VinciWorks

Understanding the EU AI Act

A guide to the world's first comprehensive Artificial Intelligence (AI) regulation. How will it impact your company?

Introduction

In December 2023, the EU's three branches - the European Parliament, the Council of the EU and the European Commission - reached a provisional agreement on the EU's proposed Artificial Intelligence Act (AI Act). Once it passes - it hasn't been fully approved yet - it will be the world's first comprehensive legislation on AI and could set a standard for laws enacted in other parts of the world.

The EU's AI Act will prevent operators from certain uses of the technology and will require transparency on its use. While countries worldwide are taking legislative action to try and rein in AI, the EU's AI Act will likely be a game changer. Similar to the way that the EU set a high standard for data protection with the General Data Protection Regulation (GDPR), which said anyone who wants to do business in the EU must respect EU laws, the impact of the AI Act will likely be as strong and will inspire a host of similar regulations throughout the world.

Any company with an AI system in the EU market or an AI system that impacts people in the EU will need to be compliant with the regulation. Of course, it is unclear how AI will evolve.

We are just beginning to see what this technology can do. But with this regulation it is clear that the EU is serious about the technology and wants to protect users' fundamental rights and freedoms.

After a final deal is reached on the AI Act, it will come into force by 2025. Its widespread impact means that companies need to start understanding now what will be expected of them.

And that is why we created this guide. It will help you understand what the EU AI Act is, how that will affect the global response to regulating AI and, most significantly, it will clarify what impact all this will have on your company.

The EU AI Act - A first glance

The EU's AI Act is a comprehensive new law designed to regulate the use of artificial intelligence systems in the European Union. It is one of the world's first regulatory attempts to control the use of AI as this rapidly evolving technology has incredibly broad societal and economic implications.

The AI Act will likely set a global standard for countries who want to both harness the potential benefits of AI while mitigate against its possible risks.

> **"Used wisely and widely, AI promises huge benefits to our economy and society. Therefore, I very much welcome today's political agreement by the European Parliament and the Council on the Artificial Intelligence Act. The EU's AI Act is the first-ever comprehensive legal framework on Artificial Intelligence worldwide. So, this is a historic moment. The AI Act transposes European values to a new era. A commitment we took in our political guidelines for this Commission mandate – and we delivered."**
>
> *European Commission President Ursula von der Leyen*

The AI Act reached a provisional agreement in December, 2023. Once the consolidated text is finalised in early 2024, the majority of the AI Act's provisions will apply two years after its entry into force.

The regulation aims to ensure that AI systems placed on the European market and used in the EU are both safe and respect fundamental rights and EU values. It is hoped that the regulation will stimulate investment and innovation on AI in Europe.

The Act establishes obligations for high-impact general-purpose AI (GPAI) systems that meet certain benchmarks, like risk assessments, testing and incident reports. It also mandates transparency by those systems like creating technical documents and summaries about the content used for training. Another important element of the Act is that people will have the right to launch complaints about AI systems and receive explanations about decisions on "high-risk" systems that impact their rights.

Learn more

The main elements of the agreement consists of:

1. rules on high-impact general-purpose AI models that can cause systemic risk in the future, as well as on high-risk AI systems

2. a revised system of governance with some enforcement powers at EU level

3. extension of the list of prohibitions but with the possibility to use remote biometric identification by law enforcement authorities in public spaces, subject to safeguards

4. better protection of rights through the obligation for deployers of high-risk AI systems to conduct a fundamental rights impact assessment prior to putting an AI system into use

Learn more

A brief history of AI

A timeline of the development of AI

1930s and 40s	Golden age of thinking machines in science fiction
1950	Alan Turing releases his paper, *Can Machines Think?*
1956	Dartmouth Summer Research Project on Artificial Intelligence, Logic Theorists, the first AI program, is released and John McCarthy coins the term artificial intelligence
1966	*Eliza*, an AI program that can conduct simple conversations, is developed
1970	Marvin Minsky, the father of AI, states: "In three to eight years we will have a machine with the general intelligence of an average human being."
1980	Edward Feigenbaum creates expert systems
1986	Carnegie Mellon builds Navlab, the first autonomous car
1997	World chess champion Gary Kasparov defeated by IBM's Deep Blue, a chess playing computer program and Dragon Systems releases its speech recognition software
2000	Honda releases ASIMO, an artificially intelligent humanoid robot
2002	i-Robot releases Roomba, an autonomous robot vacuum cleaner
2011	Apple releases Siri, a virtual assistant that uses a natural-language user interface (Amazon Alexa and Google are released within the next few years)
2022	ChatGPT is released
2023	The EU AI Act receives provisional approval

