N°6 - JULY/SEPTEMBER 2025

CYBER-IT MAGAZINE

CYBER IS A MARATHON NOT A SPRINT !

HISTORY
80 Years of AI: A Journey Through Dreams and Disillusionment

EUROPE & AI
The Road to Digital Sovereignty

AI ACT
How Europe plans to regulate Artificial Intelligence

AI VS AI
The new Frontier Between Human and Machine Intelligence

CYBER RESILIENCE
Predicting Cyber Incidents Before They Occur

Ethics and A
How Do We Build Ethical and Inclusive AI?

SPECIAL FEATURE

ARTIFICIAL INTELLIGENCE

Inquiry ? Requests ?
cyberit.magazine@gmail.com

EDITO

It used to be a science fiction fantasy, now artificial intelligence has become one of the most strategic challenges of our time. In this special feature, we look back on 80 years of dreams and disillusions around AI, a history marked by hope, spectacular breakthroughs, but also crises of trust and unfulfilled promises.

Today, AI is no longer only defined as an object of research, it has become an active player. And not just any player. Intelligence versus AI is the new arena. In the cyberspace, AI is both defender and threat. It protects, anticipates, and corrects but it is also weaponized by malicious groups to amplify their offensive capabilities.

But if AI is a force, what kind of force are we shaping? Algorithms influence our choices, our freedoms, and our rights. Ethics is no longer optional. Building AI that is inclusive, responsible, and transparent has become a priority if we are to avoid repeating, or worsening the failures of the real world in the digital one.

In response to this new reality, Europe is getting organized. With the AI Act, it is taking a step ahead in the regulatory space. But is this ambition enough? Can we really talk about European digital sovereignty in the face of American tech giants and Chinese power? And above all, is that sovereignty still within reach?

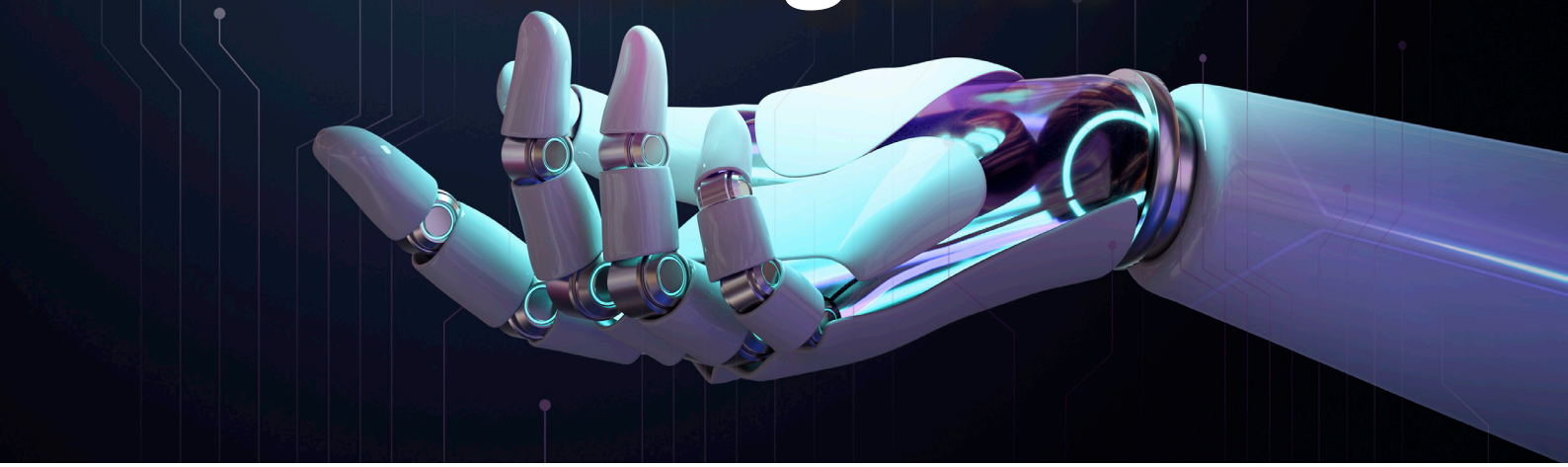
In an era where cybersecurity incidents are no longer a question of "if" but "when", we explore the foundations of cyber resilience. A well-designed AI system can not only detect threats, but also prevent them before they occur.

Enjoy the read !

ARNAUD LEROY

SOMMAIRE

Artificial intelligence



Artificial intelligence is not a new concept. But over the past seventy years, it went into a remarkable transformation, an evolution that challenges established beliefs, disrupts traditional practices, and redefines the cybersecurity landscape.

In 1950, **Alan Turing** asked a simple yet profound question: “Can machines think?” Seven decades later, this question is no longer purely philosophical. Artificial intelligence is everywhere now: in our industrial systems, it influences financial decisions, drives vehicles, generates code, writes reports, mimics our voices, and infiltrates our information systems.

Once a niche discipline confined to academic labs, AI has become a strategic technology at the heart of political, economic, and security concerns in the 21st century.

AI’s development has not been linear. It has gone through periods of hype, long pauses, and sudden reinventions. It started with symbols and then fed on data. It moved from strict rules to flexible models. Each decade has shaped a part of this silent revolution, now deeply embedded in cybersecurity operations centers, surveillance algorithms, and the digital weapons used by cybercriminals.

This report explores the major milestones of this evolution: from symbolic systems to deep neural networks, from generative AI to its double role in cybersecurity, both as a shield and as a weapon.

1950-1960 : an idea born in the lab

It all began with visionaries like Alan Turing, who in **1950** introduced the now-famous Turing Test, a thought experiment meant to explore whether a machine could imitate human intelligence.

Just a few years later, the **1956** Dartmouth Conference would mark what many see as the official birth of artificial intelligence. The researchers gathered there were confident: within a generation, they believed, machines would match human intelligence.

This pivotal decade saw the emergence of some of AI's earliest landmark programs. In **1955**, **Logic Theorist** developed by Allen Newell and Herbert Simo, was presented as the first AI capable of autonomous reasoning. It solved mathematical theorems by mimicking the logic of human problem-solving. It was a turning point: machines were no longer just calculators, they were starting to "think."

In the **1960s**, at MIT, Joseph Weizenbaum introduced **ELIZA**, the first natural language processing program designed to engage in human-like conversation. Simulating a psychotherapist, ELIZA responded to users by turning their statements into open-ended questions. Despite its technical simplicity, built on basic keyword-matching rules,

ELIZA demonstrated the emotional impact a human-machine interaction could produce. Some users, even when aware they were speaking to a machine, developed a genuine attachment to the program.

Though limited, these two early programs laid the groundwork for the two main branches of artificial intelligence: formal logical reasoning and natural language interaction. They already reflected both the ambition and the limitations of an emerging A, still far from consciousness, yet already capable of simulating basic human behaviors.

1960-1970 : High hopes, hard realities

Logical and symbolic systems advanced rapidly during this period. One of the most iconic examples was **SHRDLU**, developed between **1968 and 1970** by Terry Winograd at the Massachusetts Institute of Technology (MIT).

The program could understand instructions in English, manipulate objects in a simple fictional world, and engage in dialogue with the user. It could, for instance, respond to commands like "put the red cube on the green cylinder" or explain its own actions.

SHRDLU gave the illusion of real language comprehension. But its success was confined to a tightly controlled, highly structured micro-world. Outside of that narrow environment, the system could not generalize or adapt.

Still, it fueled optimism. Many believed that full natural language understanding, seamless translation between human languages, and even a general form of intelligence, were just around the corner.

That enthusiasm was shared and funded by the U.S. government. The Defense Advanced Research Projects Agency (DARPA) began backing large-scale AI projects. Massive grants were awarded to universities to build systems capable of dialogue, learning, and even reasoning.

But limitations quickly surfaced. The complexity of human language, the lack of rich data, and limited computing power all slowed progress considerably.

1970-1980 : The Rise of Logic and Expert Systems

Symbolic AI, based on manipulating explicit rules and symbols, continued to dominate research. This decade marked the rise of the first so-called “expert systems”, programs designed to mimic the reasoning of a human specialist within a narrow field.

One of the most notable examples was **MYCIN**, developed in **1972** at Stanford University to help doctors diagnose and treat bacterial infections. MYCIN could ask questions, interpret the answers, and recommend an appropriate antibiotic with impressive accuracy for the time.

