

WTF AI: What the hell is happening here?

Marcel Alexander Mayr @marcelamayr - 2025

Chapter 1: The Spark in the Silicon – From Early Dreams to Global Power

Something has changed in the last 2 years. Maybe you feel it sometimes, too? This crackling in the air, as if you were standing with one foot in your familiar world—the smell of rain on asphalt, the laughter of a friend—and with the other already in something new, invisible, that somehow feels faster, more intangible, and alien—an ever-more important world of data, devices, and algorithms that seem to permeate everything. The age of Artificial Intelligence has begun. It has outgrown my bookshelf, no longer a place visited only by science fiction authors. It exists in the here and now, it's almost the air we breathe, and beyond that, a mirror in which we increasingly encounter ourselves—whether we want to or not.

To understand what is quietly shaping your life, your decisions, and even your dreams today, I want to take a little journey through time with you, back to the roots of this digital earthquake. Far back. To a time when computers were still room-sized behemoths and the idea of a thinking machine sounded like pure fantasy. The year is 1950-something: A psychologist at Cornell University, Frank Rosenblatt, is tinkering with a strange apparatus. He calls it the **Perceptron**. A clunky box full of wires, potentiometers, and small motors, fed with punch cards. It was the first serious attempt to build a machine that could learn—rudimentary, yes, but still: learn. Rosenblatt dreamed of machines that could recognize faces, understand languages, perhaps even develop their own consciousness. What tension and excitement that must have been—this first spark of artificial cognition in a world that had just entered the atomic age. It was a moment of almost naive optimism, a reach for the stars.

But disillusionment followed swiftly. The Perceptron was extremely limited. It failed at tasks that would have been easy even for a toddler. Critics like Marvin Minsky and Seymour Papert published an influential book in 1969, "Perceptrons," in which they demonstrated the fundamental mathematical limits of these simple networks. What followed was the first so-called "AI winter"—a phase of stagnation in which the dream of the thinking machine seemed to be over.

But beneath the surface, the embers remained alive. A few researchers continued to work in secret. They searched for ways to train more complex, multi-layered neural networks that came closer to the complexity of the real world. This is where Geoffrey Hinton enters the stage, along with pioneers like Yann LeCun and Yoshua Bengio. Hinton and his team refined a method called backpropagation.

Maybe you remember learning something new as a child—riding a bike, or playing an instrument. Every mistake, every fall, every wrong note was a piece of information you could learn from. You changed something, adjusted your behavior, and then tried again.

Backpropagation is something like this internal learning plan for the machine. A method to learn from every 'misstep' and do it a tiny bit better the next time—only repeated millions of times, with the patience that we humans sometimes lack. It's an elegant feedback loop that allows even deep networks to recognize complex patterns from vast amounts of data.

This insight was the theoretical breakthrough. Finally, there was a method to train more complex artificial brains. But there were still two enormous hurdles: computing power and data. The blueprint for a space station had been invented, but only wood and nails were available.

Then, in the first decade of the new millennium, something remarkable happened. Two developments came together and created a perfect storm—almost overnight, as if someone had ignited a turbocharger:

The first turn was the explosion of computing power through **graphics processing units (GPUs). These chips, originally developed for video games, proved to be the perfect engines for neural networks.**

The second development was supplied by all of us: **The internet became an inexhaustible ocean of data. Every Google search, every uploaded photo, every tweet became potential training material. Projects like ImageNet, a huge database with millions of labeled images, provided the high-quality fuel.**

Around 2012, the rocket became visible to everyone in the sky. At the annual ImageNet competition, a team led by Geoffrey Hinton achieved a spectacular breakthrough with a neural network called AlexNet. It was the proof: Deep Learning, learning with deep networks, worked—and it worked dramatically better than anything before it.

That was the Big Bang of the modern AI revolution. Suddenly, the major tech corporations invested massively in research. AI labs shot up like mushrooms. Progress accelerated at a breakneck pace. **Machines learned to understand language, recognize images, and play games at a superhuman level. AlphaGo, developed by DeepMind, defeated the world's best human player in the highly complex strategy game of Go—a moment that stunned many experts. It learned by playing millions of games against itself, developing moves that no human had ever conceived.**

In parallel, the revolutionary so-called Transformer models, like those used by ChatGPT, were developed for language processing. The rise of these models marked the next phase: AI learned not only to recognize but also to generate—text, images, music, code. With systems like ChatGPT, Google's Gemini, or MidJourney, AI finally became a mass phenomenon. Millions of people could now interact directly with a machine that wrote astonishingly coherent texts, painted impressive pictures, or could even program. **Since then, development has continued to race ahead: just recently, Google demonstrated a video generator called Veo 3, which creates photorealistic videos from simple text commands, complete with sound and lip-syncing. Synthetic content is thus practically indistinguishable from real content.**

What was once a dream, the bold ideas of Rosenblatt and Hinton, is now reality. No longer ghosts in the machine, but active forces that permeate our world: helpful, irritating, sometimes unsettling. They are changing what it means to be human. And you? You are right in the middle of it.