Learn more

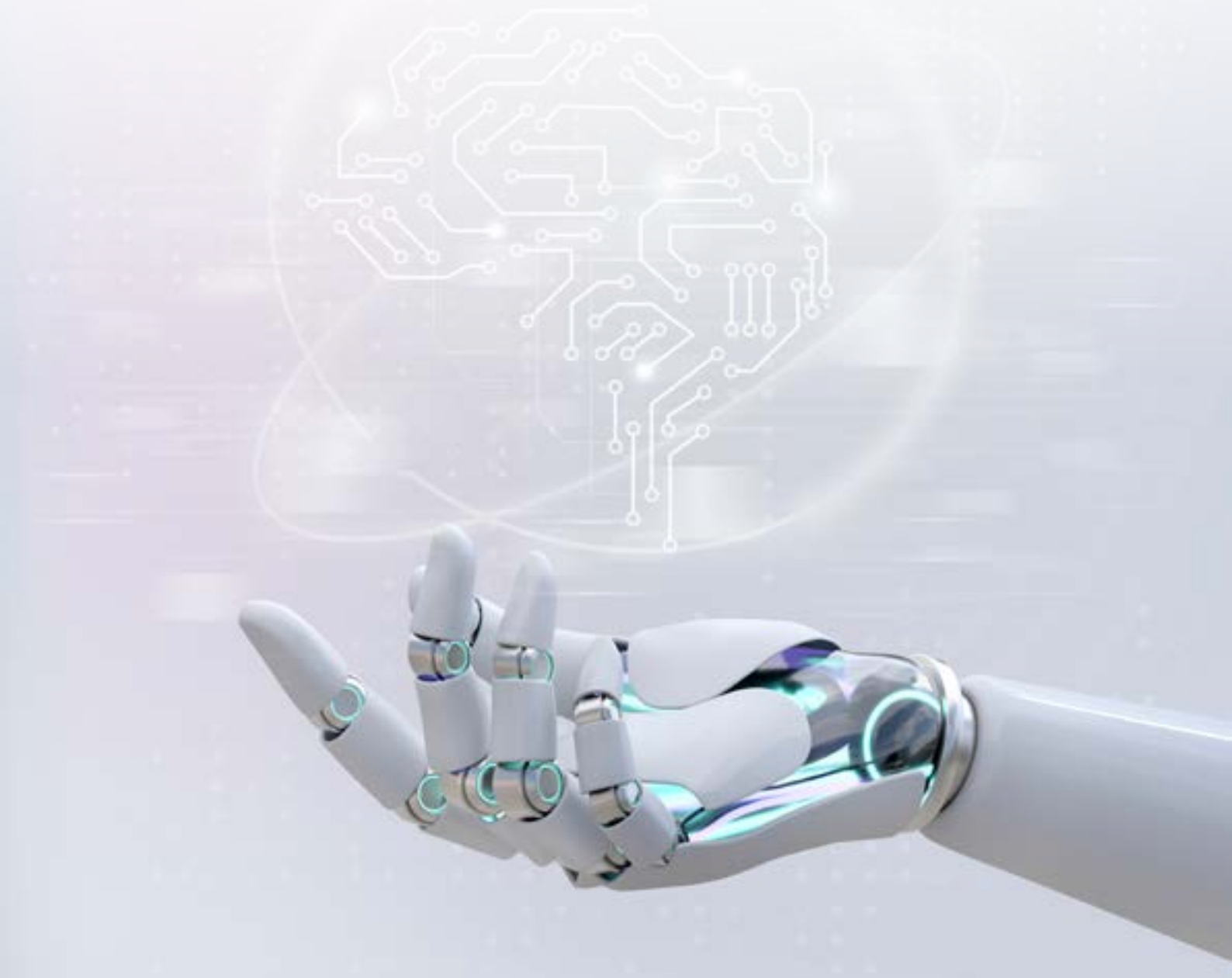
Where did AI come from?

In 1955, John McCarthy, one of the original creators of AI, defined AI as:

"The science and engineering of making intelligent machines."

In the AI Act, artificial intelligence is defined as:

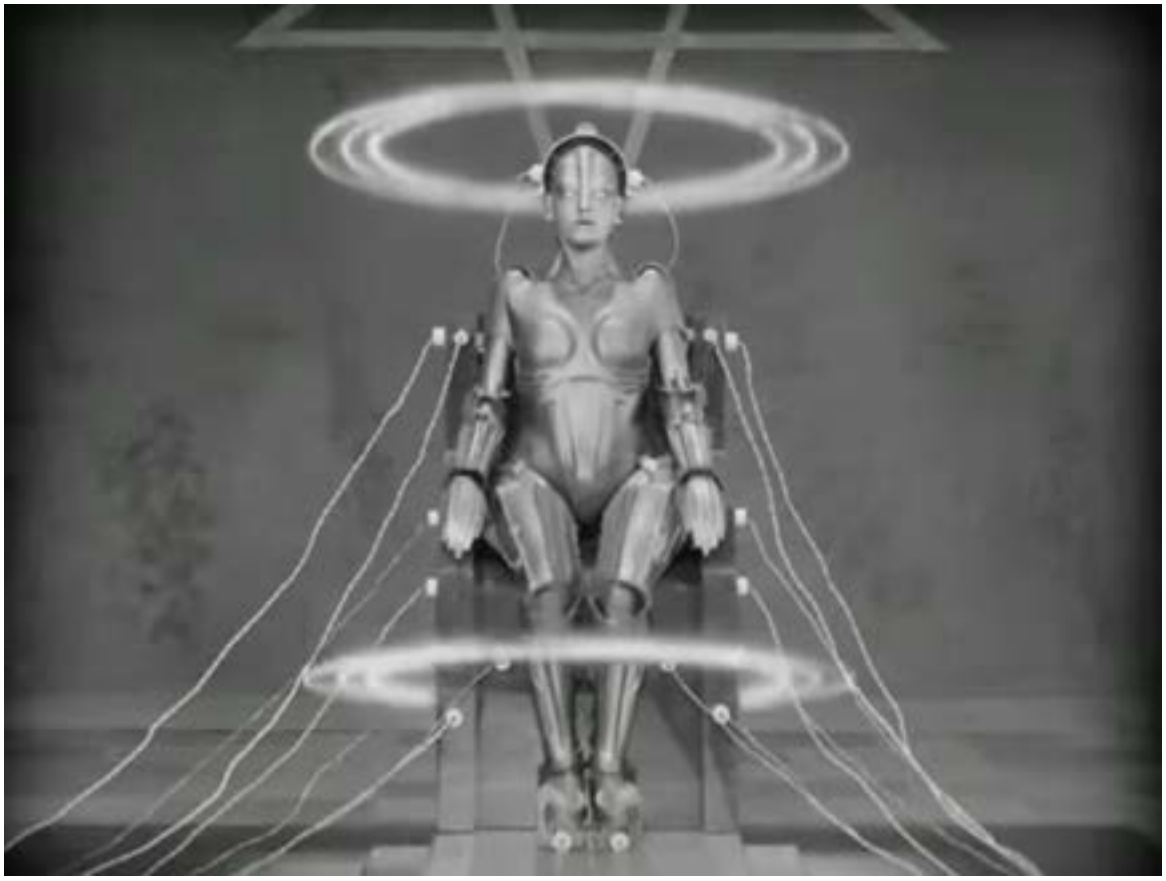
"a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments."



Learn more

How did we get from there to here?

The idea of artificial intelligence has its roots in the human imagination, taking us back hundreds of years. Think of Mary Shelley's Frankenstein which kicked off the science fiction genre back in 1818. Artificial intelligence was next presaged in Thea von Harbou's 1925 novel Metropolis that was made into a movie by Fritz Lang two years later. It featured Maria, the *Maschinenmensch*, German for machine-human, a fictional humanoid robot. This was the beginning of the golden era of science fiction that ran from the 1930s to the 1940s.



The *Maschinenmensch* from the 1927 film *Metropolis*

By the 1950s, scientists and mathematicians - and even some philosophers - were intrigued by the concept of a form of intelligence from a machine, not a human. Among them was Alan Turing, a British mathematician and philosopher who explored the mathematical possibility of artificial intelligence. For Turning, it was obvious: If humans can use information and reason to solve problems and make decisions, why can't machines? In a paper in 1950 titled Can Machines Think? Turing considered not only how to build intelligent machines but also how to test their intelligence. He

Learn more

Alan Turning, the father of computer science

eventually developed what would be called the Turing test, which analyses a machine's ability to show intelligent behaviour that is equivalent to, or indistinguishable from, intelligence shown by a human.