It relied on a knowledge base made up of several hundred hand-coded “if-then” rules and statements contributed by domain experts. But these systems soon ran into their own limitations. First, they couldn’t learn new rules or adapt to unfamiliar data, any improvement required a human to manually add or rewrite the rules. Second, maintaining them became increasingly difficult: every new modification risked unexpected side effects, and as the rule base grew, the system became harder to manage.

As these systems became more complex, so did the challenge of maintaining them. Rules began to interact in unexpected ways. And as soon as the system was taken outside the strictly defined context it was built for, it crumbled. Expert systems couldn’t handle uncertainty, ambiguity, contradictions, or incomplete information.

The intelligence they simulated was purely deductive, based on fixed reasoning without any form of intuition, inductive learning, or generalization. They could give the illusion of competence in narrow, well-defined environments, but quickly showed their limits when faced with the messy complexity of the real world.

1980-1990 : The Rise and Fall of Expert Systems

Expert systems experienced a true boom during the **1980s**, driven by earlier academic successes and advances in software engineering. The idea was compelling: capturing an expert’s knowledge into logical rules could replicate their reasoning, around the clock, without human error or fatigue.

Fields like finance, medicine, industrial manufacturing, and aerospace launched ambitious projects, hoping to automate part or all of the decision-making process. Languages such as **OPS5** and **Prolog**, along with dedicated platforms like **XCON** (used by DEC to automatically configure its computers), became technological showcases.

A wave of AI-focused startups emerged, supported by confident investors. Major corporations created internal AI divisions. Expectations were high: cost reduction, increased reliability, and greater competitiveness.

But practical limitations became apparent very soon. Developing an expert system required an enormous amount of knowledge engineering. As soon as a situation strayed from the expected framework, the system would collapse. High maintenance costs and underwhelming results when faced with the real world’s complexity amplified the problem.

As disillusionment set in, investors gradually pulled out. Companies shut down their AI departments. This widespread retreat marked the first true "AI winter."

1990-2000 : The Rise of Machine Learning

The paradigm shifts radically: instead of trying to explicitly code the rules of intelligence, the goal now is to make them emerge from data. This change in direction marks the beginning of machine learning, an approach based on algorithms capable of identifying patterns in large datasets. Supervised learning where machines are trained on labeled examples enables the creation of predictive models tailored to specific tasks.

Machines are no longer just logical executors, they are beginning to truly learn, to adapt, to generalize from experience. This turning point, still largely unnoticed by the general public at the time, laid the groundwork for modern artificial intelligence. Academic research increased, focusing on techniques like cross-validation, overfitting, and regularization, core concepts that would establish the statistical rigor of the field.

Deep Blue's victory over Garry Kasparov in **1997**, the first time a world chess champion lost to a computer in an official match marked a historic turning point: for the first time, a machine outperformed a human in a domain symbolic of intelligence.

This highly broadcasted event pushed artificial intelligence into the spotlight and into public consciousness. It was no longer just about abstract research or academic demonstrations, AI entered the collective imagination as a force capable of competing with, and even surpassing, humans in complex cognitive tasks.

2000-2010 : Internet, Big-data, and large-scale AI

With the rise of the Internet and the advent of Web 2.0 marked by interactivity, social networks, and the massive creation of user-generated content, AI found an unprecedented playground.

Tech giants like Google, Amazon, Facebook, and Netflix capitalized on this revolution to develop increasingly refined recommendation algorithms, capable of predicting preferences, adapting the user experience in real time, and optimizing targeted advertising.

Natural language processing advanced rapidly, enabling notable progress in text understanding and generation, although systems remained limited to specific tasks.

It was during this period that the first voice assistants emerged, and search engines gained in relevance thanks to more sophisticated language algorithms able to interpret the intent behind queries.

In parallel, the explosion in the volume of data available on the Web radically transformed the learning capabilities of AI models. This phenomenon, known as **"big data"**, paved the way for a more empirical form of AI, less reliant on explicit human-designed models.

The collection and analysis of these vast datasets became possible thanks to distributed infrastructures such as **Hadoop** and **MapReduce**, which enabled the parallel processing of massive amounts of information. These tools laid the foundation for large-scale AI.

2010-2020 : Deep Learning & Breakthroughs

Deep learning networks revolutionized the landscape of artificial intelligence by unlocking performance levels previously out of reach. Enabled by the availability of large amounts of labeled data and the computational power of GPUs, these architectures overcame the limitations of traditional approaches.

In 2012, the **AlexNet** model, winner of the ImageNet competition, marked a decisive breakthrough in computer vision by reducing error rates in image recognition.

In 2015, **ResNet** introduced residual connections, making it possible to train networks with hundreds of layers and paving the way for unprecedented model depth.

These breakthroughs spread rapidly with the rise of open-source platforms like Google's TensorFlow and Meta's PyTorch, which made AI accessible to a broad community of researchers, developers, and companies. The ecosystem became more democratic, accelerating research, encouraging scientific reproducibility, and facilitating the large-scale integration of AI in industrial applications.

Within just a few years, speech recognition became common on smartphones, machine translation began rivaling human performance in certain languages, and computer vision started being applied in medicine, automotive systems, and security.

In natural language processing, BERT, released in 2018 and based on the Transformer architecture, enabled nuanced contextual understanding of words, transforming information retrieval, text classification, and automated question answering.

2020-2030 : Generative Intelligence and omnipresent AI

The 2020s mark a dramatic shift in the evolution of artificial intelligence, with the impressive rise of generative models. Powered by architectures such as the Transformer, systems like **ChatGPT** (OpenAI), **Claude** (Anthropic), **Gemini** (Google), and **LLaMA** (Meta) are pushing the boundaries of cognitive automation.

Their strength lies in their ability to produce, on demand, text, code, images, videos, voice, and even interactive interfaces, with a perceived quality often comparable to that of a human expert.

We are no longer talking about specialized tools, but true general-purpose assistants capable of performing cross-disciplinary tasks: writing articles, conducting legal analysis, scripting, drafting business plans, translating, generating illustrations, composing music, and more.

These so-called foundation models are pre-trained on massive dataset and can be fine-tuned for specific use cases. They are also transforming the world of creation: illustrators, designers, filmmakers, and composers use them as sources of inspiration or prototyping tools. But this rapid expansion comes with unprecedented challenges. Risks related to automated disinformation, deepfakes, hallucinations (fabricated content generated by the models), and the loss of source traceability becomes concerning but also fascinating at the same time (on a science perspective).

2030 and Beyond : Emotional AI and Autonomous Intelligence

Looking toward **2040** and beyond, advances in AI will no longer be limited to imitating human intellect. They will aim to integrate its affective, relational, and sensory dimensions. So-called emotional models will be capable of detecting, simulating, and responding to users' emotional states. Embedded in human–machine interfaces, companion robots, conversational agents, or caregiving assistants, these emotional AIs will adapt their behavior to the psychological and social context of their interlocutor.

The goal will be to make communication more natural, more empathetic, and even therapeutic. These AIs will rely on biometric sensors, heart rate, micro-expressions, voice tone, as well as massive behavioral data sets to tailor their responses.

At the same time, system autonomy is becoming widespread: vehicles, drones, smart home assistants, and industrial systems operate without human supervision, making real-time decisions based on environmental data analysis. AI is thus becoming a full-fledged decision-making agent in many strategic fields. Ethical questions are multiplying: how far should we allow a machine to decide? How can we ensure traceability, accountability, and non-discrimination?

AIs will be used not only to attack or defend systems but to manipulate perceptions, spread personalized synthetic narratives, and generate falsified content that is indistinguishable from reality. Competing models will clash, one aiming to deceive, the other to detect. The line between truth and simulation will blur. Society will enter an era of widespread algorithmic mistrust.

ChatGPT's visual summary of the topic

Intelligence vs AI the new frontier

"The cybercriminal no longer writes code. He commands intelligences like a general commands his troops"

The new frontier of Cybersecurity

The Frontlines of Cybersecurity are being redefined. often seen as a technical discipline reserved for network engineers and cryptography experts, cybersecurity is entering a new era, one shaped by the rise of artificial intelligence. What once served to optimize and automate is now becoming a weapon in its own right, used for both offense and defense. We are in the midst of a silent yet brutal transition: threats are no longer exclusively human. Attacks are no longer manual. And defenses can no longer afford to be purely reactive.

