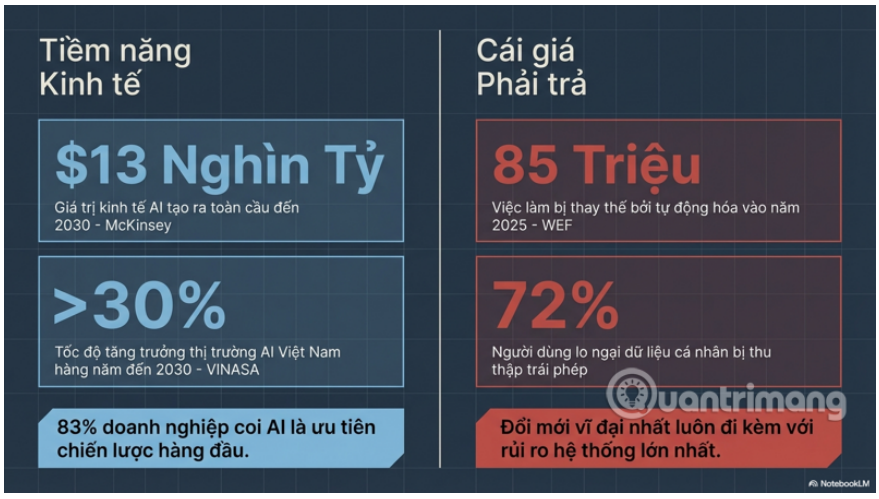
Risks of Artificial Intelligence (AI): A Comprehensive Analysis in the Context of Modern Technology

Artificial intelligence offers great potential but also carries significant risks regarding employment, privacy, security, and ethics, requiring careful control.

1. Introduction: Artificial Intelligence in Modern Life

In less than a decade, artificial intelligence (AI) has transformed from an academic concept confined to laboratories into a dominant force in every aspect of life. Virtual assistants like Siri and Google Assistant, and large-scale language models like ChatGPT and Claude, are present in millions of devices every day. AI analyzes medical images to detect cancer earlier than doctors, coordinates city traffic, reviews bank loan applications, drafts legal contracts, and even composes music and paintings.

In Vietnam, the digital transformation wave is drawing AI into every sector – from smart agriculture and e-commerce to online education and public services. According to reports from VINASA and the Ministry of Information and Communications, the Vietnamese AI market is projected to grow by over 30% annually between 2024 and 2030. This figure reflects high expectations, but also raises a pressing question: what are the associated risks, and are we prepared to address them?



This article does not aim to deny the enormous benefits of AI, but rather approaches it from a systematic risk analysis perspective – because only by understanding the potential risks can we responsibly utilize the full potential of this technology.

2. Overview of the main risk groups of AI

Risks from AI don't originate from a single source, but rather span multiple dimensions. They can be categorized into the following main groups:

Economic and labor risks

High level

Widespread automation, income polarization, and the disappearance of traditional occupations.

Privacy and data security risks

High level

Comprehensive surveillance, exploitation of personal data, and violations of human rights.

Cybersecurity and military risks

High level

Autonomous weapons, AI-accelerated cyberattacks, information warfare.

Risks of bias and discrimination

Average

The model learns from skewed data, amplifying social inequality.

Moral and legal risks

Average

Accountability loopholes, lack of a legal framework, and a black box in the decision-making process.

Risk of dependence and loss of control

Average

Humans are losing their ability to make independent judgments, posing a risk to super-intelligent AI systems.

Environmental risks

The level is increasing.

Data centers and large-scale models consume enormous amounts of energy.