A few years later, a group of scientists developed Logic Theorist, possibly the first AI program. It was presented at the Dartmouth Summer Research Project on Artificial Intelligence, which was where John McCarthy coined the term artificial intelligence and set in motion excitement and interest in the field of AI research.

Over the next 20 years AI development progressed rapidly. Computers got increasingly more sophisticated. They could store more data, they were faster and they got cheaper. Government agencies funded all sorts of AI research. In 1966, researchers developed some of the first actual AI programs, including Eliza, a computer program that could have a simple conversation with a human. By 1970, Marvin Minsky, a computer scientist who some consider the father of AI, was quoted in Life Magazine: "In three to eight years we will have a machine with the general intelligence of an average human being."

Stanford University's John McCarthy

Learn more

It didn't quite work out that way. It took a little more time to develop computers with enough storage and processing power for AI to continue to flourish. But throughout the 1970s and 1980s, AI researchers did make some major advances. In 1980, Edward Feigenbaum led one of the next big AI breakthroughs: He created "expert systems," a computer program that could simulate the judgement of a human expert in a particular field.

Then in 1986, Carnegie Mellon built Navlab, the first autonomous car. By the 1990s AI started to achieve real goals. One of the most well known achievements was in 1997 when the then-reigning world chess champion Gary Kasparov was defeated by IBM's Deep Blue, a chess playing computer program.

Carnegie Mellon's 1986 Navlab

At the same time, a dizzying number of developments happened in AI. Dragon Systems developed speech recognition software that was implemented on Windows. Cynthia Breazeal created Kismet, a robot that could recognize and display emotions. By 2000, Honda had released ASIMO, an artificially intelligent humanoid robot. And then, in a sign of the times, Steven Spielberg released A.I. Artificial Intelligence, a movie set in a futuristic, dystopian society starring David, a humanoid child that is programmed with anthropomorphic feelings. In 2002 AI got real: i-Robot released Roomba, an autonomous robot vacuum that cleaned while avoiding obstacles.

Over the next few years, AI took some big leaps forward. "Machine reading" or autonomous understanding of text was developed, object recognition software was enhanced and Google started developing a driverless car. In 2014 it managed to pass Nevada's self-driving test.

Learn more

Gary Kasparov playing IBM's 'Deep Blue' in 1997 © AFP

Into the 2010s, a more advanced form of machine learning known as deep learning allowed AI to tackle even more complex tasks. AI systems were used for things like image recognition, natural language processing and machine translation

AI innovation has since progressed rapidly with it slowly becoming embedded in our day-to-day lives. Apple released Siri, a virtual assistant that uses a

Amazon's Alexa - putting AI in every home

Learn more

natural-language user interface to observe, answer and recommend things to its human user. Microsoft released its own virtual assistant, Cortana, and in 2014, Amazon created Amazon Alexa, a home assistant that developed into smart speakers that function as personal assistants.

In 2016, Google, not to be outdone, released Google Home, a smart speaker that uses AI to help users remember tasks, create appointments, and search for information by voice.

OpenAI CEO Sam Altman asking Congress to regulate AI

By the early 2020s, language models like GPT-3 created by OpenAI started to generate excitement in the AI world by creating text that was very similar to human writing. It opened up a whole new world of possibility for AI.

Research in AI is ongoing, with scientists and engineers continuously pushing the boundaries of what AI can do. As technology evolves, new breakthroughs are expected in the near future.

The status of AI now

Artificial intelligence has transformed industries and our daily lives.

Here's how:

- ✓ Machine learning's advancements means that AI systems can process huge amounts of data and improve their performance. This has led to enhanced image recognition, natural language processing and even autonomous vehicles.
- ✓ Natural language processing (NLP) is creating tools like GPT-4, a more sophisticated ChatGPT that can generate even more natural and human-like text.
- ✓ Autonomous systems powered by AI are being used in self-driving cars and drones.
- ✓ AI is revolutionising healthcare by utilising AI algorithms to help in diagnosing illnesses and speeding up processes in drug development.
- ✓ AI is being used to personalise user experiences on streaming services, social media platforms and e-commerce sites with AI algorithms using user's behaviours to offer recommendations.
- ✓ Businesses are using AI for jobs like predictive analytics, fraud detection and supply chain optimization.

Learn more

AI explained: Types of AI

AI is divided into two broad categories.

Weak AI is designed to perform a specific task or set of tasks. It is "weak" because it is limited in its capabilities and can only perform the task it was designed for. It is trained on a specific dataset and can only provide answers based on that dataset.

> Within weak AI, there is **conversational AI** which is trained on large datasets of human interactions and provides responses in a limited series of conversational turns and **generative AI** which is based on large language models, trained on huge datasets of language use and generates new (or at least remixed) content in response to user queries (think OpenAI's GPT)

Strong AI or artificial general intelligence (AGI), is designed to mimic the cognitive abilities of a human being. It is called "strong" because it has the potential to be as intelligent as a human being and can perform a wide range of tasks, rather than being limited to a specific set of tasks.

AI applications

AI could impact every sector thanks to its many different applications.

These include:

Natural Language Processing (NLP) which allows computers to understand and generate human language. This leads to translations or spam filtering.

Machine Vision which allows computers to identify and interpret visual content. This leads to self-driving cars, facial recognition and object detection.

Machine Learning (ML) which allows computers to learn from data and improve their performance over time. This is used for predictive analytics, fraud detection and recommendation processes.

Robotics which deals with robot design, construction and operation. This is used in manufacturing, healthcare and even space exploration.

Learn more

The good, the bad and the future

In recent years, AI has become increasingly common in both business and daily life - with many people not even fully aware that AI is changing the way they go about their day. From virtual assistants to online shopping to driving to streamlining production processes, AI is powering so much of what we do.

AI can give your company an edge or at least ensures it stays apace. The Appen State of AI Report for 2021 stated,

> "The AI industry continues to grow rapidly year-over-year, to the point where organisations that haven't yet invested in their own AI initiatives are at risk of being left behind."

Companies increasingly utilise AI to streamline their internal processes, their customer-facing processes and applications. Implementing AI can help your business achieve its results faster and with more precision by eliminating human error, automating repetitive tasks, sifting and analysing data and providing unbiased decision-making capabilities.

What's the downside?