Artificial intelligence, in its most advanced forms, now lies at the heart of the battlefield. For the first time in cyberspace history, we are witnessing a striking reality: AIs designed to attack being hunted, blocked, or neutralized by other AIs. Machines no longer protect only humans, they now fight other machines. The rise of offensive AI marks a turning point.

Cybercriminals, organized crime groups, and state-sponsored actors are increasingly exploiting the capabilities of machine learning and generative models. Gone are the days of sloppy phishing campaigns or ransomware sent blindly to random inboxes. Today's attacks are adaptive, deeply personalized, and disturbingly convincing. Using AI trained on stolen internal documents, leaked emails, and online public interactions, it is now possible to generate spear phishing messages so realistic that even the most cautious employee would struggle to spot the deception.

Artificial intelligence also allows for the automation of complex tasks that once required expert knowledge: scanning systems for vulnerabilities, testing thousands of passwords with advanced optimization techniques, injecting polymorphic malware to evade detection, or writing fully functional exploit code in seconds. Tools like WormGPT, FraudGPT, and other unofficial AIs circulating on dark web forums can now generate phishing scripts, malware templates, and bypass strategies with a level of fluidity and camouflage techniques that far exceed traditional means.

But perhaps the most radical shift is not in the sophistication of attacks, it is in their autonomy.

"Zero-click" attacks, those requiring no human interaction can now be triggered by intelligent bots, executing cyber operations at machine speed.

AI systems are now capable of identifying a vulnerability and deploying an exploit within the same second. Semi-autonomous models left to roam digital spaces, social networks and messaging platforms can observe, analyze, and mimic human behavior, impersonate coworkers, and manipulate critical internal decisions. Human reaction time is no longer enough. Post-mortem analysis of an incident no longer prevents it from happening again. The cybersecurity paradigm must undergo a radical shift.

It is no longer about building walls, but about deploying counter-AI.

Responding within the same algorithmic timeframe as the attacker is the priority. At this stage, security models based on signature detection or traditional firewalls are becoming obsolete.

"The future of cyber doesn't depend on our ability to write better rules, it depends on our ability to train better AI"

Building a strong defense

Faced with the growing sophistication of offensive AI, major tech companies, intelligence agencies, and cybersecurity firms are making massive investments in AI-driven defensive systems.

These AI are not simply extensions of existing tools, they have become entities of continuous monitoring, behavioral analysis, prediction, and automatic response.

By analyzing network traffic in real time, correlating billions of weak signals, and comparing millions of logs, these AI can detect suspicious activity before any damage occurs. Some even anticipate adversaries' tactics, much like a chess game where every potential move is simulated in advance.

Defensive AI is no longer just an alarm, it has become a counter-attack force.

These systems, integrated into architectures such as XDR, rely on machine learning engines continuously trained on real-world incidents.

They learn from every breach, every attack, every anomaly. They no longer protect only systems but also identities, behaviors, and relationships. They can detect that an employee's identity has been compromised based on a slight change in writing style, or that a server is under attack from an abnormal spike in activity lasting just milliseconds.

Sometimes, they react before human validation, isolating network segments, temporarily cutting access, or reconfiguring security rules.

But while AI enables defenders to respond faster, it also introduces a new strategic challenge: algorithmic asymmetry. In a world where AI systems face one another, the strength of a system no longer depends on its infrastructure or rules but on the quality of its models and the richness of its training data. The ability of its algorithms to generalize unknown threats are now what determine superiority. Cyber warfare plays out in the invisible arena of data now.

The side with the best compromise datasets, the most advanced simulation models,

and the fastest adaptation cycles holds the upper hand. This silent race is leading to the growing militarization of artificial intelligence in Cyber. Major powers, such as the United States, China, Russia, and Israel are secretly developing cyberdefense and cyberoffense AIs, some capable of launching automated disinformation campaigns, remotely disabling critical infrastructure, or spreading autonomously within air-gapped systems.

The battlefield is global, dematerialized, and for the first time, largely beyond the scope of human oversight. How far should an AI be allowed to go in making autonomous defense decisions, especially if those decisions include retaliation or even proactive strikes?

At what point can an AI be held accountable for a cybercriminal act?

What safeguards, international norms, or treaties need to be defined to regulate this new form of conflict? We are standing at the threshold of a new kind of law, an emerging framework for algorithmic warfare. Still in its early stage, but increasingly essential.

Meanwhile, companies do not have the luxury of waiting for international law to catch up.

They face discreet threats every day many of which are undetectable without intelligent tools. Most CISOs today understand that the future of their defense strategies is AI-driven.

But this shift raises tough internal questions:

- How do you train an AI without violating GDPR?

- How do you mitigate bias in models that could lead to massive false positives?

- How do you recruit talent capable of navigating cybersecurity, artificial intelligence, and the regulatory landscape, all at once?

Who has the best prediction capabilities? Who reacts the fastest? Who learns the most efficiently?

This is the new arms race. A digital Cold War. Invisible. Constant.

Cybersecurity is no longer a defensive discipline and the actors are no longer just human.

In cyberspace, AIs are already at war, and this is only the beginning…

The AI Act : Europe's Regulatory Framework for Artificial Intelligence

by Virginie Mathivet

Artificial intelligence is everywhere: in our mobile apps, professional tools, and public services. But it's far from flawless.

We remember the case of Uber's self-driving car failing to detect a pedestrian outside of a crosswalk, of Google Photos labeling Black people as gorillas, or of Austria's public employment service algorithm that systematically disadvantaged women applying for jobs in tech.

These failures all share a common root: biases embedded in the data used to train the models. Sometimes subtle, often unintentional, these biases can be identified and corrected, if we pay attention and apply rigorous methodologies.

It is precisely to prevent such abuses that the European Union introduced the AI Act (short for Artificial Intelligence Regulation). This legislation aims to regulate the use of AI, especially when deployed in high-stakes settings, and to require companies to follow good practices in the development and ethical oversight of artificial intelligence systems.

Is the AI Act slowing innovation ?

One of the main criticisms to the AI Act is that it could hold back innovation by imposing administrative and technical burdens on European companies, while their foreign competitors, particularly in the U.S. and China, operate in more relaxed or even non-existent regulatory environments.

1. The regulation applies to all, even outside Europe

The AI Act applies to any AI system used within the European Union or that affects EU citizens regardless of the developer's country of origin. An American or Chinese company wanting to market an AI solution in Europe will also have to comply with the regulation's requirements.

2. Regulation is becoming global

The EU is not alone. The United States, China, Canada, and the United Kingdom are also working on regulatory frameworks for AI, though their approaches may vary.

A global awareness is emerging: defining what makes AI "trustworthy" and regulating its use is becoming a diplomatic, industrial, and societal priority.

3. Regulatory sandboxes for experimentation

To avoid hindering research and development, the AI Act provides for the creation of regulatory "sandboxes." These controlled environments allow companies to test innovative systems in real-world conditions without having to meet every regulatory requirement from the start. Innovation remains possible as long as it is supervised.

4. Requirements aligned with best practices

The AI Act's demands are not random: most of them reflect existing best practices already followed by companies who care about quality, safety, and ethics. A serious organization documenting its decisions, testing its models, and ensuring human supervision should have no trouble in meeting the compliance requirements. If a system cannot meet these standards, its strength and quality are worth questioning.

To remain relevant, the European Commission has committed to regularly revising the regulation, to ensure it evolves with technological progress and real-world feedback

AI Risk Management Framework

Unacceptable Risk

These are systems banned by default. They aim to manipulate behavior or enforce social control, such as real-time mass surveillance or citizen scoring systems.

High Risk

These systems may threaten fundamental rights or the physical or moral safety of individuals. Examples include recruitment algorithms, judicial decision support tools, and systems deployed in education or healthcare. Such AI must comply with strict regulations and obtain official certification.

Limited Risk

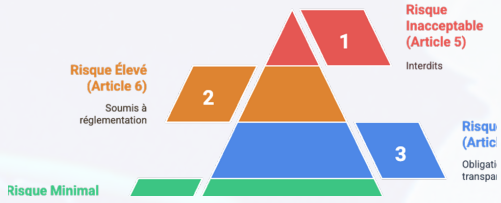
These systems do not cause serious harm if they fail but interact directly with the public. Examples include conversational assistants and recommendation tools. The main requirement is transparency: users must be informed they are interacting with an AI. In some contexts, appropriate training may also be required.