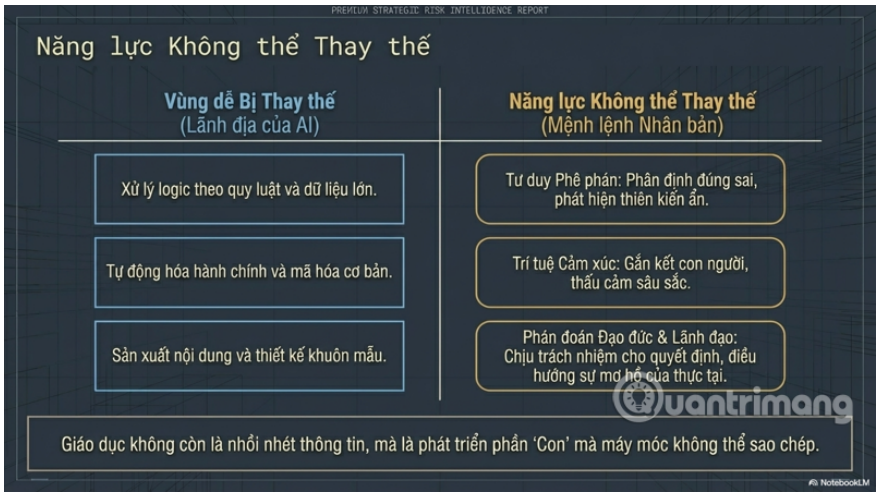
3. Labor market risks and job losses

3.1. Automation and the wave of labor replacement

This is the most discussed and also the most readily apparent risk in everyday life. The history of industrial revolutions has shown that technology always creates new jobs while destroying old ones. However, AI is different in one core respect: it not only replaces manual labor but also knowledge labor.

Occupations at highest risk include: accounting and auditing (as AI software processes millions of transactions in seconds), customer service and call center staff (with increasingly sophisticated AI chatbots), translation and interpretation (with multilingual models), data entry and administrative document processing, basic data analysis, and even some positions in journalism, graphic design, and programming.

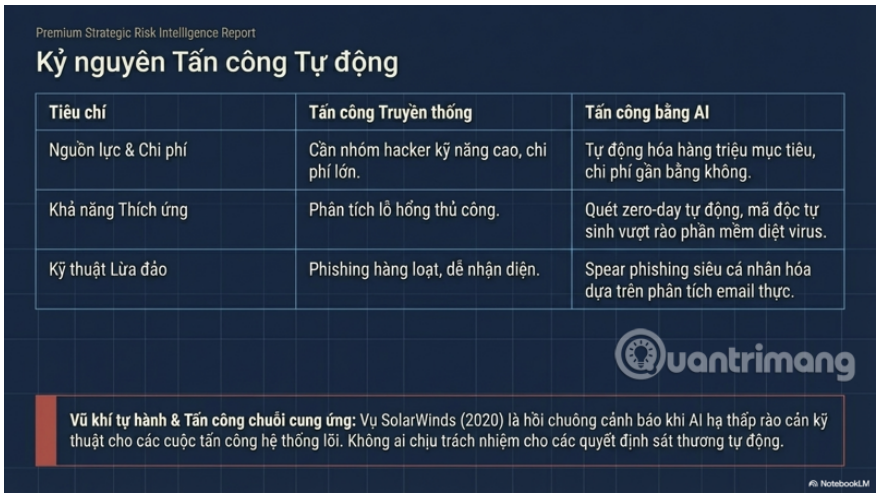
According to a Goldman Sachs study (2023), approximately 300 million global jobs could be affected by AI-generated tools. Of these, two-thirds will be partially automated, and one-quarter could be completely replaced.



3.2. Income Divergence and Economic Inequality

The risks lie not only in job losses, but also in the widening gap between those capable of mastering and applying AI and those without access to digital skills training. Economists call this trend "skill polarization"—the labor market is stretched to two ends: on one side, highly skilled AI jobs with exceptional pay, and on the other, simple manual labor jobs that cannot be automated, while the middle ground—the middle class—shrinks significantly.