One of the biggest concerns experts cite for companies using AI is around consumer data privacy and security. But there are other, complicated issues that can arise. The results that AI produces depend on how the data is designed and what data it uses, could lead to decisions that are intentionally or unintentionally biassed. AI can be used in facial recognition equipment or for online tracking and profiling of individuals. In addition, AI enables merging pieces of information a person has given into new data, which can lead to results the person would not expect.

Learn more

Imbalances of access to information could also be exploited. A company can use a person's data to predict behaviour and adapt their message. It is also sometimes unclear to consumers if they are interacting with AI or a person.

And that is just the beginning. As AI gets more sophisticated and as AI tools proliferate, the need to regulate this ever-evolving technology has become evident. Concerns about AI and its potential dangers have been raised by industry professionals, prompting calls for action. Over 50,000 signatories signed a letter in March, 2023 urging an immediate halt in the development of "giant" AIs and the establishment of robust AI governance systems.

Even the CEO of ChatGPT's creator, Sam Altman, appealed to the US Congress to regulate the new technology:

> **"I think if this technology goes wrong, it can go quite wrong. And we want to be vocal about that. We want to work with the government to prevent that from happening."**

The future of AI regulation

In April 2021, the European Commission proposed the AI Act, the first EU regulatory framework for AI. AI systems that can be used in different applications would be analysed and classified according to the risk they pose to users. The different risk levels would mean more or less regulation. In an indication of how quickly AI technology is evolving, at the time of the proposal, the EU law did not give much attention to generative AI systems which can produce text, images and video in response to prompts - like ChatGPT. Under the latest version generative AI is included and has transparency requirements.

In December 2023, the EU took the first step in passing one of the first major laws to regulate AI. This could set a standard for policymakers around the world as they deal with how to manage this rapidly developing technology.

The EU AI Act

The AI Act is designed to minimise risks to consumers and compliance costs to providers while ensuring that AI systems on the EU market are safe, investors can have legal certainty and innovation in AI is encouraged.

It features a risk-based approach, with four different risk classes, each of which covers different use cases of AI systems. While some AI systems are banned entirely, barring narrow exceptions, the Act imposes specific obligations on the providers and deployers of high-risk AI systems, including testing, documentation, transparency, and notification duties.

Different risk levels, different rules

The legislation establishes obligations for both providers and users depending on the level of risk from AI. AI systems will need to be assessed to determine the risk they pose.

Learn more

The 4 levels of risk

1. Unacceptable risk

Unacceptable risk AI systems are systems considered a threat to people and will be banned.

They include:

- ✓ Cognitive behavioural manipulation of people or specific vulnerable groups
  > Such as voice-activated toys that encourage dangerous behaviour in children
- ✓ Social scoring as in classifying people based on behaviour, socio-economic status or personal characteristics
- ✓ Biometric identification and categorisation of people
- ✓ Real-time and remote biometric identification systems, such as facial recognition
- ✓ Social scoring for public and private purposes
- ✓ Exploitation of vulnerabilities of persons, use of subliminal techniques
- ✓ Individual predictive policing
- ✓ Emotion recognition in the workplace and education institutions
- ✓ Untargeted scraping of internet or CCTV for facial images to build-up or expand databases

Some exceptions may be allowed for law enforcement purposes. “Real-time” remote biometric identification systems will be allowed in a limited number of serious cases, while “post” remote biometric identification systems, where identification occurs after a significant delay, will be allowed to prosecute serious crimes and only after court approval.

Learn more

2. High risk

AI systems that negatively affect safety or fundamental rights will be considered high risk and will be divided into two categories:

- ✓ AI systems that are used in products falling under the EU's product safety legislation. This includes toys, aviation, cars, medical devices and lifts.
- ✓ AI systems falling into specific areas that will have to be registered in an EU database:

> Management and operation of critical infrastructure
>
> Education and vocational training
>
> Employment, worker management and access to self-employment
>
> Access to and enjoyment of essential private services and public services and benefits
>
> Law enforcement
>
> Migration, asylum and border control management
>
> Assistance in legal interpretation and application of the law

All high-risk AI systems will be assessed before being put on the market and also throughout their lifecycle.

Learn more

3. General purpose and generative AI

The AI Act considers systemic risks which could arise from general-purpose AI models, including large generative AI models. These can be used for a variety of tasks and are becoming the basis for many AI systems in the EU. Some of these models could carry systemic risks if they are very capable or widely used. For example, powerful models could cause serious accidents or be misused for far-reaching cyberattacks. Many individuals could be affected if a model propagates harmful biases across many applications.

Generative AI, like ChatGPT, would have to comply with transparency requirements:

- ✓ Disclosing that the content was generated by AI
- ✓ Designing the model to prevent it from generating illegal content
- ✓ Publishing summaries of copyrighted data used for training

High-impact general-purpose AI models that might pose systemic risk, such as the more advanced AI model GPT-4, would have to undergo thorough evaluations and any serious incidents would have to be reported to the European Commission.

4. Limited risk

Limited risk AI systems should comply with minimal transparency requirements that would allow users to make informed decisions. After interacting with the applications, the user can then decide whether they want to continue using it. Users should be made aware when they are interacting with AI. This includes AI systems that generate or manipulate image, audio or video content, for example deepfakes.

The vast majority of AI systems currently used or likely to be used in the EU fall into this category. Voluntarily, providers of those systems may choose to apply the requirements for trustworthy AI and adhere to voluntary codes of conduct.

Learn more

EU AI Act: High-risk use cases

Certain critical infrastructures such as in the fields of road traffic and the supply of water, gas, heating and electricity

Education and vocational training such as to evaluate learning outcomes and steer the learning process and monitoring of cheating

Employment, workers management and access to self-employment such as to place targeted job advertisements, to analyse and filter job applications and to evaluate candidates

Access to essential private and public services and benefits such as healthcare, creditworthiness evaluations and risk assessment and pricing for life and health insurance

Certain systems used in the fields of law enforcement, border control, administration of justice and democratic processes

Evaluation and classification of emergency calls

Biometric identification, categorisation and emotion recognition systems (outside the prohibited categories)

Recommender systems of very large online platforms are not included, as they are already covered in other legislation

Learn more

Does the AI Act apply to my company?

The AI Act will apply to both public and private organisations both within and without the EU as long as the AI system is placed on the EU market or its use affects people located in the EU.

The AI Act can apply to both the providers (such as a developer of an AI tool) and deployers of high-risk AI systems. In addition, certain obligations are foreseen for providers of general-purpose AI models, including large generative AI models.