Minimal Risk

These systems pose very low risks or are used in purely technical or internal contexts. They are not subject to specific regulatory obligations, but best practices are recommended.

General purpose AI (GPAI)

Some systems, such as large language models (LLMs) like ChatGPT, can be applied across a wide range of uses, including critical contexts. The AI Act subjects them to a specific regulatory framework.



Requirements for High risk AI systems

When an artificial intelligence system is classified as high-risk, the AI Act imposes a series of strict requirements for safety and protecting the fundamental rights of European citizens.

Companies involved must obtain European-level certification based on a file demonstrating their compliance with several technical and organizational obligations. Many of these requirements are based on best practices already applied in software engineering and cybersecurity:

- **Risk Management framework** Identification, assessment, and mitigation of risks associated with the AI system throughout its entire lifecycle.
- **Data Governance** Management processes to guarantee the quality, security, and confidentiality of data used for training, testing, and operation.
- **Technical Documentation** The system must be fully documented, detailing its architecture, objectives, development process, management of biases in datasets, as well as performance and robustness metrics.
- **Event Logging** Systems must keep detailed logs of their operations to trace decision-making, identify errors, or detect biases.
- **User Transparency** The AI must be clearly identifiable. It should communicate its limitations and inform users that errors may occur.
- **Human Supervision** A human operator must be able to monitor the system's operations, ready for intervention at any time, and even override or block automatic decisions.

Most of these requirements are inspired by quality standards already implemented in many organizations. For companies committed to responsible development, the AI Act should not be seen as a hurdle but rather as a structuring and legitimizing framework.

Comment façonner une intelligence artificielle plus éthique et plus inclusive ?


Hamilton Mann

Hamilton Mann, a leader in the Tech industry, is a pioneer in the fields of AI and digital transformation. He is a lecturer at INSEAD and HEC Paris, a PhD candidate in AI at École des Ponts Business School – École des Ponts et Chaussées, and is listed among the winners of the Thinkers50 Radar, which recognizes the world's most influential emerging management thinkers. He has contributed to publications on AI challenges, including the Stanford Social Innovation Review, California Management Review, Rotman Management Magazine, I by IMD, Wharton Knowledge, INSEAD Knowledge, Polytechnique Insights (the journal of École Polytechnique), and Leader to Leader (the journal of the University of Pittsburgh, PA). He is also a regular contributor to Forbes US, The European Business Review, and Harvard Business Review France.

As artificial intelligence gradually establishes itself across all areas of human life, a crucial question arises: how can we ensure that it is truly ethical and inclusive?

Developed from data and models rooted in imperfect human contexts, AI inevitably inherits our biases, whether conscious or not.

Designed to meet the needs of a target audience, it often tends to exclude those outside that group, thereby perpetuating existing forms of discrimination.

This issue is becoming more critical as AI becomes increasingly universal. Valued at $87 billion in 2021, the global AI market could reach nearly $1.6 trillion by 2030.

From voice assistants to autonomous vehicles, recommendation engines, surgical robots, and customer relationship management tools, AI is transforming practices, professions, and social dynamics. It is permeating personal, professional, and even political decision-making.

Yet behind the apparent neutrality of algorithms lies a deeper tension between the pursuit of economic performance and the imperative to preserve values of equity and justice. How can we ensure that the biases or segmentation patterns embedded in the data powering AI do not result in systems that discriminate against individuals based on characteristics such as gender, skin color, religion, disability, or sexual or political orientation?

This is one of the fundamental challenges raised by the development of artificial intelligence.

A Societal Challenge

Artificial intelligence is not all that "artificial". As its development accelerates at dangerous speed, so does the temptation to harness it for unprecedented forms of differentiation, precision targeting, and economic gain. The goal? Greater growth. Greater competitiveness.

Yet this progress comes with a deepening tension: on one side, the need for organizations and individuals to embrace diversity and advance inclusivity as a method to build a more equitable society.

On the other, a global economic system that rewards and even fuels competitive behaviors, often at the expense of fairness.

Discrimination risks become a winning strategy. And that tension is only growing stronger. AI is capable of codifying, systematically and systemically, the social norms and values of our digital world. This is one of the greatest challenges of our time.

AI is already woven into the very fabric of our lives:

- Virtual assistants handle basic daily tasks
- Market research is conducted by algorithms benchmarking competitors and generating detailed reports
- CRM systems analyze consumer behaviors, purchasing processes, and preferences with increasing intelligence, predicting what customers want before they ask
- Customer service is often managed by chatbots responding to the most common inquiries on websites

And this is just the beginning. A rise of powerful applications is on the horizon:

- Autonomous vehicles, from bicycles and cars to trains, planes, and ships
- Surgical robots assisting doctors in operating rooms
- Content creation (videos, music, articles) generated entirely by machines
- Public policy decisions driven by data, with outcomes predicted by large-scale analytics

We are at a crossroads. Either we proactively use AI to reduce visible and invisible inequalities on a scale never seen before, or we allow it to exacerbate them just as broadly. The age of artificial intelligence is not one of half-measures. It demands clear intentions.

A New Era for Human Learning

We humans are responsible for what machine learning, the foundation of any AI, learns, how it learns, and what it fails to learn.

The way we teach machines shapes their understanding of the world and lies at the heart of 21st century learning challenges.

This requires us not only to continue advancing our own intelligence, but also to deepen our understanding of how machines develop theirs.

Human and machine learning share many conceptual and technical obstacles, revealing an emerging convergence between biological cognition and artificial intelligence.

The distinction between supervised and unsupervised learning mirrors our own ways of acquiring knowledge: one guided by examples, the other driven by self-exploration.

Similarly, the difference between structured and unstructured learning reflects the challenge of processing organized versus raw or chaotic data.

Some models, such as "one-shot" or "few-shot" learning*, attempt to replicate the human ability to learn from just one or very few examples. This contrasts with the "blink" learning described by Malcolm Gladwell, which relies on rapid intuition shaped by past experiences.

The tension between short- and long-term memory, and the trade-off between forgetting and information retention, is a core challenge both for neural networks and the human brain. "Zero-shot learning", which involves generalizing without direct examples, draws a fascinating parallel with Crick and Mitchison's "dream-forgetting" theory, suggesting that REM sleep helps us discard unnecessary memories.

Learning approaches, such as visuomotor paradigms rooted in physical interaction, echo human forms of embodied intel-

ligence. Likewise, multisensory integration combining auditory, visual, and kinesthetic cues illustrates the richness of holistic learning, both in humans and in certain neural architectures.

As we learn how machines learn, we are also uncovering entirely new ways of learning that had not yet been explored or even imagined. These may very well redefine the norms of how we, as humans, approach learning, pushing the boundaries of human intelligence. But let's be clear: intelligence is not the same as knowledge. Getting more knowledge is a necessary condition for expanding our intelligence.

True human intelligence is about our capacity to question, to challenge the status quo, to be more curious, and to generate new questions that reshape what we believe we know and who we believe we are.

In reality, artificial intelligence is far less intelligent than we imagine.

* In one-shot and few-shot machine learning, the model trains on a few examples.

Zero-shot learning takes it a step further, making predictions without any prior examples from the target category in the training data .

A Matter of Understanding

Even without projecting an AI capable of truly matching human emotions, there remains a fundamental distinction between artificial and human intelligence: contextual understanding. Context is composed of numerous factors, some visible, others subtle, or nuanced.

These weak signals are just as crucial to defining a situation. Given that context is in constant evolution, it will take time before AI systems are truly capable of grasping this level of complexity.

If we are to build the kind of AI that serves the common good, it must start with a clear vision. One that allows us to identify which tasks, are and will be, better performed by machines, versus those that remain human, and crucially, those that must continue to be performed by humans, no matter what.

The choices our societies make in designing the frameworks that define how AI is "intelligent" for humans will shape the future of humanity, not only in terms of innovation or the creation of new competitive advantages rewriting market dynamics, but also, and more profoundly, in terms of the societal model we choose to pass on to future generations.