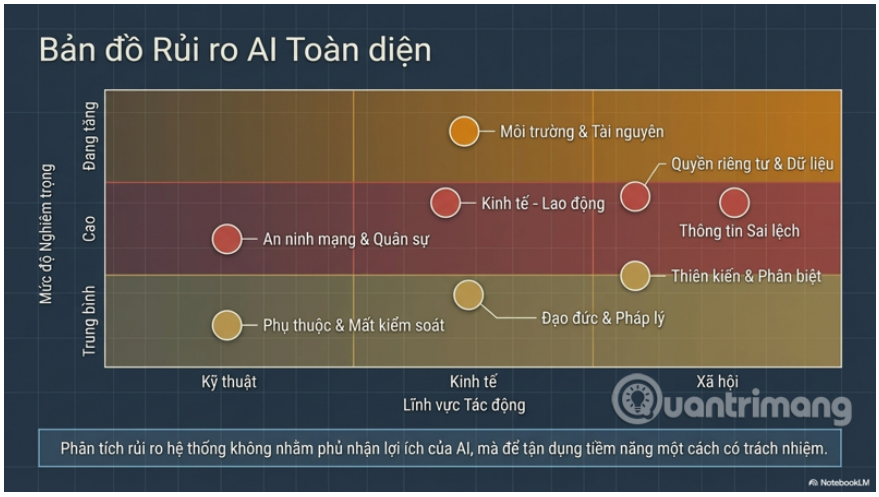
In developing countries like Vietnam, this risk is particularly serious as the exported workforce in the textile, electronics assembly, and business process outsourcing (BPO) industries could be significantly impacted by robots and AI in the next decade.



3.3. Jobs that AI is currently replacing and will replace in the future.

Beyond the aforementioned sectors, AI is also penetrating fields previously thought to be "immune" to automation. In healthcare, AI has achieved accuracy equal to or surpassing that of doctors in analyzing X-ray and MRI images. In law, AI systems can research case law, draft contracts, and predict litigation outcomes. In education, AI-personalized learning paths can partially replace the role of tutors. In media and marketing, automated content generation tools are directly challenging copywriters, content creators, and designers.

4. Risks related to privacy and data security.



4.1. AI as a comprehensive surveillance machine

To function effectively, AI needs data – and the more data, the smarter the model. This is the root of the global privacy crisis. Facial recognition systems deployed at airports, train stations, and on streets in many countries are capable of tracking the movements of millions of people without their explicit consent. AI-powered cameras not only record images but also analyze behavior, emotions, and even predict the intentions of the person being observed.

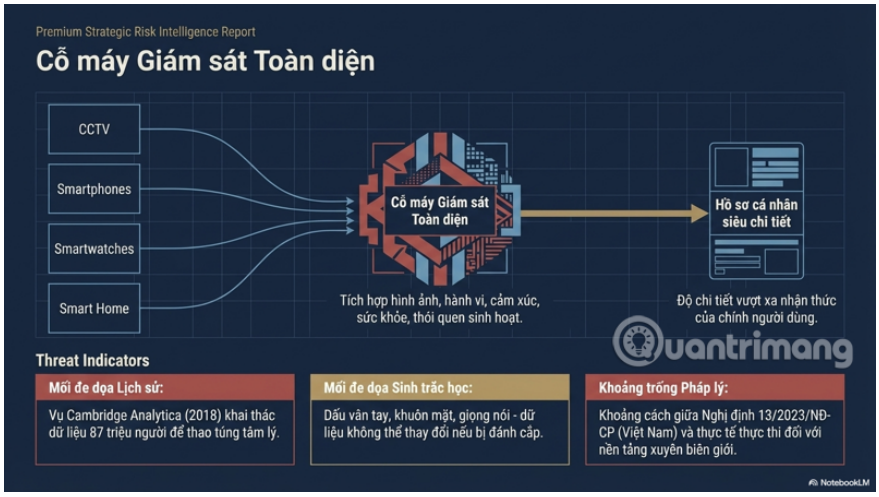
Smartphones, smartwatches, smart home devices – each device is a continuous data collector of users' habits, health, social networks, and political views. When this data is integrated and analyzed by AI, a personal profile is created with a level of detail that even the user themselves may not be aware of.