Providers of free and open-source models are exempted from most of these obligations. This exemption does not cover obligations for providers of general purpose AI models with systemic risks.

Obligations also do not apply to research, development and prototyping activities preceding the release on the market, and to AI systems that are exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities.

Learn more

My company is a provider of a high-risk AI system. What do I need to know?

Before placing a high-risk AI system on the EU market, providers must subject it to a conformity assessment. This will enable them to demonstrate that their system complies with the mandatory requirements for trustworthy AI. This involves:

- ✓ data quality
- ✓ documentation and traceability
- ✓ transparency
- ✓ human oversight
- ✓ accuracy
- ✓ cybersecurity
- ✓ robustness

This assessment has to be repeated if the system or its purpose are substantially modified.

AI systems that are safety components of products covered by EU legislation will always be deemed high-risk when subject to third-party conformity assessment.

Biometric systems always require a third-party conformity assessment.

Providers of high-risk AI systems will also have to implement quality and risk management systems to ensure their compliance with the new requirements and minimise risks for users and affected persons, even after a product is placed on the market.

High-risk AI systems that are deployed by public authorities or entities acting on their behalf will have to be registered in a public EU database. If those systems are used for law enforcement and migration they will have to be registered in a non-public part of the database that will be only accessible to relevant supervisory authorities.

Authorities will support post-market monitoring through audits and with the ability to report on serious incidents or breaches of fundamental rights obligations. In case of a breach, the requirements will allow national authorities to have access to the information needed to investigate whether the use of the AI system complied with the law.

Learn more

The penalties

EU member states will have to develop effective, proportionate and dissuasive penalties, including administrative fines, for AI systems that are put on the market or in use that breach the requirements of the AI Act.

The Act sets out thresholds that need to be taken into account:

- ✓ Up to €35m or 7% of the total worldwide annual turnover of the preceding financial year (whichever is higher) for infringements on prohibited practices or non-compliance related to requirements on data
- ✓ Up to €15m or 3% of the total worldwide annual turnover of the preceding financial year for non-compliance with any of the other requirements or obligations of the regulation, including infringement of the rules on general-purpose AI models
- ✓ Up to €7.5m or 1.5% of the total worldwide annual turnover of the preceding financial year for the supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request

> For each category of infringement, the threshold would be the lower of the two amounts for SMEs and the higher for other companies

In order to coordinate national rules and practices in setting administrative fines, guidelines will be set by the Commission.

Learn more

When will the AI Act be fully applicable?

Following its adoption by the European Parliament and the Council, the AI Act will enter into force on the 20th day after its publication in the official Journal. It will be fully applicable 24 months after entry into force, with a graduated approach.

- ✓ At 6 months, Member States will phase out prohibited systems
- ✓ At 12 months, obligations for general purpose AI governance becomes applicable
- ✓ At 24 months, all rules of the AI Act become applicable including obligations for high-risk systems
- ✓ At 36 months, obligations for high-risk systems apply

How will the AI Act be enforced?

Member States hold a key role in the application and enforcement of the AI Act. Each Member State should designate one or more national competent authorities to supervise the application and implementation, as well as carry out market surveillance activities.

To increase efficiency and to set an official point of contact with the public and other counterparts, each Member State should designate one national supervisory authority, which will also represent the country in the European Artificial Intelligence Board.

Additional technical expertise will be provided by an advisory forum, representing a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society and academia.

In addition, the Commission will establish a new European AI Office, within the Commission, which will supervise general-purpose AI models, cooperate with the European Artificial Intelligence Board and be supported by a scientific panel of independent experts.

Learn more

The European Artificial Intelligence Board - What is it and what will it do?

The European Artificial Intelligence Board consists of representatives of national supervisory authorities, the European Data Protection Supervisor, and the Commission. Its role is to facilitate a smooth and effective implementation of the new AI Regulation.

The Board will issue recommendations and opinions to the Commission regarding high-risk AI systems and on other aspects relevant for the effective and uniform implementation of the new rules. It will also support standardisation activities in the area.

Will it work? It's complicated

In an indication of just how complicated artificial intelligence is, the EU debate on the AI Act was incredibly contentious. Officials were divided over how strictly to regulate AI systems in an effort to protect consumers and control this new technology versus the fear of hampering innovation and preventing EU start ups from catching up to their American competitors.

France's president Emmanuel Macron stated that the new EU legislation could hamper European tech companies compared to their rivals in the US, UK and China.

"Regulation must be controlled, not punitive, to preserve innovation," he stated.

Executive vice president of the European Commission Margrethe Vestager responded that the new AI Act would "not harm innovation and research, but actually enhance it." She believes the Act will help tech start ups by creating "legal certainty" as they build their technology.

The final terms of the Act need to be ratified by EU member states. Macron's comments may indicate a new battle is brewing. France, Germany and Italy are considering changes to the law.

There is also the fact that many aspects of the policy will not be in force for up to 2 years - a lifetime in AI development. Who knows where the technology will be then? The initial draft of the AI Act from 2021 had to be rewritten to accommodate the technological breakthroughs that enabled generative AI applications like ChatGPT.

Learn more

The new regulations will be closely watched on a global scale. They will impact mega AI developers like Google, Meta, Microsoft and OpenAI, and smaller companies that are expected to use the technology in sectors like education, health care and banking. And then there are the government agencies that are also using AI systems in everything from criminal justice to determining who gets public benefits.

While there were calls to have the industry self regulate, it's difficult to imagine how that would play out. But the AI Act will involve regulators across 27 countries. It will necessitate hiring new experts and legal challenges are likely as companies test the new rules in court. GDPR has been criticised for being unevenly enforced and the AI Act could suffer a similar fate.

> "The AI Act aims to strengthen Europe's position as a global centre of excellence in AI from the laboratory to the market, ensure that AI in Europe respects our values and norms and harness the potential of AI for industrial use."
>
> Thierry Breton, European Commissioner for Internal Market and Services

How will the AI Act support innovation?

The European Commission notes that the AI Act could increase interest in AI by providing users with the ability to trust organisation's AI systems. It will also help with legal certainty and rules, so AI providers can access bigger markets. Companies in compliance with the AI Act will also have to do real world testing so they will have a "controlled environment" to test their innovative technologies

Racial and gender bias in AI

AI uses data created by people as a starting point, so invariably it could inherit human flaws such as bias based on age, gender or race.

Can the AI Act fix it? Maybe.