Chapter 2: The Ghost in the Machine – How AI Shapes Our Lives Today

The fire that once began with a small spark now burns brightly. And its light falls directly into your living room. Onto your desk. Into your pocket. AI is no longer "out there." It has become an invisible hand—a force that organizes your digital life, guides you through traffic jams, makes decisions—and sometimes throws everything into disarray. So let's take a closer look at where this former ghost in the machine is active today and what traces it leaves behind.

For example, in our daily lives. Your smartphone is a small AI powerhouse. Facial recognition, autocorrect, voice control, personalized feeds, real-time navigation, music recommendations—all of these are AI systems, trained to understand your needs. This AI is like an invisible butler. Isn't it amazing when your music service plays a song that perfectly matches your mood? It's sometimes almost uncanny, as if someone is listening or watching very closely.

And this butler also has access to areas with far greater significance. In healthcare, things are happening that would have sounded like science fiction just a few years ago. **Systems like Google's Med-PaLM analyze medical reports and lab results, and suggest diagnoses—often with a precision that surprises even experienced specialists. It's as if a super-doctor**

has been made to memorize a library of all medical knowledge, a doctor who recognizes patterns in data that have eluded us until now. At the same time, DeepMind's AlphaFold has predicted the structure of over 200 million proteins—a massive breakthrough that could immensely accelerate the development of new drugs and therapies. And just recently, researchers used AI to discover a new class of antibiotics against multi-resistant germs. The promise is a medicine tailored to each individual. But: who has access to these prognoses? And doesn't this technology lead to a two-tiered healthcare system? The ethical challenges are at least as great as the opportunities.

The upheavals in the world of work are similarly dramatic. AI is penetrating highly skilled fields. **Lawyers use AI for contract analysis, journalists for research. Programmers work with tools like GitHub Copilot, which write code for them. And autonomous AI agents like Devin are waiting in the wings to implement entire projects on their own.** Large companies are already announcing that they will not fill thousands of administrative positions. So the question is no longer if, but how quickly AI will change human labor. This creates enormous uncertainty and forces us to rethink our image of humans as exclusive creators. There is pressure to adapt everywhere—and with it, the question: Is AI a tool, a partner—or possibly a successor?

It becomes particularly explosive in the political sphere. This is about the very foundations of our democracy: information, opinion formation, trust. We are experiencing a normalization of disinformation, fueled by AI. Deepfakes make it possible to make politicians say or do things that never happened. **In early 2024, an AI-generated voice that sounded like President Joe Biden called voters and urged them not to go to the primary election—a clear attempt at election manipulation. And with new video generators like Sora or Veo 3, we face a flood of potentially misleading video material that requires almost no effort to create.** This fundamentally undermines trust in visual and auditory evidence. When you look at your feed today, you've surely seen a video and wondered: Is this really real? That millisecond of doubt. That is the new reality, in which you must actively decide who and what you still believe.

At the same time, social media algorithms act as an accelerant for social division. AI systems learn at lightning speed what captivates us—often content that triggers anger, fear, or outrage. The results are filter bubbles and echo chambers. AI thus becomes a tool of fragmentation. Trust in established media is declining, and conspiracy theories are spreading more easily. Our social media feed may seem like a warm, familiar place. But what if this space was designed by architects who keep us entertained—while simultaneously shielding us from conflicting perspectives?

The fuel for all of this is our data. Every online search, every click, every location stamp leaves a digital trace. The research of Michal Kosinski showed how these data can be used to create detailed psychological profiles that reveal more about us than we realize. These profiles are the capital of the digital age. Companies like Google and Meta earn billions from them. We live in a system of surveillance capitalism, in which our most intimate behaviors are turned into a commodity. And we often tacitly agree to this—for the illusion of convenience. We become transparent and manipulable, click by click.

But why does this work so well? Because AI meets a partner that makes its job easy: the human psyche. We are not rational machines. **Cognitive psychology, from Daniel Kahneman to the present day, has proven how susceptible we are to cognitive biases. Our brain prefers to operate in energy-saving mode, the fast, intuitive "System 1." It loves simple stories and quick validation. It's the place for errors in thinking—for biases.** The confirmation bias makes us seek confirmation. The availability heuristic makes us consider dramatic information more likely, and at the same time, we overestimate ourselves (overconfidence bias).

This tendency toward cognitive laziness is thus the ideal breeding ground for algorithmic manipulation. The personalized feeds and emotional headlines appeal to our System 1. Critical

thinking, fact-checking—that is the domain of the strenuous "System 2." And the digital environment is often designed to keep us from that effort.

Or from the effort of interpersonal relationships. For some time now, we have seen the emergence of AI-powered forms of relationships. Chatbots like "Replika" offer perfect partners on demand—always understanding, never contradictory. Are we unlearning the ability for genuine, imperfect human connection when we get used to the seamless perfection of a simulation? Reports of a young person who ended their life based on the instructions of a chatbot, or the desire to marry an AI, are no longer science fiction. They show how deeply this technology can intervene in our emotional lives.

It is high time to ask important questions.

