AI Was Never New. It Just Started Talking to Us Directly.

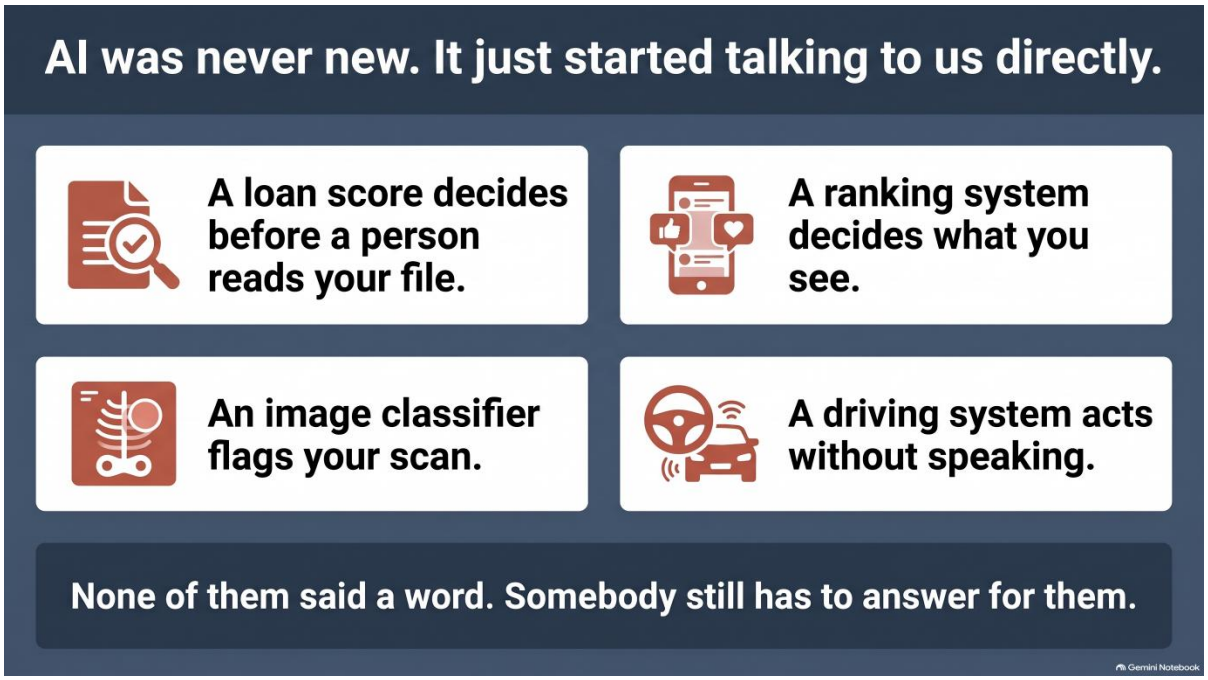
It has been scoring loans, ranking feeds, and reading scans for decades. The chatbot is just the first version that speaks, and nobody settled who answers for the rest.

A woman applies for a car loan on a Tuesday morning, and before a human being reads her name, a model has scored her. On her phone, a ranking system has already chosen which six of the four thousand things published overnight she will actually see. At the clinic that afternoon, a classifier flags a region of her scan before the radiologist opens the file. On the drive home, the car keeps itself in the lane.

She has interacted with artificial intelligence perhaps a dozen times before lunch. Not one of those systems said a word to her.

If you asked her whether she uses AI, she would probably say she tried a chatbot once and found it useful for writing an email.

The gap between where artificial intelligence actually operates and where people believe it lives is the whole reason I do this work. So here is the story of what AI is, told in the order it happened. It is told from where the consequences land rather than from where the product launches happen. Her Tuesday is the map. We are going to walk it backward through fifty years and arrive where we started.



First, we wrote the rules down

The dominant early approach was to tell the machine exactly what to think.

Symbolic AI and expert systems worked from explicit rules and encoded knowledge, written in the shape of a conditional. Some of them still run today, wherever the rules are genuinely known and consistency matters more than nuance. If this condition exists, examine these possibilities, and if these criteria are met, reach this conclusion. Specialists sat with engineers and spelled out how a diagnosis was reached or how a loan was approved. Other approaches ran alongside it, in search, planning, and automated reasoning, and all of them shared one property.

A human wrote every rule.

Nobody had to ask who held authority, because the logic was legible and somebody had typed it. If the system reached a conclusion you disagreed with, you could open it, find the line, and argue with the person who wrote it.

We lost that property next, and we have not recovered it.

Then we let it learn from examples

The next move was to stop writing rules and let the machine find them.

Machine learning shows a system a great many examples and lets it infer the statistical relationships inside them. Transactions labeled fraudulent or legitimate, customers who stayed and customers who left, applicants who repaid and applicants who defaulted. The system learns which patterns tend to accompany which outcomes, then applies that to someone it has never seen.

This is the technology that scored her loan.