When properly designed and used, many AI proponents believe that AI systems can actually reduce bias and existing structural discrimination, and possibly even lead to more equitable and non-discriminatory decisions. This would involve making sure that the data being used to "train" the algorithm is free of bias, or that the algorithm can recognize bias in the data and bring the bias to a human's attention.

Learn more

According to the European Commission, the AI Act's mandatory requirements for all high-risk AI systems will help achieve less racial and gender bias. The Act requires that AI systems are technically robust and that false positive/ negative results will not disproportionately affect protected groups.

High-risk systems will also need to be trained and tested with representative datasets to minimise the risk of embedded unfair biases and ensure that these can be addressed through bias detection, correction and other mitigating measures.

Data racism and bias in AI, European Network Against Racism

Compliance systems will have to ensure these AI systems are regularly monitored and potential risks are promptly addressed.

Learn more

The AI Act & GDPR

GDPR was enacted in May, 2018. Its aim was to provide people with more control over their personal data and give organisations guidelines on how they should collect, process, and store data. AI systems often rely on processing massive amounts of data, including personal data, to improve their performance, so GDPR principles are crucial to consider when designing and implementing AI systems.

With its focus on safeguarding fundamental rights and emphasising transparency and fairness, the AI Act can be seen as a complement to the GDPR. AI often involves personal data processing, necessitating compliance with both regulations, potentially involving data protection impact assessments (DPIAs). High-risk AI systems face stricter regulations under the AI Act, but GDPR requirements remain for personal data processing.

AI testing under the AI Act also aligns with GDPR. Both require organisations to demonstrate compliance and maintain records. Substantial fines and penalties for non-compliance are the consequences under these regulations, depending on the violation's severity.

Ultimately, the goal of both these regulations is protecting personal data and lawful AI development. Organisations will have to meet their obligations under both the GDPR and the AI Act.

Learn more

Case studies: GDPR fines for AI

France's data protection agency pursues Clearview AI

Clearview AI, a US-based startup was accused of massive privacy violations after it scraped selfies off the internet and used people's data to build a facial recognition tool. The EU's GDPR clearly sets out conditions for processing personal data lawfully and Clearview was found to have breached a number of requirements set out in the law by France - which was joined by the UK, Italy and Greece. Whether Clearview will ever pay any of the fines imposed remains unclear because the company is not cooperating with EU regulators.

In December 2021, France's CNIL (its data protection agency) said that Clearview had breached the GDPR by unlawfully processing several tens of millions of citizens' data and failing to provide locals with data access rights. Clearview did not comply with the breach order and the CNIL slapped Clearview with another breach and issued the biggest fine it possibly could under the GDPR. The regulation allows for fines of up to 4% of global annual turnover or €20 million, whichever is higher.

In May, 2023 CNIL added an overdue penalty payment of €5.2 million on Clearview. The CNIL also instructed Clearview not to collect and process data on people located in France without a proper legal basis and to delete data of individuals whose information it had processed unlawfully, after fulfilling any outstanding data access requests.

To date, the company has not paid any of its outstanding fines. Its official response is that it does not have a place of business in France and contends that it is not subject to the GDPR.

Learn more

Italy's data protection agency fines Uber

Italy's data protection authority, the Garante, fined ridesharing platform Uber 4.2 million euros for data processing violations. The Garante said it found a subsidiary of Uber processed users' personal data without consent and without notifying supervisory agencies. Approximately 57 million users worldwide were affected. The Garante said the personal data included contact information, Uber account data, location and relations with other users. The Garante also said Uber's management companies based in the Netherlands and US violated the personal data protection code for Italian users.

Learn more

AI regulation around the world

Around the world, countries are taking their own approach to managing AI. From China to Canada to Brazil, countries are creating regulations to manage the ever-evolving technology.

As Sam Altman, OpenAI's CEO, said when he called for regulation of the industry:

> "If someone does crack the code and build a superintelligence ... I'd like to make sure that we treat this at least as seriously as we treat, say, nuclear material."

Learn more

AI regulation in the US

In the US, a number of states are working on their own approaches to regulating AI, while discussions at the national level are gaining momentum. White House officials, including vice president Kamala Harris, met with Big Tech CEOs in early May to discuss the potential dangers of the technology.

The US National Telecommunications and Information Administration is seeking to find out if there are measures that could be put in place to provide assurance "that AI systems are legal, effective, ethical, safe, and otherwise trustworthy."

The White House Office of Science and Technology Policy released the Blueprint for the Development, Use, and Deployment of Automated Systems (also known as the Blueprint for an AI Bill of Rights) on October 4, 2022. In contrast to the EU's draft AI Act, this Blueprint is not legally binding and outlines five principles aimed at minimising potential harm caused by AI systems. On August 18, 2022, the National Institute of Standards and Technology (NIST) published the second draft of its AI Risk Management Framework for public feedback. The original version of the framework dates back to March 2022 and is based on a concept paper from December 2021. The AI Risk Management Framework is designed to assist companies involved in the development or deployment of AI systems in assessing and mitigating risks associated with these technologies. It consists of voluntary guidelines and recommendations, emphasising that it is non-binding and explicitly not intended to be interpreted as a regulation.

To date, there has been no serious consideration of a US version of the EU AI Act or any sweeping federal legislation to govern the use of AI, nor is there any substantial state legislation in force, despite some state privacy laws that may extend to AI systems that process certain types of personal data.

Learn more

AI regulation in the UK

The UK has released a white paper with an outline of the principles the AI industry should adhere to, but without a legislative framework as of yet.

The white paper outlines that AI has shown significant societal benefits, such as advancements in medicine and efforts to combat climate change. For instance, DeepMind, a UK-based company, has developed AI technology that can predict the structure of various proteins, aiding scientific research and the development of life-saving medicines.

The UK government recognizes AI as a critical technology and aims to lead the international conversation on AI governance. Sir Patrick Vallance's Regulation for Innovation review emphasises the need for timely government intervention to create a clear regulatory environment that attracts foundational AI companies. While AI presents immense opportunities, it also poses risks to physical and mental health, privacy, and human rights. Addressing these risks and building public trust are essential for widespread adoption of AI and maximising its benefits.

The UK has issued principles on a non-statutory basis and utilising existing regulators' expertise and includes feedback from regulators, industry, and academia which may lead to a future potential statutory duty.

The UK framework does not anticipate an AI regulator at present, but is keen to collaborate with international partners to develop interoperable measures and ensure compatibility between approaches to AI regulation.

Learn more

AI regulation in Brazil

The newly elected leftist government of Lula in Brazil is in the process of developing its first law to regulate artificial intelligence. In December 2022, a Senate panel presented a report containing studies on AI regulation, along with a draft for AI regulation.