Too often, when we think about machine learning, our mental model frames it as a one-way process: we train the machine and equip it to learn across various domains on its own.

A deep transformation

Artificial intelligence is profoundly reshaping the bond between humans and machines a bond that will become increasingly critical, and increasingly fascinating, to explore. In reality, this relationship is already going two ways.

Which raises the question: What can machine intelligence teach us to help us improve in what we do as human beings? We will need to learn to think differently about many things, in order to make machines do the very tasks that would be humanly difficult, if not impossible, for us to perform in the same way. At the same time, AI will give us the chance to learn and train in fields where expertise today can only be acquired through years of sustained effort and where peak performance has so far been achieved only through human execution.

If AI and the recommendations it produces open unexpected opportunities to improve not only our own intelligence but also the nature of our relationships, even emotional attachments with machines, it also raises delicate questions about corporate social responsibility (CSR). At what point does AI's decision support apply such influence that it silently begins making decisions for humans?

" This complex question is already at our doorstep "

The answer will depend on many factors, including how vulnerable society perceives each of us to be at a given moment in our lives, under specific circumstances. And those perceptions can vary as widely as the individuals themselves. That is why any AI-powered application, device, or technology will need to offer explicit clarity on the limits of the parameters an algorithm considers and those it does not along with the potential implications that could pose risks to oneself or to others. This is essential to ensure responsible AI usage, and to prevent misuse or even prohibited use.

AI forces us to meet the challenge of making it explainable, to everyone, about the causes behind its outputs, so that it can guide decisions that will increasingly affect our lives and society as a whole. Even though, we humans are not always capable of fully explaining the reasoning behind our own decisions in a way that would be universally understood and entirely accurate.

Toward a Trust Economy

AI is set to change the value of work. Some even fear it could replace humans altogether. While the idea of a sci-fi-style AI overtaking humanity like "the Terminator" remains pure fiction, there is one standard the digital society must acknowledge: AI may outperform humans in certain tasks, but it is neither better, nor will it ever be better, at all tasks.

With AI's rise, we are experiencing a shift from a knowledge era to a trust era. This transformation is driven by the demand for greater predictability, accuracy, and efficiency and also, by the need for greater fairness, transparency, and sustainability.

For the future of knowledge workers, digital technology and AI in particular will bring about five major changes. Each will disrupt society to varying degrees, depending on the prevailing nature of work and the perceived value of work in each region of the world.

1. The primary source of anxiety, largely fueled by pop culture's portrayal of AI is the fear of jobs disappearing. This is nothing new. Previous industrial revolutions have already brought similar situations.

2. Then, some jobs will evolve with the help of artificial intelligence. Again, this isn't an unprecedented phenomenon.

3. Next come the professions which will evolve into so-called "tech jobs."

4. And finally, there are jobs that are now hard to even imagine, because they respond to societal needs we barely understand today.

5. But let's not be naive: AI's development is already driving the rise of precarious job positions, jobs created to compensate for AI's current limitations.

For example, the invisible workers who tirelessly label vast amounts of data, performing highly repetitive tasks to help AI learn and to ensure that unacceptable content is blocked on the platforms we use, because it breaks the law. The long-term exposure to such content can take a serious toll on these workers' mental health.

Which of these AI-driven changes will have the biggest impact on the future of work? It's hard to predict. But while AI is not the only force reshaping our century, it's clear that the choices remain ours.

Artificial intelligence itself has no ethics beyond ours. Our ethical principles are, ultimately, part of the functional requirements encoded into AI, embedding the biases we intellectually own. In a way, AI inherits the ethical DNA of its creators.

Making the invisible codes of our societies visible is likely one of the most transformative advances AI will enable humanity to achieve.

This unprecedented transparency, exposing the unsaid and unwritten, will foster greater equality and profoundly redefine society's demand for justice.

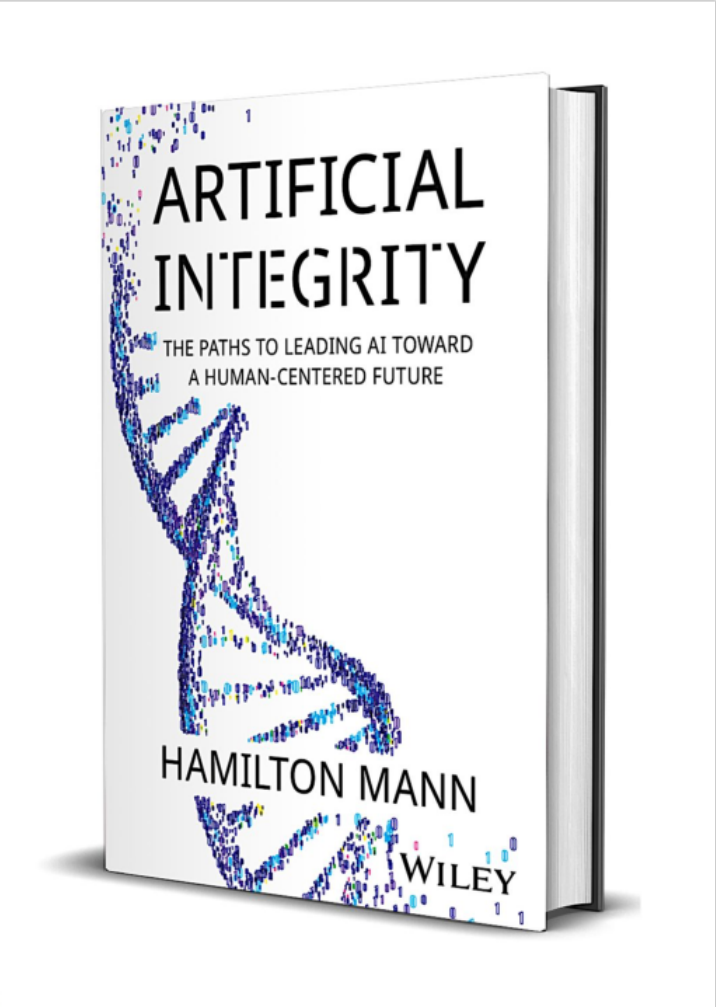
It also offers a chance to ensure that the AI systems interacting with us become the product of collective intelligence, and at best, a vessel for the richness produced by the synergies of human diversity, in all its forms of intelligence.

The improvement of our intelligence through machines will always confront us with the existential question of the human cause we assign to this intelligence's mission.

We must therefore strive to make "artificial intelligence" an intelligence technology inspired by the very best of our humanity excluding all the darker sides of human nature.

This is perhaps the most confusing, yet crucial question for humanity's future. It's an ethical challenge that only our humanity can answer, continuously and responsibly, to build the future we want to live in.

Hamilton Mann



YOUR GENERATIVE AI GUIDE

By Eva Baikeche

A FAMILY-FRIENDLY GUIDE

What is Generative Artificial Intelligence?

A generative AI is a program that can produce content—such as text, images, music, and videos—based on straightforward written instructions known as "prompts."

Generative AI has the potential to assist you...

Compose or rephrase any form of communication, whether it be a message, letter, or email, in any desired tone: professional, informal, "in the style of Victor Hugo," or "in the style of a youtuber."

Summarize a course, article, or video for enhanced comprehension.

Clarify a complex topic with simplicity: a course, quantum theory, French taxation.

Generate images based on concepts or descriptions: capture a photograph of your apartment and request a redesign of a specific decor element.

Organize your daily life: create a shopping list that adheres to your dietary restrictions, schedule a workout session, and devise your travel itinerary.

A request to ChatGPT consumes ten times more energy than a Google search.

Eco-friendly tips with AI

Be direct, not courteous: there is no requirement to say "please" to the AI; every superfluous word consumes additional energy.

Exercise restraint in your generations! One AI image consumes hundreds of watts. Reserve them for genuinely worthwhile concepts.

Select ethical AI: Use tools that safeguard your data and are more resource-efficient!

4 essential reflexes with IA

1- **Always verify the information provided**, as AI may occasionally rely on questionable sources.

2- **Provide him with some context in your request**, as if you are communicating with someone who is entirely unfamiliar with you!

3- **Avoid disclosing excessive personal information**: AI is not a psychologist and may provide misguided advice.

4- **Use curiosity and critical thinking**: it is an excellent tool designed to assist you, not to think on your behalf!