No engineer wrote the rule moving her application into the queue it landed in. The rule emerged from records of people who came before her, so whatever was true about the past is now quietly true about her Tuesday.

Sometimes the examples come labeled. Sometimes they arrive with no labels and the system hunts for clusters and anomalies on its own. Sometimes the machine builds its own training signal out of the raw material. That is what makes it possible to learn from enormous piles of text and images without a person tagging every one. And sometimes the system learns by acting and collecting a reward, adjusting toward whatever gets rewarded.

That last one carries the first real warning in this story. A system trained on reward optimizes for the reward rather than for the intention behind it. If the number you chose is a

poor stand-in for what you care about, the machine pursues that number with great efficiency. It takes you somewhere you never intended to go.

I have watched that happen in rooms with no machines in them at all. Pick the wrong metric and a whole organization will chase it off a cliff while reporting excellent results the entire way down. The machine simply does it faster and without anyone noticing the direction.

Then it learned to rank, and it decided what she would see

Somewhere in the same period, these systems took over the question of what people see.

Recommendation and ranking decide which posts, products, videos, advertisements, search results, and people appear in front of a person, and in what order. That is the system that chose her six items out of four thousand. It optimized for engagement, or relevance, or revenue, or watch time, depending on what its owners set.

I spent years working inside those systems before anyone in my field called them artificial intelligence. They were algorithms, and algorithms were plumbing.

They were never plumbing. A system does not need to write a sentence to change what a person believes. It only needs to decide which sentence appears first.

Then we gave it depth, and it learned to see and hear

Stack the learning in many layers and the machine starts finding patterns nobody could have described in advance.

Deep learning did not invent computer vision or speech recognition, since both existed before it. What deep learning did was transform their performance and their scale. A system could now recognize a face, transcribe a conversation, and separate suspicious tissue from healthy tissue in an image.

This is the technology that flagged her scan.

It is also where the field stopped being able to fully explain itself. You can inspect a neural network's weights, and that will not tell you why it flagged this region and not the one beside it.

Here is the part worth keeping. The same technical capability inspecting a part on a production line can identify a person walking down a street. The engineering barely changes. The consequence changes completely, and the machine cannot tell the difference, because the difference is not a technical property of anything.

Then came the architecture that changed the shape of everything

In 2017, researchers introduced the transformer.

It learns relationships across a sequence, including relationships between things far apart in it, which in language meant connecting a word to another word many sentences away. The result was not simply better translation. Models stopped being built for one narrow job.

You could now train a very large model on very large amounts of material and adapt it to many different tasks afterward. We call those foundation models, and they changed the economics of the entire field. Instead of training something new for every problem, an organization starts from a broadly capable base and shapes it.

The efficiency casts a shadow, and the shadow is a governance problem.

When thousands of applications stand on the same foundation, a flaw in the foundation does not stay in one place. It propagates into every product built on top of it, including products built by people who never examined it and had no way to.

Then it learned to talk, and everyone finally noticed

Large Language Models are trained on enormous quantities of language, and at the core of that training is a deceptively simple task: predict what comes next from context. At sufficient scale, the task produces a system able to write, summarize, translate, analyze a document, draft software, and hold a conversation.

This is the moment the public met artificial intelligence, and the meeting did something strange. It made the field visible and invisible at once. A chat window is legible in a way a credit model never was. So the chat window became the whole category in the public mind, and the credit model went back to being furniture.

The same moment did a second thing, and it is why I never let fluency stand in for reliability anywhere in my own work. The model produces statistically plausible language, and plausibility is not truth. An answer can arrive polished and confident while carrying a fabricated source, a statistic nobody ever published, or a qualification quietly dropped from the original. It does not sound uncertain when it is uncertain, because sounding certain is part of what it learned.

When I started working seriously with these systems, the first thing that broke was sourcing. I asked for evidence and got an answer reading beautifully with nothing behind it. Everything I have built since answers some version of that failure. The first thing I built was a second system whose only job was checking the first one's sources.

The part almost everyone gets wrong

Here is the distinction I would most want carried out of this story.

A model is not a system.

The model is the component performing the inference. The system is everything around it. The data it draws from, the retrieval feeding it, the tools it can call, and the software it can run. The interfaces it touches, the monitoring, the human review, and sometimes the physical machinery at the end.

The language model may decide what step comes next. The system is what searches the web, queries the database, writes the file, sends the message, moves the money, or turns the wheel.

Consequence does not occur at the model. Consequence occurs where the system touches the world.

This is why evaluating a model exhaustively still leaves you with no idea what it has been wired to.

Then it stopped answering and started acting

For most of this story the machine returned something and stopped.

An agent does not stop. It takes an objective, breaks it into steps, searches, calls software, and compares results. It makes intermediate decisions, revises its plan, talks to other systems, and continues until it judges the objective complete. Several agents can coordinate on the same work.