The Cambridge Analytica scandal (2018) served as the first wake-up call: the data of 87 million Facebook users was exploited to build AI models targeting campaign psychology. Since then, the scale and sophistication of data exploitation have increased exponentially.

4.2. Data Leaks and AI Attacks

Companies and governments collect vast amounts of data to train AI, which means creating incredibly lucrative data repositories for hackers. A single leak from a large company could expose sensitive information on hundreds of millions of people. Particularly concerning is biometric data (fingerprints, facial recognition, voice) – unlike passwords, it cannot be altered if stolen.

Furthermore, AI is also being used to attack security systems: automating vulnerability scanning, creating adaptive malware, and personalizing spear phishing attacks with frightening precision. A phishing email crafted by AI, based on analysis of hundreds of victims' real emails, is almost indistinguishable from a genuine email.



4.3. Lack of a legal framework to protect users.

While the EU already has GDPR and is implementing the EU AI Act, most countries – including Vietnam – are in the process of building and perfecting their legal frameworks for AI. Decree 13/2023/ND-CP on the protection of personal data in Vietnam is an important step forward, but there remains a significant gap between legal regulations and practical implementation, especially when dealing with cross-border platforms.

5. Cybersecurity risks and the threat of AI weaponization.

5.1. AI in Information Warfare

Cyberspace has become the fifth battlefield after land, sea, air, and space. AI is fundamentally changing the nature of cyber conflict. Instead of needing thousands of highly skilled hackers , a small group can deploy automated AI systems to attack millions of targets simultaneously. The ability to mass-produce fake content – ?? from articles and social media accounts to videos – allows for the deployment of opinion manipulation campaigns at near-zero cost.

Attacks on critical infrastructure such as power grids, water supply systems, hospitals, and banks are becoming increasingly sophisticated thanks to AI. Attackers can use AI to find and exploit zero-day vulnerabilities faster than any human security team can patch them.

5.2. Autonomous weapons and control risks

Autonomous drones, combat robots, and AI-guided smart missiles are being developed by many major powers. The risks lie not only in the possibility of these weapons being attacked and hijacked by the enemy, but also in the more fundamental question: who is responsible when an autonomous weapon system mistakenly kills civilians? No soldier gave the order, and no current legal mechanism is sufficient to answer this question.

Over 3,000 AI and robotics researchers have signed a pledge not to develop lethal autonomous weapons, but this remains a voluntary act while the AI ??arms race continues at the national level.

5.3. Self-generating malware and supply chain attacks

AI-generated code can be abused to write malware, creating new variants of ransomware that constantly bypass antivirus software. Software supply chain attacks – injecting malicious code into widely used open-source libraries – become easier and more common when attackers have AI support. The SolarWinds attack (2020) served as a warning about the devastating impact a supply chain attack can have – and AI is lowering the technical threshold required to carry out similar attacks.

6. Risks of bias, discrimination, and inequality

6.1. Data bias, model bias

A fundamental principle in AI is "garbage in, garbage out"—bad training data will produce a bad model. But even more dangerous is when the data reflects historical and social biases that exist in the real world. If historical data shows that women are less likely to be promoted than men in a particular industry, and the AI ??model learns from that to select personnel, it will reproduce and even reinforce gender inequality.

Facial recognition systems have been shown to have significantly higher error rates for women of color compared to white men, due to the lack of representative training data. AI-powered credit analysis systems can inadvertently discriminate geographically – rejecting poorer areas – even without directly inputting race or income data.

6.2. Bias in Justice and Healthcare

In the U.S. justice system, the COMPAS software is used to predict the likelihood of recidivism to aid in decisions about acquittal or imprisonment. A ProPublica study (2016) found that the software mispredicted a high rate of recidivism for African Americans compared to white Americans. This isn't a technical error – it's a systemic bias that's AI-powered and legitimized by the technology's apparent "objectivity."

In healthcare, AI models trained primarily on Western data may provide less accurate diagnoses and treatment recommendations for Asian, African, or patients with specific health conditions – exacerbating healthcare inequalities.