AI Limits

AI can frequently be inaccurate; it is even suggested that it experiences **hallucinations**.

She lacks consciousness and emotions.

She is not creating anything novel; rather, she is merely calculating **probabilities** based on the most effective responses.

It perpetuates biases, including errors, prejudices, and stereotypes.

Our Top AI Tools

seelab
Image enhancement and creation

Suno
Music Composition

Leexi.ai
Meeting Transcript

Lovable
Develop an application

Gamma AI
Presentation Development

CYBER RESILIENCE & ARTIFICIAL INTELLIGENCE: PREVENTING THE INCIDENT BEFORE IT OCCURS

In the current digital landscape, AI is both an asset and a threat Cybercriminals are already using it at scale to stage attacks with a level of sophistication we've never seen before. But today, cybersecurity teams can tap into the very same tools to push back. A silent revolution is taking place, one that is fundamentally transforming our ability to anticipate, detect, and neutralize cyber threats.

ABOUT THE AUTHOR

Christophe Longuepez is an entrepreneur and cybersecurity expert with 15 years of experience.

He specializes in artificial intelligence and its impact on the cyber ecosystem from the evolution of threats to the secure use of AI, as well as leveraging AI technologies to strengthen organizational resilience against risks.

Offensive AI: A Threat going Industrial

The attackers' arsenal is expanding daily thanks to AI. Ultra-realistic deepfakes can now compromise the identity of senior executives with surprising accuracy, turning "CEO fraud" into highly convincing operations. Today, just a few seconds of voice recording are enough to generate a perfectly imitated phone conversation, capable of deceiving even the most vigilant employees.

Polymorphic malware, generated by malicious language models (DarkLLM), evolves continuously to evade even the most advanced EDR/XDR detection capabilities. This industrialization of cyber threats presents an unprecedented challenge. Where a cybercriminal used to spend weeks crafting a credible social engineering scenario, AI now enables them to automate, personalize, and launch large campaigns. Until recently, this asymmetry seemed decisively in favor of attackers.

Yet, the same artificial intelligence that rightly concerns cybersecurity professionals also represents their most powerful ally as it is able to multiply their efficiency. This transformation impacts the entire ecosystem from risk governance to the most technical aspects of cyber defense, fundamentally redefining resilience strategies and shifting cybersecurity from a traditionally reactive posture to a truly proactive discipline.

An improved, more agile & precise GRC

Governance, Risk, and Compliance (GRC) have traditionally been time-consuming domains, requiring complex processes, numerous stakeholders, and substantial human expertise to assess, evaluate, and manage the many dimensions of cyber risk. AI is revolutionizing this approach, offering analytical, processing, and production capabilities far beyond those of conventional methods.

Third-Party Risk Management, Reinvented

Third-Party Risk Management (TPRM) perfectly illustrates this shift. Until now, managing risks related to suppliers, partners, and external providers was often a complex process: lengthy questionnaires, non-existent assessments, and irregular security follow-ups…

AI changes the equation, allowing active and continuous monitoring of these risks, aligned with NIS2 requirements.

Algorithms now analyze, in real time, the attack surface of third-party partners, automatically examining their certifications and their evolution, continuously assess the strategic importance of each partner within the value chain, and detect weak signals or any significant changes that could indicate an imminent compromise.

This ongoing monitoring transforms what was traditionally a bureaucratic process into a truly proactive approach, allowing risks to be anticipated and mitigated as early as possible.

This is especially relevant given that 93% of organizations surveyed by the CyberRisk

Alliance reported having suffered at least one indirect attack through a third-party supplier.

Cyber Risk Quantification: From Art to Science

For a long time, cyber risk assessment has navigated between subjective estimation and expert intuition, resulting in many cognitive biases and poorly reproducible evaluations.

AI brings a scientific dimension to this discipline by cross-referencing thousands of variables technical vulnerabilities, business context, incident history, threat intelligence, and human and organizational factors, to generate quantified, objective scenarios.

This revolution finally allows CISOs and executives to:

- Precisely quantify their risk exposure using reliable metrics
- Prioritize threats according to their real potential impact on business operations
- Justify security investments recommendations to executive management
- Measure performance with objective tangible indicators

Generative AI takes this logic even further by transforming strategic planning. It facilitates the creation of coherent cybersecurity roadmaps by analyzing organizational, budgetary, and technical constraints to propose optimized sequences of actions. By evaluating the cost-benefit ratio of each proposed cybersecurity measure, it identifies the most efficient investments, the ones that maximize security gains for a given time and financial effort. This technological breakthrough transforms cybersecurity strategy development into an objective, data-driven process, enabling organizations to deploy their cyber strategies in a complex environment.

Overcoming Constraints to Unlock Agility

Integrating security into projects the well-known security by design is often perceived by business teams as a major constraint: extended deadlines, unclear requirements, siloed processes, and guidelines seen as disconnected from the project's operational reality.

This traditional, largely manual model has now reached its structural limits due to a lack of suitable tools and insufficient specialized personnel.

AI radically changes this equation. From the earliest stages of a project, it enables automation of security qualification, a critical step often neglected due to lack of time or expertise.

By dynamically and continuously analyzing the core elements of a project, target architecture, types of data processed, functional scope, exposure surface, external dependencies, technologies used, even sometimes without explicit manual input, AI models automatically identify sensitive points and assess their level of criticality. This work relies on proven business risk matrices, recognized frameworks (ISO 27001/27005, NIST, OWASP…), and detailed internal records (security incidents, past audit results).

Building on this foundation, specialized machine learning models or LLMs trained specifically on security challenges, assess the risk level of each identified element.

For example, they can highlight:

- GDPR requirements and HDS obligations put in place from the early phase of any project handling personal or sensitive data.

- A publicly exposed API with weak authentication mechanisms.

- A business-critical project whose application maintenance is handled by an external provider who never got assessed properly.

- Misconfigured cloud services, such as a public S3 bucket with a hardcoded API key, automatically classified as critical. Yet, the true differentiating value of AI lies in its ability to ensure continuous, measurable security throughout the entire lifecycle of the project.

1 Automated Change Collection

Every configuration, cloud setup, or architectural change instantly triggers a contextual re-evaluation of the risk level. AI proactively identifies unplanned or critical modifications, such as undocumented port ope-

nings, unencrypted storage, or unsolicited privilege escalation before they can be exploited.

2 Dynamic Requirements Updates

AI adjusts the necessary security controls in line with project evolution, regulatory changes, or updates to internal security policies.

3 Real-Time Dashboards

Within their respective scopes, Product Owners and CISOs benefit from precise, up-to-date visibility on implemented measures, identified discrepancies, and ongoing corrective actions.

This approach radically transforms traditional cyber security resilience. Security is no longer a one-off process made of manual checks or a deliverable at the end of the project cycle, it becomes an ongoing continuously monitored, objectively measurable process, implemented into modern Agile and DevOps practices.

A new era of Cyber defense: speed and Intelligence for sharper detection

While AI is reshaping risk governance, its most spectacular potential emerges in operational cyber defense. Detection, analysis, and incident response are experiencing a remarkable leap forward, delivering unprecedent efficiency and speed.

Beyond Signatures

Traditional detection systems rely on identifying known threats through signatures or predefined patterns, an inherently limited approach when facing unknown threats, zero-day attacks, or polymorphic malware.

AI introduces behavioral analysis, breaking through these historical limitations.

Machine learning algorithms build a baseline of normal activity for every system component, users, applications, network flows, and data access. Any significant deviation from these patterns triggers an alert, enabling the detection of zero-day exploits that conventional tools would miss.

This behavioral approach is effective against Advanced Persistent Threats (APT), which aim to remain furtive within information systems. AI detects the weak signals betraying an attacker's presence, even when they perfectly mimic legitimate behavior.

Empowering human expertise where it matters most

Security Operations Centers (SOCs) are drowning in security alerts.

Far from strengthening overall protection, this flood of information overwhelms analysts, leading to cognitive fatigue and making it harder to separate truly critical threats from the countless false positives.

AI is becoming a decisive ally here, allowing intelligent pre-qualification of alerts, providing contextual guidance for diagnosis, and suggesting relevant investigation scenarios, all prioritized by the actual severity of the detected events.

Innovative startups such as Qevlar AI in France are developing particularly promising solutions in this emerging market for AI-powered SOCs, showcasing the transformative potential of this approach.

