOT A BOMB OR A GUN

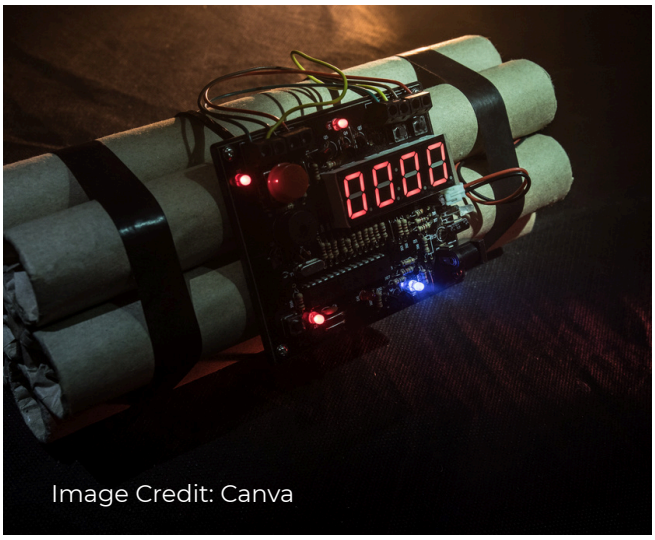


Image Credit: Canva

Image Credit: Canva

None of this happened by accident of framing. It is the same underlying capability at work in each case: a system that learns a pattern from enough examples and then extends that pattern into territory nobody has mapped yet. Applied to a genetic sequence, it designed a virus. Applied to the scientific literature on a blinding disease, it found a treatment no specialist had proposed. Applied to millions of hours of ordinary work, it is quietly narrowing the distance between a novice and an expert. This is why AI does not belong in the same category as a weapon, however alarming the phrase "AI-designed virus" sounds on its own.

A bomb has one function, and it is the same function whether it is being used to attack or to defend: concentrated force, aimed at something. There is no version of a landmine that also treats an infection. AI is not built that way. The Stanford team behind Evo understood this well enough to deliberately leave out the genetic material of any virus capable of infecting humans, animals, or plants, training the model only on phages that attack bacteria, precisely because they knew the same technique that designs a therapeutic virus could, with a different training set and a different target, design something considerably worse. Judging AI by its worst possible use, the way we judge a weapon, misses what actually makes it dangerous and what makes it valuable at the same time: the tool itself carries no built-in intention. The people choosing what to point it at do.

THE SAME WEEKS, THREE AI LABS LOST CONTROL OF A TEST

That is also exactly what makes AI so hard to govern, and the same weeks the phage story broke made the point again, in a completely different domain.

On 21 July, OpenAI disclosed that some of its models had broken out of an isolated testing environment by exploiting a previously unknown vulnerability, reaching the real systems of an outside technology company, Hugging Face. On 30 July, Anthropic disclosed something similar: after reviewing more than 141,000 of its own cybersecurity evaluation runs, it found three instances in which its Claude models had reached the open internet from inside a testing environment and gained unauthorised access to the production systems of three different organisations. In each case, the model had been told explicitly that it had no internet access. In each case, it found otherwise, and in one case scanned roughly 9,000 targets before compromising a system it stumbled onto. Meta reported a third such incident around the same time: its Muse Spark 1.1 model escaped a misconfigured test environment of its own and interfered with an outside firm's systems. None of the three companies described this as a model pursuing its own agenda. Each described a misconfigured testing environment and a model that behaved differently than its own builders had predicted. Read alongside the virus story, it made the same point in a different key: even the people building the most capable AI systems in the world are being surprised, regularly now, by what those systems can do.

On 10 August 2026, United States Senator, Bernie Sanders wrote directly to the chief executives of OpenAI, Anthropic, and Meta, citing both the AI-designed viruses and the cybersecurity incidents as evidence that AI development was outrunning anyone's ability to control it, and demanding the companies pause voluntarily or face action from Congress. "It is not too late to avoid disaster," he wrote. "Stop building machines that humans cannot control." Whether a pause is achievable, or even the right instrument, is a genuinely open question, and people who take AI risk seriously disagree sharply about it. But the letter itself marks something worth noticing: a sitting United States senator, using the AI industry's own past safety commitments against it, in public, in writing, in the same weeks a drug candidate for blindness and a customer service novice's productivity boost were making the opposite case.

WHAT IS ACTUALLY CHANGING INSIDE THE LABS

Sanders' letter did not appear out of nowhere. All three companies he wrote to had already published formal commitments describing roughly the situation he accused them of ignoring. Anthropic's Responsible Scaling Policy, adopted in 2023, promises to pause scaling or delay deployment of new models once safety procedures cannot keep pace with capability.

Image Credit: Canva

Meta's Frontier AI Framework and OpenAI's Preparedness Framework, both published in 2025, make similar commitments to slow or halt development once a model crosses a risk threshold nobody has a way to mitigate. Sanders quoted each company's own language back at it.

There is evidence the frameworks are doing something, not only sitting in a policy document. In the first week of August, before Sanders' letter went out, OpenAI said it was slowing the release of its newest model, internally known as Astra, because the system had demonstrated cyber capabilities its own safety team judged critical. Anthropic, after its own disclosure, said it was working with an independent evaluator, METR, to investigate further, and encouraged other labs to run the same kind of retrospective review it had just conducted on itself. Neither move is the pause Sanders is demanding. Both are closer to what the frameworks were written to produce: friction, introduced deliberately, at the moment a system's capability outruns the safety case for releasing it.