7. Risks of misinformation – Deepfakes and industrial-scale fake news

7.1. Deepfake: When you can't believe your eyes.

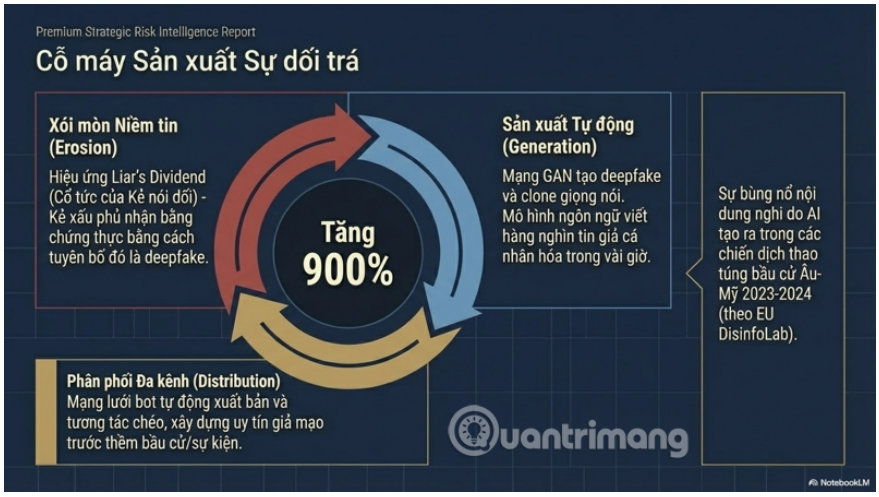
Deepfake technology uses geointeractive neural networks (GANs) to create increasingly indistinguishable fake images, videos, and audio. In 2023–2024, a series of fake videos of world leaders announcing unprecedented policies circulated on social media before being removed. Voice cloning technology allows for the replication of voices from just a few seconds of audio samples, creating fraudulent calls impersonating relatives or business leaders.

The most dangerous threat is deepfake in the context of elections. Fake videos of candidates saying things they would never say, disseminated in the final hours before voting day, can have serious consequences without sufficient time for verification and correction.

7.2. Mass Production of Fake News

Before AI generation, producing fake news required time, manpower, and funding. Today, a large-scale language model can generate thousands of fake news articles personalized to the geography, demographics, and psychopaths of a target audience in just hours. Fake news websites, fully automated by AI, can publish hundreds of articles daily, building a credible history before launching a disinformation campaign.

The EU DisinfoLab report notes a 900% surge in the amount of content believed to be AI-generated in disinformation campaigns surrounding major elections in Europe and the US between 2023 and 2024.



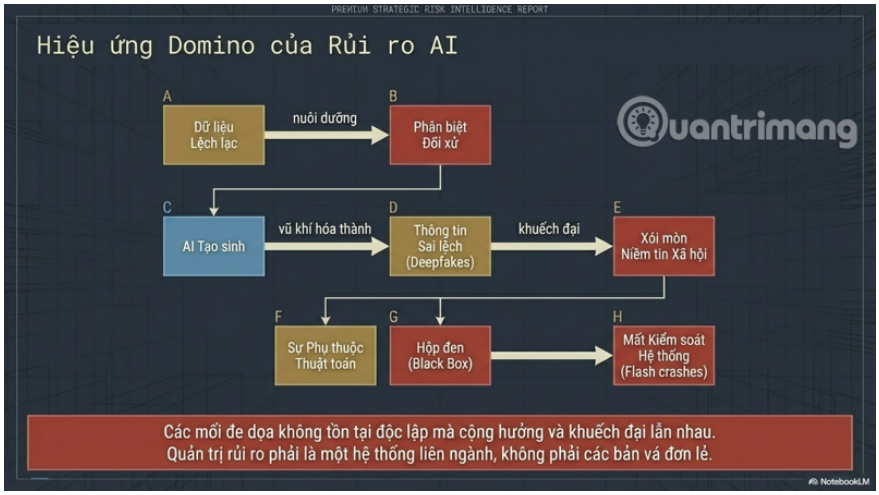
7.3. The "Liar's Dividend" Effect – Exploiting Suspicion

A less-noticed but equally dangerous side effect: the existence of deepfakes allows malicious actors to negate genuine evidence by claiming it is deepfake. As people increasingly doubt the authenticity of all videos and audio, trust in the media and institutions crumbles – precisely what information manipulators aim to achieve.