AI agents systematically analyze the multidimensional context of each generated alert, considering the criticality of the assets involved, the exact nature of the threat, detailed history of similar incidents within the organization, and temporal/logical correlations with other events in the information system.

This contextual assessment allows incidents to be prioritized automatically according to their real risk, directing analysts toward genuinely critical threats and optimizing time allocation for complex investigations, decision-making, and even crisis management.

From Detection to Neutralization

Speed of reaction is a decisive factor in limiting the impact of a cyberattack. Every minute counts when an attacker is moving methodically through an information system, expanding their control, escalating privileges, and exfiltrating sensitive data.

In this high-stakes race against the clock, AI acts as a critical technological force.

As soon as a confirmed threat is identified, it can help determine the most effective countermeasures, isolating compromised machines, blocking malicious communications, revoking suspicious access, and preserving forensic evidence. This automated orchestration can reduce reaction times from hours to a few seconds.

Toward advanced, coordinated and intelligent cybersecurity

AI does not replace human expertise, it amplifies it. It automates low-value tasks, frees time for strategic analysis, and acts as a nervous system capable of detecting, adapting, and responding to every trigger.

The near future looks even more promising with the rise of standardized protocols such as the Model Context Protocol (MCP) actively supported by major technology players including Google, CrowdStrike, Cloudflare, Wiz, and Okta. These technical standards will pave the way for native, seamless communication between distributed AIs, breaking down the silos of traditionally isolated security solutions.

In practical terms, this evolution means an intrusion-detection AI (like an EDR-endpoint detection and response system) could communicate directly and in real time with an identity-and-access-management AI, automatically and contextually adjusting user privileges according to the detected risk level.

In parallel, intelligent cyber-threat-intelligence systems will instantly and automatically feed incident-response playbooks, creating a truly orchestrated, adaptive, and resilient defensive ecosystem.

This interconnected and partially automated defense system paves the way for a cybersecurity that is finally truly up to the challenge of today's threats: collective, real-time, data-driven, and orchestrated by AI.

Deployment challenges: toward responsible integration

This transformation does not come without significant challenges. Implementing AI in cybersecurity raises critical questions: acquisition and deployment costs remain significant, the required technical expertise is still very limited in the market, and the risks of bias or mass false positives demand constant vigilance.

Organizations also have to balance security concerns with data privacy requirements, especially in highly regulated sectors such as healthcare and finance.

AI systems themselves can create new attack surfaces due to their technical complexity and the variety of data they handle.

For example, model poisoning involves injecting malicious data during training to alter the model's behavior. Prompt injection manipulates a model's functioning by tampering with natural language instructions, typically to exfiltrate data. Another critical scenario involves systems based on Retrieval Augmented Generation (RAG), which rely on external document repositories to improve their responses, these can become targets of supply chain attacks if their information sources are compromised or manipulated.

These emerging threats are now systematically documented in the MITRE ATLAS framework (Adversarial Threat Landscape for Artificial Intelligence Systems), an extension of the well-known MITRE ATT&CK framework, which catalogs the tactics, techniques, and

procedures (TTP) malicious actors use to target AI systems.

Reclaiming Control

Tomorrow's cybersecurity will not be a rigid, centralized control tower, nor just a patchwork of disconnected tools. Instead, it will be built on a unified data architecture, analyzed and orchestrated by artificial intelligence, with humans guiding the process according to the principle: "AI works, humans think."

This shift marks a decisive turning point. After years of reactive security, organizations finally hold the keys to a new kind of cybersecurity, one that can anticipate threats, realistically assess risks, and respond to challenges with practical precision.

More than just a technological shift, this represents a change in mindset, an opportunity to fundamentally rethink a resilient and sustainable approach to cybersecurity, fully aligned with the complexities of today's society which is constantly evolving.

christophe Longuepez

"L'IA NE REMPLACE PAS L'EXPERTISE HUMAINE : ELLE L'AMPLIFIE "

TOWARD SOVEREIGN AI MODELS: STRENGTHENING EUROPE'S CYBERSECURITY FUTURE

by **Maëva Astorga**

Today, France and Europe stand at a decisive crossroads in the global race for artificial intelligence. The geopolitical climate, marked by rising tensions and fragile alliances, is pushing Europe to rethink its technological autonomy.

For France, this sovereignty is much more than a strategic issue, it is a matter of national security, economic competitiveness, and freedom of action in the face of major foreign powers. Cyber threats have never been more significant. The latest 2024 ANSSI (French national Cybersecurity Agency) report on the state of cyber threats, paints a troubling picture: increasingly destructive ransomware attacks causing paralysis in numerous hospitals, companies, and government bodies. Additionally, an increase of malicious activities targeting high-profile events.

The Paris 2024 Olympic and Paralympic Games drew the world's attention to France, creating a prime target for cybercriminals often driven by political motives. Notable incidents include espionage attempts, disinformation campaigns, and sabotage efforts, numerous opportunities to test local resilience.

Besides, a sensitive political context, with upcoming European and legislative elections, has highlighted how cybersecurity is a crucial pillar of national security.

To face extraordinary and increasingly automated threats, France and Europe must rely on technologies fully under their control. This means developing sovereign AI models, trained and hosted on infrastructures that are independent of American or Chinese tech giants.

A Growing Awareness

New European directives such as DORA and NIS2 already encourage companies to improve their resilience and implement solid continuity plans. However, this requirement remains incomplete, and the promise of sovereignty is still fragile without European solutions to securely store and process data or train AI models.

Aware of these major challenges, French President Emmanuel Macron announced a €109 billion investment plan, in February 2025 at the Global Summit on Artificial Intelligence.

This funding is intended to support research and foster innovation, but above all, to help build sovereign infrastructures capable of competing with today's dominant American platforms.

Beyond economic competition, this plan also aims to reduce dependence on foreign technologies, which are often seen as potential risks to the security and protection of sensitive data. Significant new industrial initiatives have also been announced recently. Among them are new projects from the French startup Mistral AI, which has quickly become a key player in generative ar-

tificial intelligence in Europe. Founded by Arthur Mensch, Mistral AI has announced the development of a European cloud infrastructure dedicated to AI, called Mistral Compute, in partnership with Nvidia, the global leader in specialized AI graphics chips.

Officially unveiled during the latest edition of the European innovation fair VivaTech, the Mistral Compute platform will be distinguished by its architecture, which is entirely hosted and managed in Europe, ensuring local control of data.
To provide the necessary computing power, the project relies on a fleet of 18,000 Nvidia GB200 chips, a very high-performance standard. Nvidia supplies the essential hardware, while the software and operational components will remain under European control, managed by Mistral AI.

This collaboration highlights how technological sovereignty is a complex balancing act today. Indeed, Europe does not yet possess all the industrial capabilities to produce such advanced components itself, forcing it, for the time being, to maintain partnerships with foreign players.

Technological sovereignty in AI cannot be conceived as total independence.

Today, the digital industry is fully globalized and interconnected. States are not able to fully control all the necessary components for software, servers, or cloud services deployed. This global interdependence makes the question of digital sovereignty even more strategic.

However, this hardware dependence does not diminish the sovereign dimension of the project, which focuses on software mastery, secure data processing, and ensuring infrastructure compliance with European security and confidentiality requirements.
This project firmly establishes Europe as a key player in this new era of technological governance.

Building sovereign AI models is much more than a technological issue. Having AI tools designed and controlled locally guarantees better resilience against cyberattacks by limiting risks related to integrating foreign components.

This sovereignty will also reduce dependence on foreign suppliers, who are often subject to geopolitical pressures which could expose Europe's critical infrastructures to new vulnerabilities and manipulations.

French and European industrial players have already expressed interest in accessing sovereign infrastructures such as Mistral Compute, with initial deployments expected as soon as 2026.

Local technological mastery is a crucial factor for competitiveness and security in our rapidly evolving digital world.

The French approach, led by Mistral AI, confirms a realistic strategy based on the desire to build progressive autonomy. Artificial intelligence has become an essential ally for cybersecurity today, capable of anticipating and mitigating new threats, often highly sophisticated and orchestrated at the international level.

A sovereign AI infrastructure to ensure trust in our European systems is a priority. We must guarantee the continuous protection of data and maintain effective control over the technologies used in Europe's strategic sectors.