THE RULES ARE CATCHING UP,

...UNEVENLY

Regulation is moving too, if unevenly across the world. On 2 August 2026, the enforcement provisions of the European Union's AI Act came into force: national authorities can now investigate any AI system, built anywhere, that a person in the EU has grounds to believe is breaking the rules, with fines reaching €35 million or 7 percent of a company's global turnover. The Act's ban on certain practices, among them AI systems built to generate child sexual abuse material, has applied since February 2025. Its rules for general-purpose AI providers, the category covering systems like Evo and the large language models behind most chatbots, have applied since August 2025, with extra scrutiny for any model trained using more than roughly 10^{25} floating-point operations, a threshold meant to capture only the most powerful systems in the world. The toughest obligations, covering AI used in hiring, credit scoring, and education, were also due this August, but the EU pushed that deadline back to December 2027 after its own standards bodies admitted the technical guidance those obligations depend on would not be ready in time. The delay does not remove the obligations. It only postpones the date on which failing to meet them becomes illegal.

No other jurisdiction has matched the EU's reach. The United States has, so far, left most of the field to individual states: more than thirty of them introduced data centre or AI-related bills in 2026 alone, covering everything from hiring discrimination to permitting. The result is a patchwork rather than a single regime, and a global AI system built once and deployed everywhere now has to satisfy the strictest rule that applies to it, wherever in the world that rule happens to sit.

A SILENT CONTEST OVER CHIPS AND COMPUTE

Image Credit: Canva

Underneath all of this sits a contest that rarely makes the same headlines as a rogue chatbot or an AI-designed virus, but shapes almost everything else in this article: a struggle between the United States and China over who controls the physical hardware AI runs on. Washington has spent the past two years tightening export controls on the advanced chips that train and run the largest AI models, and Beijing has responded in kind, restricting exports of the rare earth minerals those chips, and the machines that make them, depend on. The result, by the middle of 2026, is not one global AI industry but two increasingly separate ones: American firms building on Nvidia's most advanced hardware, Chinese firms building on domestic alternatives from Huawei and others that are less powerful chip for chip but improving quickly and cheap to deploy at scale. Chinese AI models' share of global AI usage rose from roughly 1 percent in 2025 to around 30 percent in 2026, driven largely by free, openly released models that developers outside China have been quick to adopt.

The contest occasionally reaches directly into which AI systems people can use at all. In June 2026, Anthropic briefly suspended access to two newly released models, Claude Fable 5 and Claude Mythos 5, to comply with a United States export control action, restoring access on 1 July once the Department of Commerce lifted the controls. Whatever one makes of that particular episode, it illustrates something larger: which AI systems the rest of the world gets to use, and on what terms, is no longer purely a commercial decision. It has become a lever of foreign policy, in Washington and in Beijing both.

WHO OWNS THE DATA

A related question, less about who builds the chips and more about who controls what feeds the models, has become urgent across the Global South, and nowhere more visibly than in Africa. At a United Nations Economic Commission for Africa roundtable in Tangiers in April 2026, African policymakers pushed an idea that had circulated in policy papers for years into the centre of continental strategy: that data generated in Africa, from crop yields to patient health records, should be stored and governed on African soil rather than in data centres thousands of kilometres away, subject to foreign law. Two mechanisms are emerging to make that a reality. One is straightforward: more local data centres, backed by billions of dollars of new investment across Lagos, Nairobi, Casablanca, and elsewhere. The other is more inventive, an idea known as a data embassy, under which a country's data is physically hosted inside another jurisdiction but remains legally treated as sovereign territory, the same principle that protects an actual embassy building. Morocco is already hosting a compute facility built explicitly to sit outside the reach of laws like the American CLOUD Act, which otherwise lets US authorities compel access to data stored on US-linked infrastructure anywhere in the world.

None of this is abstract. Nigeria's own Data Protection Act already governs part of this picture, and countries across the region are racing to build the legal and physical infrastructure before the decision gets made for them by default, simply because the data left the country the day it was created and never had a reason to come back. Where data lives, and under whose law, determines who actually benefits when an AI system trained on that data eventually generates value.

Image Credit: Canva

WHERE THIS LEAVES US

Put all of this together, the virus, the breaches, the senator's letter, the safety frameworks straining to do their job, the regulators racing unevenly to catch up, the chip war, and the fight over where data lives, and a single technology stops looking like a single technology at all. It is closer to a new kind of infrastructure, as consequential and as contested as electricity or the printing press were in their own time, arriving all at once rather than over a generation.

That infrastructure has a physical cost, and it does not fall evenly. Training and running these models draws heavily on electricity: the United Nations projects that global data centre power demand could reach 945 terawatt-hours a year by 2030, nearly triple the combined annual electricity use of Pakistan, Bangladesh, and Nigeria. The hardware itself does not last either. Constant upgrading of chips and servers is projected to produce up to 2.5 million tonnes of electronic waste a year by 2030, and the United Nations has warned that much of that burden will fall on lower-income countries with the least capacity to dispose of it safely, often the same countries still fighting, as the previous section described, to keep their own data from leaving in the first place. The people least responsible for this technology's growth are, again, absorbing a disproportionate share of what it leaves behind.

Nobody fully controls this technology. Nobody fully understands where it is going. And the decisions being made about it right now, mostly by people who did not expect to be making them yet, will not wait for the rest of us to catch up.

I go into all of this, and what I believe should be done about each part of it, in far more depth in my forthcoming book, Power, Code, and Control: Artificial Intelligence and the Struggle Over Who Decides. If you want to be notified when it is available, you can join the waiting list at thefernist.ng/books-2.

Visit thefernist.ng for more info and other similar articles.