8. Moral hazard and accountability

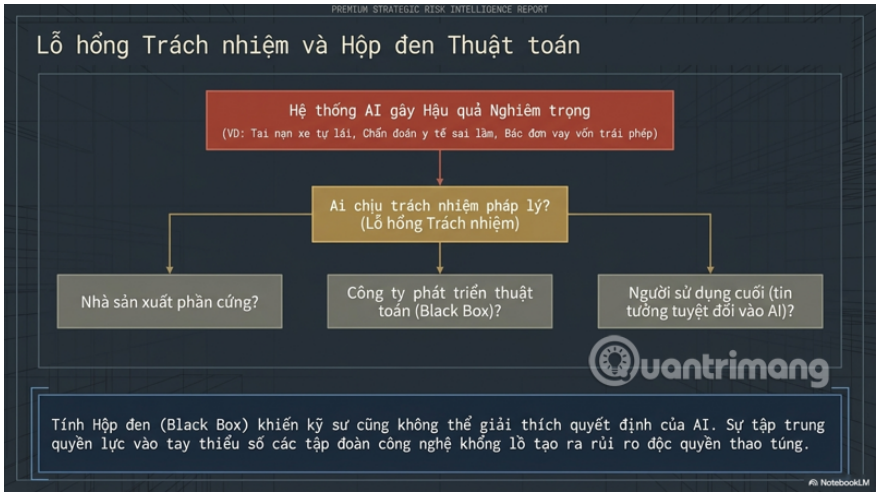
8.1. The black box of decisions and lack of transparency.

One of the core ethical challenges of modern AI is its "black box" nature – many deep learning models make decisions that even the engineers who created them cannot fully explain. When AI rejects your loan application, refuses your child's college admissions, or suggests a sentence for you – and cannot explain why in a reasonable way – this is a serious issue concerning human rights and social justice.



8.2. Gaps in Responsibility

When a self-driving car causes an accident, who is responsible: the car manufacturer, the AI ??software development company, the passengers, or the car itself? When AI in a hospital makes a wrong diagnosis and doctors trust and follow it, who is legally responsible? The current legal system is built on the assumption of intentional human behavior – there is no clear mechanism to hold an autonomous system accountable.



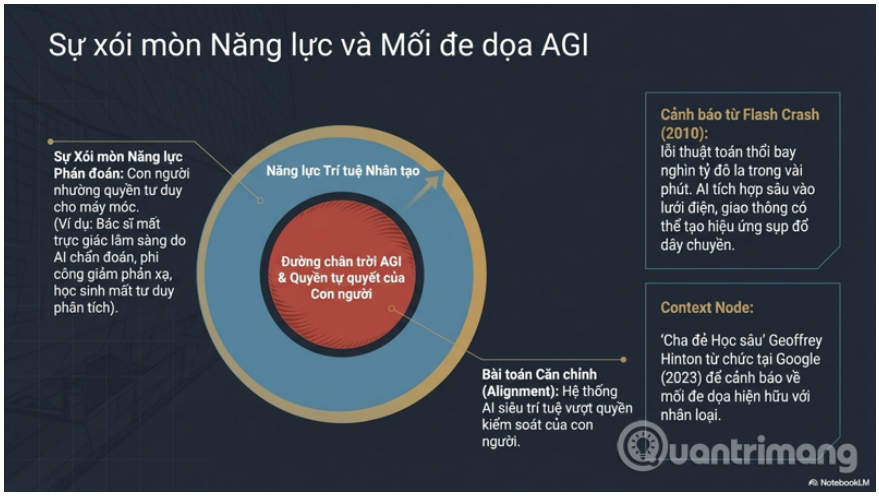
8.3. Concentration of Power and the Risk of AI Monopolies