The question that governs a chatbot is whether you can trust the answer. The question that governs an agent is what authority you delegated to it. It is a much larger question, and a much harder one to answer honestly after the fact. Stanford's AI Index for 2026 documented where this stands. Agent success on real computer tasks rose from roughly 12 percent to 66.3 percent in a year, landing within six percentage points of human performance. The same figure means agents still fail about one attempt in three.

A system right two times out of three, acting without asking, is not a system anyone should look away from.

Then it got a body

The last turn in the story so far: intelligence left the screen.

Models are now wired into vehicles, robots, drones, and industrial machinery. This is what held her car in the lane on the drive home, and it is the least conversational AI she touched all day.

The same 2026 AI Index documented the other half of the picture. Robotic manipulation succeeds at 89.4 percent in simulation, while robots complete about 12 percent of real household tasks. The laboratory and the world remain very far apart.

The gap is not a reason to relax. It is the reason to pay attention now, while the deployment decisions are being made. A language model in error produces a wrong paragraph, and a paragraph can be edited. An embodied system in error produces a wrong physical event, and nothing edits a physical event.

Capability is also concentrating. The compute, the models, the distribution, and the capital increasingly sit with a small number of owners. Their decisions about allocation and access arrive as announcements rather than as arguments anyone contests beforehand.

The character missing from the whole story

Put the sequence together.

A field that started with rules a person wrote, then moved to patterns nobody wrote and gained perception nobody can fully explain. It standardized onto foundations everything else is built on, then learned to speak persuasively whether or not it is right. It learned to act without returning for permission, acquired a body, and concentrated into a small number of hands.

At no point in that sequence did anyone answer the question of who decides.

Three words get used as if they were one word, and they are not. Ethical AI asks whether this should be done, Responsible AI asks who answers when this fails, and AI Governance asks who decides, by what authority, at what checkpoint.

Responsible AI is where most of the industry actually lives, and it is real engineering. Testing, monitoring, bias evaluation, documented limits, security, controls, and named organizational owners. I am not interested in diminishing it, because I depend on it.

The argument here is narrower than the usual criticism, and it is the one I have not been able to get past. Having oversight processes, accountability mechanisms, and responsible-AI controls does not establish binding human decision authority over the individual consequential output. A program can be excellent and still have no named person who holds

the authority to stop this decision, about this woman, on this Tuesday. The checking inside it is done by one machine validating another, by agents running the pipeline, or by a human in the loop. None of those three reliably produces that person.

Human in the loop is the phrase that hides the gap best. Someone can watch a dashboard, receive an alert, review a recommendation, click approve, or sit on a committee, and none of it establishes the person governs anything. Presence is participation, and authority is governance. Organizations buy the first routinely and report the second.

So I wrote the third one down, and Checkpoint-Based Governance is the framework built around it. AI Governance exists when a qualified human holds binding authority at specific checkpoints, with personal accountability for the outputs that pass through. The accountability has to reach a named person through moral, professional, civil, or criminal channels. If it cannot reach anyone, what exists is process.

I wrote it because no standards body has published a formal, standalone definition of the term as an official position. Plenty of institutions describe governance, assign governance functions, and publish governance principles. ISO, NIST, the OECD, UNESCO, and the EU AI Act all do some version of it. None of them locks a definition making binding individual human authority at the consequential checkpoint the threshold. A word carrying so much institutional weight with so much definitional room is a word an organization can attach to almost anything.

Back to Tuesday morning

None of this means the machine should be pulled out of the loan, the feed, the scan, or the car. The capability is real and it reaches further than any person reaches alone. Some AI appropriately stays at Responsible AI permanently, because a ranking system cannot have a named human reviewing every output and should not pretend otherwise.

It means a checkpoint would have to sit somewhere, proportional to what the system can do to a person. Sometimes it sits upstream, where a named human approves the policy, the boundaries, and the acceptable level of automation. Sometimes it has to sit immediately before the action. The location moves, and the accountability should not vanish while it moves.

Increasing capability does not create authority. A system that becomes faster, more accurate, more autonomous, or more fluent has not answered the question of who should be permitted to decide.

So go back to her. Suppose the model was wrong about her, the ranking system buried something she needed, and the classifier missed what it was looking at.

Somebody should be able to say who had the authority to let that happen, and who answers for the result.

If nobody can say, then nobody governed it. That is the story, and we are still in the middle of it.

The full reference version, with the types of AI, the technical layers, and the governance argument in detail: Artificial Intelligence: Technology, Capability, Governance, and Human Accountability

Related: Checkpoint-Based Governance: Presence Is Not Authority · AI Governance Has No Formal Definition. Here Is One. · Why You Cannot Program or Prompt Governance Into AI

Benchmark figures cited are from the Stanford Institute for Human-Centered Artificial Intelligence, AI Index Report 2026, Technical Performance.

Basil C. Puglisi, MPA
A Human-AI Collaboration

#AIassisted using HAIA Ecosystem