The main aims of the legislation are to safeguard the rights of individuals affected by AI systems, categorise the level of risk associated with these systems, and establish governance measures for companies that provide or operate AI systems.

The draft shares similarities with the European Union's (EU) draft AI Act. The definition of AI systems in the Brazilian draft closely aligns with the EC's draft definition. Similar to the AI Act, the draft proposes risk categories and corresponding obligations. Prohibited AI systems include those that exploit vulnerabilities of specific groups of individuals with the intention to harm their health or safety. Social scoring by public entities and the use of biometric identification systems in publicly accessible spaces are also prohibited, except when explicitly authorised by specific laws or court orders, such as for criminal investigations.

The Brazilian draft, like the AI Act, identifies high-risk systems that are sensitive to fundamental rights. These include AI systems used in critical infrastructure, education and vocational training, recruitment, autonomous vehicles, and biometric identification. The list of high-risk systems can be adjusted by a designated authority, and such systems will be publicly listed in a database.

Learn more

The draft grants data subjects rights against providers and users of AI systems, regardless of the risk level. These rights include access to information about their interactions with AI systems, the right to receive explanations for decisions made by AI systems within 15 days of request, the right to challenge decisions that significantly affect their interests or have legal effects, the right to human intervention in decisions made solely by AI systems, the right to non-discrimination and correction of biassed outcomes, and the right to privacy and protection of personal data.

Governance measures are also addressed in the draft, akin to the AI Act. Providers and users of AI systems are required to establish internal structures and processes that ensure the safety of AI systems. High-risk AI systems necessitate more stringent measures, such as conducting publicly available AI impact assessments, which may need to be periodically repeated.

Additionally, the draft includes provisions related to reporting serious security incidents to the competent authority and regulations concerning civil liability. Like with GDPR, the penalties for non-compliance vary depending on the violation, but can include fines of up to 50 million Brazilian reals (approximately 9 million euros) or up to 2% of a company's turnover.

Learn more

AI regulation in China

In January 2022, China promulgated two laws specific to AI applications. While the provisions regarding the management of algorithmic recommendations for internet information services (Algorithm Provisions) have been in effect since March 2023, the provisions for managing deep synthesis of internet information services (Draft Deep Synthesis Provisions) are still in the drafting stage.

Algorithm Provisions

These regulations aim to address the misuse of algorithmic recommendation systems and include provisions concerning content management, tagging or labelling, transparency, data protection, and fair practices. Additional regulations are applicable in specific areas, such as those related to minors or e-commerce services. Non-compliance may result in fines ranging from 10,000 to 100,000 RMB (approximately 1,570 to 15,705 US dollars).

Learn more

Draft Deep Synthesis Provisions

These provisions seek to regulate "deep synthesis" technologies, particularly in combating deep fakes. With the exception of fair practices, these laws cover all the aspects mentioned above. Additionally, certain obligations for online app store operators are included. The maximum penalties are the same as those specified in the Algorithm Provisions.

China's Cyberspace Administration (CAC) concluded its consultation on the draft Administrative Measures for Generative Artificial Intelligence Services on May 10, 2023. This draft regulation mandates a "safety assessment" for new AI products developed in China before their release to the public. Specifically, the regulation requires AI-generated content to be truthful and accurate, while prohibiting content that undermines state power, contains terrorist or extremist propaganda, promotes violence, contains obscene or pornographic information, incites ethnic hatred or discrimination, or disrupts economic and social order. The regulation also requires AI service providers to take measures to prevent the generation of false information and harmful content. If inappropriate content is generated, providers must update their technology within three months to prevent similar content from being produced. Non-compliance with the regulation may result in fines, service suspensions, or criminal investigations.

Additionally, China has implemented regional legislation concerning AI. On September 6, 2022, the Shenzhen government published China's first city-level AI regulation, known as the "Regulations on Promoting Artificial Intelligence Industry in Shenzhen Special Economic Zone." Shanghai also published a provincial law on AI development called the "Shanghai Regulations on Promoting the Development of the AI Industry" which came into force in October 2022.

Learn more

AI regulation in Japan

With the second largest IT sector in the world, Japanese AI regulations are strongly correlated to the major project "Society 5.0". Behind this is the ambition to counter social problems (such as the ageing population) with innovation.

In 2019, the Japanese government's Social Principles document outlined the basic principles of an AI-capable society. The first part contains seven social principles that society and the state must respect when dealing with AI: human-centricity, education/literacy, data protection, ensuring safety, fair competition, fairness, accountability and transparency, and innovation. However as yet, Japanese measures are not legally binding.

Learn more

AI regulation in Canada

The Canadian federal government introduced draft law C-27, known as the Digital Charter Implementation Act 2022, in June 2022. Part 3 of this legislative package includes the Artificial Intelligence and Data Act (AIDA), which serves as Canada's inaugural AI Act.

AIDA's purpose is to regulate the international and interprovincial trade of AI systems by mandating certain individuals to take measures aimed at reducing the risks of harm and biassed outcomes associated with high-performance AI systems. The Act also establishes provisions for public reporting and grants the Minister the authority to order the disclosure of records related to AI systems. Furthermore, the legislation prohibits specific practices concerning the handling of data and AI systems that could cause severe harm to individuals or their interests. The legislation is likely to come into force later in 2023.

Learn more

AI regulation in Australia

The Australian federal government has outlined its intention to regulate artificial intelligence, saying there are gaps in existing law and new forms of AI technology will need “safeguards” to protect society. The government is considering whether to adopt AI risk classifications, like those being developed in Canada and the EU. They currently have a voluntary guide known as the 8 Artificial Intelligence (AI) Ethics Principles of Australia that are designed to ensure AI is safe, secure and reliable. They are as follows:

Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment

Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals

Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups

Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data

Reliability and safety: AI systems should reliably operate in accordance with their intended purpose

Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them

Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system

Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled

Learn more

AI regulation in Singapore

Singapore has implemented a key regulation to govern the use of AI, particularly in regard to data protection and privacy. The Personal Data Protection Act (PDPA) establishes specific obligations on organisations when it comes to the collection, use, and disclosure of personal data.

The Infocomm Media Development Authority (IMDA), an agency based out of Singapore, has developed AI Verify, an AI governance testing framework and a software toolkit. The testing framework consists of 11 international accepted AI ethics principles, and are consistent with globally recognized AI frameworks. AI Verify helps organisations validate the performance of their AI systems against these principles through standardised tests.