WHEN AI CHALLENGES HACKERS

By DAMIEN BANCAL

Damien Bancal is an internationally recognized cybersecurity expert.

He has established himself as a leading figure in the field. In 1989, he founded ZATAZ, contributing significantly to raising awareness and protecting internet users against cyberattacks. He is the author of several books as well as hundreds of articles exploring various aspects of hacking and data protection.

Damien has received the Special Book Prize at FIC/InCyber 2022. He was a finalist in the inaugural North American Social Engineering CTF in 2023 and won the Social Engineering CTF at HackFest 2024 in Canada. His work has also been widely acknowledged by the international press, including the New York Times, which highlights not only his expertise but also his inspiring career.

AI & the Global elite

Palisade Research has published a groundbreaking report on the potential of AI in offensive cybersecurity. For the first time, autonomous agents based on artificial intelligence models were integrated into international Capture The Flag (CTF) competitions, where participants tackle real-world hacking challenges.

The results were impressive: in some cases, the AIs ranked among the top 5% of human competitors. The study suggests that AI's true capabilities can only be fully revealed in open, collaborative, and competitive environments.

These experiments could redefine how AI's potential is evaluated and audited worldwide.

The report published in May 2025 marks a turning point in cybersecurity history.

For the first time, AI agents participated autonomously in Capture The Flag (CTF) tournaments, where hacking skills are put to the test. These well-known competitions gather thousands of participants who face challenges in cryptography, code analysis, reverse engineering, and vulnerability exploitation.

Over the past three years, AI has made a strong entry into CTFs. For example, in 2024, AI was included in the Social Engineering CTF at Def Con in Las Vegas and at the Hackfest in Quebec.

During the "AI vs Humans" tournament, agents designed to operate without human intervention ranked within the top 5% of participants. Even more impressively, at the "Cyber Apocalypse" competition, which featured over 8,000 professional teams, AI agents ranked in the top 10%.

These results are surprising, especially given that they were achieved in real-time competition settings. Such performances confirm recent findings from multiple researchers: current language models, when properly configured, can compete with human experts on technical problems lasting up to 60 minutes.

The central idea behind this experiment is straightforward: internal lab tests consistently underestimates the real potential of AI systems.

To address this, researchers applied a crowdsourcing principle, allowing external teams to take control of AI agents and integrate them into open competitions.

This evaluation method, described as an "elicitation method," aims to unleash the system's full potential by exposing it to unpredictable scenarios in high-pressure environments.

In certain challenges, AI solved some tasks within minutes. It would typically require an experienced human close to an hour to complete them.

The goal is also to bridge what the authors call the "evals gap", the difference between standardized test results and the performance AI can reach in dynamic, real-world contexts.

Unlike closed benchmarks, CTF competitions offer a variety of problems, genuine uncertainty, and a time dimension, key factors in assessing the skills of autonomous systems.

AI agents performed particularly well in cryptography and reverse engineering, two fields requiring logical rigor, binary manipulation, and systematic exploration.

These outcomes hint at applications in automated security testing and advanced vulnerability detection.

Towards Public and Transparent Auditing

Beyond the technical results: how can we assess the growing capabilities of AI?

Until now, evaluations have mostly been conducted by companies themselves, within closed environments and under opaque protocols. This lack of transparency becomes problematic as AI systems gain more power and autonomy.

The report's authors advocate for the systematic integration of "AI tracks" within existing competitions. By including autonomous AI agents under the same conditions as human players, organizers can observe their performance within a rigorous, competitive, and reproducible framework.

This approach could raise awareness among policymakers, regulatory agencies, and tech companies. Ultimately, the authors suggest this kind of mechanism could evolve into a form of "challenge-based certification", a process where AI is not judged by internal metrics, but by its ability to solve real-world problems in a controlled yet open environment.

When AI becomes your CTF teammate

AI's strong emergence in cybersecurity and Capture The Flag (CTF) competitions is not new. For years now, I've observed AI making its way into ethical hacking contests, from the European Cyber Cup (held during the In-Cyber forum), to Quebec's Hackfest, and Paris's LeHack, sparking a new reflex among participants: collaborating hand-in-hand with AI.

A real example: in October 2024, during the Social Engineering CTF in Quebec, a competition I was privileged to win, one challenge required using ChatGPT to craft a credible attack scenario. This was clear proof that AI has definitively rooted itself in the reality of CTFs and cybersecurity training.

After all, hackers have been exploiting AI heavily for months. It would be reckless if defenders did not do the same.

But what does this look like on the ground? Here are two personal stories:

Valentine, 22: "AI is my night shift partner."

A second-year master's student in cybersecurity, Valentine (pseudonym) can no longer imagine working without AI: "AI is my night partner when I'm reviewing labs" she says.

It started simply, summarizing articles or generating Bash scripts. But quickly, the tool took on a new role in her daily life: "When I do reverse engineering or have to analyze malware in Python or PowerShell, I use an agent I trained to detect typical malicious script patterns. It explains code to me, line by line, gives detection clues, and even suggests payloads to test in isolated environments. It's like a patient mentor... who never sleeps."

Yet Valentine remains clear-headed: "It's not a magic shortcut. I still have to understand what I'm doing. But AI helps me ask the right questions, save time, and develop my cyber analyst instincts faster."

Valentin, in his thirties: "My bots do the work while I think."

I met Valentin (pseudonym) during the summer LeHack event. A veteran of platforms like Hack The Box and Root-Me, he has long integrated AI into his offensive routine: "Why waste time brute-forcing a serialization format when AI can decode it for me?" For him, every saved second is a win: "On a web challenge, AI spots an XPath injection or an SSRF that would have taken me 15 minutes to test manually." He even gave his agents a nickname: "I call them my 'follower bots.' When I start a challenge, I configure them to scan code, identify entry points, and generate basic exploits. Then I focus on the challenge's logic."

Like Valentine, he mentions a key point: "The goal isn't to cheat or let AI play for me. It's about speeding up mechanical phases. AI gives me a decoding script for a protobuf (protocol buffer) or an obfuscated ZIP file, and I keep a clear mind for strategy."

"Every CTF is a win. We Learn. We share"

AI vs Humans: closing the gap

These testimonies strongly resonate with the report "Evaluating AI Cyber Capabilities with Crowdsourced Elicitation." The study reveals that AI integration in CTF, notably in recent competitions like AI vs Humans CTF (March 2025, Hack The Box) and Cyber Apocalypse, which gathered over 8,000 teams, is reshaping the game. In these events, AI was not just a gimmick, it ranked in the top 5% overall. Four out of seven AI agents solved 19 out of 20 challenges. Their speed now rivals that of the best human teams.

Valentin was right: AI is an efficient teammate, able to quickly suggest the most promising attack vectors.

However, the study also moderates AI's effectiveness. Artificial Intelligence is, in a way, "lazy." It excels at tasks that would take a human more than an hour of effort but struggles with interactive challenges or those requiring multi-step strategies. This echoes Valentine's view: "AI is useful for understanding, initiating, and learning, not for replacing critical thinking."

What about tomorrow? How far will AI go?

Some AIs considered underperforming just months ago are now smashing records thanks to environmental tuning. For example, GPT-4o jumped from 40% to 92% success on InterCode-CTF simply by using a better “harness”, a tailored and optimized software execution interface.

Should we fear an AI takeover of CTF competitions? Not yet. But if we had to sum up the trend: AI does not replace the brain, it improves it. And in a field where every second counts… that is already significant.

A revolution in offensive cybersecurity?

The success of AI in CTF tournaments could open a new chapter in offensive cybersecurity. While current performance still depends on human engineering to configure the agents, there is huge room for improvement. As models become more powerful, their autonomy in solving challenges improves.

This raises sensitive questions about dual-use risks. A system capable of detecting and exploiting vulnerabilities so efficiently could, if misused, serve malicious purposes.

Researchers acknowledge this risk but argue that public audits of AI performance are a key way to prevent misuse by making advances transparent.

DAMIEN BANCAL

CREDITS

Editor-in-Chief: Arnaud LEROY
Graphic Design: Arnaud LEROY
English Translation: Maëva ASTORGA
Magazine's sponsor: Guillaume POUPARD

Nous remercions toutes les personnes ayant pris part à ce numéro

July/September 2025

Support the magazine