AI is not a democratic technology. The cost of training the most advanced large-scale language models has exceeded hundreds of millions of dollars, making it only achievable by a select few giant tech companies or the governments of major powers. This creates the risk of concentrating AI power in the hands of a minority – those controlling AI will have disproportionate economic, intelligence, and influence advantages over the rest of society and smaller nations.

9. Risks of dependence and loss of control

9.1. Erosion of human judgment

As AI takes on more and more cognitive tasks—from planning and problem-solving to creativity—a subtle but profound risk is the erosion of human capacity for independent thinking. Modern pilots practice less manual flight due to autopilot systems, leading to reduced emergency response capabilities. Doctors relying on AI for diagnosis may lose their clinical intuition. Students accustomed to AI writing may lose their analytical and expressive thinking skills.



9.2. Risks from Generalized Hyperintelligence AI (AGI)

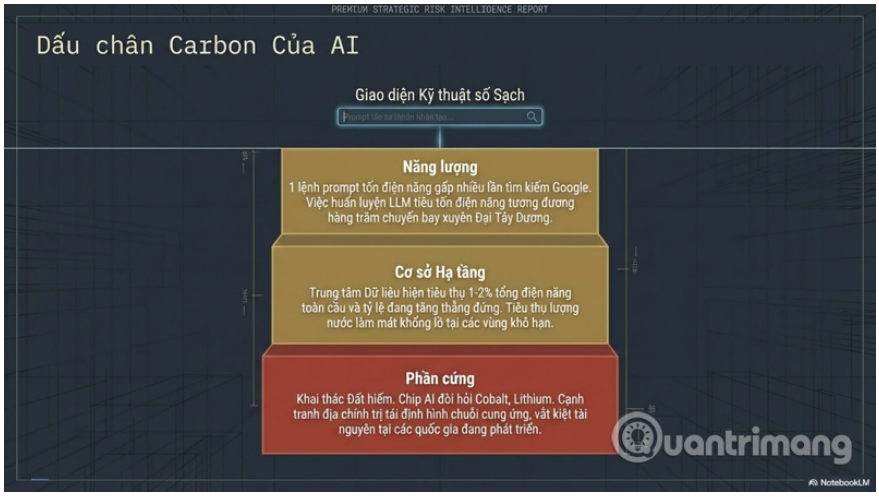
This is the most long-term and also the most debated risk. AGI – an AI with general cognitive abilities equal to or surpassing humans in all fields – could emerge within the next few decades, according to some researchers. If that happens, the question of whether we can ensure AGI acts in accordance with human values ??becomes vital. AI safety researchers at organizations like Anthropic, OpenAI, and DeepMind are working to address the "alignment" problem – aligning AI goals with the interests of humanity – but this remains an unresolved issue.

Professor Geoffrey Hinton, dubbed the "father of deep learning" and formerly of Google, resigned in 2023 to freely warn about the potential dangers of AI – including the risk that AI poses an existential threat to humanity in the future.

9.3. AI System Failure and the Chain Reaction

As AI becomes deeply integrated into critical infrastructure – smart grids, automated transportation systems, financial markets – a flaw in the AI ??model could trigger a chain reaction across the entire system. The 2010 "flash crash" in the US stock market, where automated trading algorithms caused the market to lose trillions of dollars in minutes before recovering, is a microcosm of what could happen on a much larger scale in the future.