Chapter 3: A Look in the Mirror – Navigating the Age of AI

Development is progressing at a rapid pace: in the AI community, it's often jokingly warned, "AI will never be as bad as it is today." AI systems are becoming more powerful, more autonomous—and harder to understand. It is time to ask: How do we want to live with them? How do we ensure that their potential serves the common good—without us losing control, without putting our humanity at stake? This is not a theoretical problem. It is a real, urgent challenge. And it demands answers. Not just from experts, but from all of us. From you. And from me.

One of the most pressing challenges is so-called **AI Safety**. The closer we get to an Artificial General Intelligence (AGI)—an AI with superhuman cognitive abilities—the more the question arises of whether we can still control it at all. How can we ensure that it remains benevolent toward us?

One aspect of AI Safety is the growing incomprehensibility of these systems. For example, imagine a brilliant chef who prepares the most amazing dishes. You give him the ingredients, and out comes a masterpiece. But if you ask him exactly how he did it, what his secret is, he just shrugs and says, "It just felt right." It's similar with the most complex AI models. Even their developers can often no longer trace exactly how the AI arrived at a particular answer. The path from input to output remains a mystery. Do we want to entrust our lives to such a brilliant but unpredictable chef when it comes to our health, our money, or our safety? This unpredictable behavior also manifests when language models suddenly invent facts—this is called "hallucinations." It's as if the chef suddenly used an ingredient that wasn't even in the kitchen.

This lack of transparency is further increased by new 'agent systems' like Devin, developed by Cognition AI. These agents not only react to instructions but also plan complex tasks independently. What happens when such an autonomous AI makes unforeseeable decisions, and we cannot understand the 'why' step? **Do we then just let the "code" run, hoping that everything will be fine?**

Another core problem is **alignment** with human values. How can we ensure that a highly developed, possibly superintelligent AI permanently acts in the interests of humanity? What sounds simple is unsolved. First: Which values? Human values are diverse, often contradictory. Second: How do we translate vague concepts like fairness or compassion into precise code? Third: How do we prevent an AI from independently modifying its goals or evading our control? Are we opening Pandora's box?

To put it simply: we are trying to teach a potentially superior intelligence what is 'good.' But what exactly does that mean? Your good? My good? The moral sentiment of the present day? **The debate about whether and when we will achieve AGI is in full swing. Estimates vary: some**

speak of a few years, others of decades. But the mere possibility forces us to think today about questions that could be existential tomorrow.

Research labs worldwide are working on these questions. On a political level, too, there are attempts to create rules for the use of AI, for example, with the European AI Act. But technology is developing faster than legislation. Amidst this pace, the geopolitical competition is intensifying—especially between the US and China. In this race, safety concerns can easily fall by the wayside.

Beyond these existential risks, AI is already having profound societal impacts today. Inequality is growing: between those who develop and control AI systems, and those who merely function as users or data sources. The world of work faces massive shifts. Fundamental rights like privacy are under pressure. And democracy is being challenged as the automated production of disinformation exacerbates social division and power becomes concentrated in the hands of a few tech giants.

The ease with which AI delivers perfect results also poses a double danger: the dissolution of reality, when nothing visible is believable anymore, and the escalation of performance pressure, when humans constantly compare themselves to the flawless efficiency of a machine.

What does this mean for us now? Should we resign? Should we demonize the technology? I believe we must look in the mirror and ask ourselves questions. About our values, our goals, about a meaningful future. We must try to understand what is happening in order to develop a new kind of informed autonomy.

So what can we do?

So what to do? It doesn't require a great revolution, just you. Your curiosity to understand what is happening—not as an expert, but as an aware human being. This curiosity leads to a new kind of vigilance, a radical media literacy in a world where soon everything could be a fake. Soon, you might not know if the video in which your granddaughter asks you for money is really from her—your ability to doubt will become your most important shield. This "muscle training for the mind" requires you to recognize your own thought patterns: your tendency for quick confirmation, your vulnerability to emotional bait. By leaving your digital footprints on the internet more consciously, like deliberate steps in the sand, you also protect your freedom. And this inner vigilance must have an outward effect: Your voice carries weight in the societal debate where the rules for our future are now being made. Because in the end, in a world that is becoming ever more technological, perhaps our greatest task is to cultivate precisely what a machine cannot: to feel, to love, to dream, to doubt—and to give our existence a human meaning.

Because ultimately, Artificial Intelligence is one thing above all: a mirror. It mirrors our intelligence, our creativity, our striving for progress. But also our flaws, our biases, our greed, our complacency, our fears. Looking into this mirror is not always pleasant. It forces us to confront ourselves.

But therein lies our chance. The future is not a script that has already been written. AI is not a destiny to which we must surrender. It is a force that we can shape—if we take responsibility for it. This requires courage, vigilance, and the willingness to ask uncomfortable questions. It is about using one's own mind and sharpening one's own judgment.

The machine is becoming more intelligent, that is certain. But the real question this mirror asks us is: Are we also becoming wiser in how we deal with it? Will we find the strength not only to use its answers but also to ask the right questions—of it, and above all, of ourselves? Because the greatest danger lurks not in silicon, but in our own complacency.