10. Environmental and resource risks



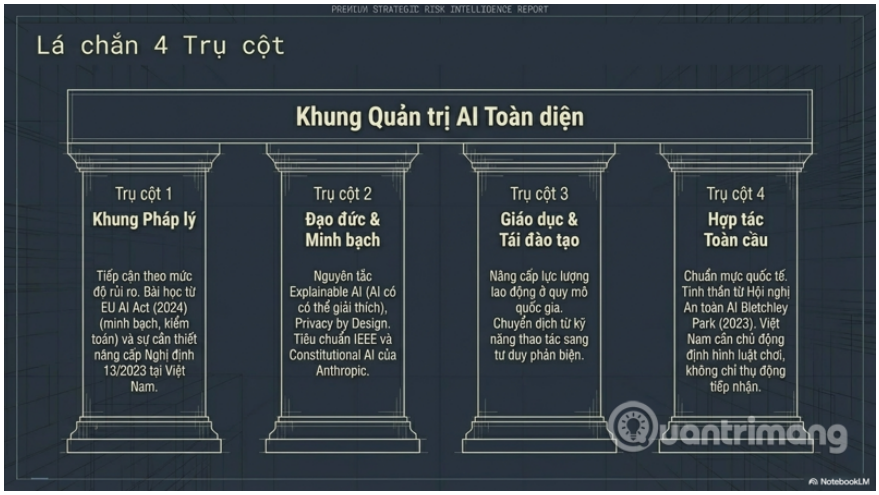
10.1. The enormous carbon footprint of AI

One risk often overlooked in discussions about AI is its environmental impact. Training a large language model consumes the same amount of electricity as hundreds of transatlantic flights. Global data centers currently consume about 1–2% of the world's electricity, and this percentage is rapidly increasing with the AI ??boom. Cooling water for data centers is also a serious issue in water-scarce regions.

10.2. Resource Exploitation and Supply Chains

AI requires hardware – AI chips demand rare earth metals like cobalt, lithium, and tantalum, much of which is mined in developing countries with worrying labor and environmental conditions. The geopolitical competition for semiconductor chips is reshaping global supply chains and creating new vulnerabilities in the digital economy.

11. AI Risk Management Framework and Solutions



11.1. Requirements for a comprehensive legal framework

The first and most crucial solution is to build a suitable legal framework that is industry-specific and adopts a risk-based approach. The EU AI Act – adopted in 2024 – is a pioneering example: it classifies AI systems according to risk levels (unacceptable, high, medium, low) and sets corresponding requirements for transparency, auditing, and accountability. Other countries need to learn from this and adapt it to their local context.

11.2. Principles of Responsible AI Development

Beyond legal regulations, the global AI community is moving towards a set of principles for responsible development: transparency (explainable AI), fairness, safety, privacy by design, and accountability. Organizations like Anthropic with its "Constitutional AI" methodology, or IEEE with its AI ethics standards, are laying the foundation for the industry.

11.3. Investing in education and retraining

Responding to labor market risks requires significant and long-term investment in education – not just digital training, but also the development of competencies that AI would struggle to replace: critical thinking, creativity, empathy, leadership, and complex problem-solving. Retraining programs for displaced workers should be prioritized at the national policy level.

11.4. International cooperation on AI governance

AI is a global challenge requiring global solutions. The first AI Safety Summit at Bletchley Park (UK, 2023) and subsequent forums mark the beginning of a necessary international collaboration process. Vietnam needs to proactively participate and contribute to shaping international standards for AI, rather than being merely a passive recipient.

12. Conclusion: Balancing potential and risk

Artificial intelligence is neither an enemy nor a savior. It is the most powerful tool humanity has ever created, and like any other powerful tool, its value depends entirely on the user and the surrounding control framework.

The eight risk groups analyzed – ranging from job loss, privacy, cybersecurity, algorithmic bias, disinformation, ethics, loss of control, to environmental impact – do not exist in isolation but are intertwined and amplify each other. Effective response requires a systemic, interdisciplinary approach and collaboration among government, business, academia, and civil society.

The most important thing is not to slow down AI, but to ensure that its development is accompanied by a commensurate development of governance capacity, social awareness, and ethical frameworks. The question is not "Is AI dangerous?" but "Are we wise enough to master AI?" – and the answer depends on our actions today.