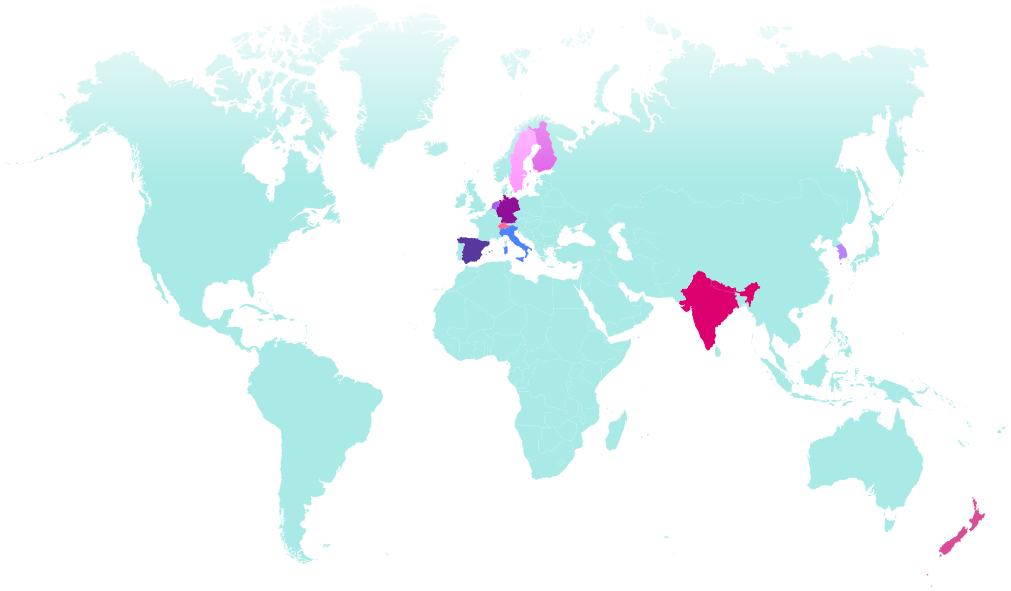
Learn more

AI regulation in France

France has developed a National Strategy for AI, composed of an ambitious national program dedicated to research, numerous policies dedicated to AI dissemination in the economy, and an ambitious massively funded national training plan on AI.

Learn more

What are other countries doing on AI?



GERMANY: Germany has proposed the Artificial Intelligence Strategy, including voluntary guidelines for ethical AI development and deployment

SOUTH KOREA: South Korea has developed voluntary guidelines for the responsible use of AI in areas such as privacy, safety, and transparency

SWEDEN: Sweden has established voluntary ethical guidelines for the development and use of AI, focusing on transparency, accountability, and non-discrimination

FINLAND: Finland has introduced voluntary ethical guidelines through the "Ethical AI Program" to ensure responsible and trustworthy AI implementation

NETHERLANDS: The Netherlands has initiated discussions on AI regulation but has not enforced mandatory regulations yet. They emphasise voluntary codes of conduct and guidelines

SPAIN: Spain has introduced a National Artificial Intelligence Strategy that includes voluntary ethical guidelines and principles for AI development and deployment

ITALY: Italy has adopted voluntary guidelines for ethical AI and is in the process of formulating national AI policies and regulations

INDIA: India has proposed the National Strategy for Artificial Intelligence, which includes voluntary guidelines for the responsible and inclusive development and deployment of AI

NEW ZEALAND: New Zealand has developed voluntary guidelines such as the "Algorithm Charter for Aotearoa New Zealand" to encourage ethical and transparent use of algorithms and AI

SWITZERLAND: Switzerland has focused on voluntary guidelines for ethical AI development, along with promoting research and innovation in AI technologies

Learn more

Where do we go from here? Collaboration v. confrontation

It's hard to predict what lies in AI's regulatory future for the simple reason that AI's technology is constantly evolving. This means its issues, of bias or copyright or lack of transparency, to name a few will continue to shape the regulatory agenda for many years to come.

Two news stories highlight the complex relationship companies have with AI and the often circuitous path they will have to take to achieve some sort of control over this nearly unmanageable technology.

In December, 2023, The New York Times sued OpenAI and Microsoft, accusing them of using millions of the newspaper's articles without permission to help train chatbots to provide information to readers. This is the first time a major media organisation in the US is suing OpenAI and Microsoft over copyright issues related to its written works.

Generative AI technologies, that generate text, images and other media from short prompts, have led other groups - writers, computer programmers - to file copyright suits against AI companies. But these companies contend that they can legally use the content for free to train their technologies because it is public and they don't reproduce the material in its entirety.

The Times has not specified an exact amount of money it is seeking in the suit. It does say that OpenAI and Microsoft should be held responsible for billions of dollars in damages for unlawful copying and using their work. The Times also wants the two companies to destroy any chatbot models and training data that use their copyrighted material.

The lawsuit could define the legal implications of generative AI technologies and have huge implications for the news industry.

In an interesting element of the case, The Times tried to work out a resolution with Microsoft and OpenAI that would involve a commercial agreement and some sort of protections around the AI systems.

At the same time, German publishing giant Axel Springer SE, which owns Politico and Business Insider, did reach a licensing agreement with OpenAI in which the company will pay Axel Springer to use its news content in the company's AI products. This collaboration is a new kind of publishing deal that will allow the ChatGPT creator to train its AI models on the news organisation's reporting.

Learn more

As part of the deal, when users ask ChatGPT a question, the chatbot will deliver summaries of relevant news stories from Axel Springer brands. Those summaries will include material from stories that would otherwise require subscriptions to read. The summaries will cite the Axel Springer publication as the source, and also provide a link to the full article it summarises.

The agreement, which involved millions of euros, reflects another direction for companies on either side of AI systems.

Getting ready (It's time)

It's time to start preparing for the AI Act, a comprehensive approach to managing the risks of this technology. While it will be two years till the Act is in force, getting ready now means that your company will more easily and effectively work within this complex regulation and remain in compliance. This involves taking a deep look at your company's operations and supply-chain management. Both AI developers and users of AI, especially high-risk AI, in a range of sectors from financial services to employment to medical devices to health care, and others should take the following steps:

- ✓ Look at who - or what department - is using AI tools
- ✓ Check if the AI tools are working with proprietary or sensitive personal data
- ✓ Define the category under the AI Act that these use cases fall - unacceptable, high, or low-to-no-risk category
- ✓ Consider whether the AI tools being developed or provided could be used in the EU market
- ✓ Consider whether your company uses AI tools that could have any exposure in the EU market
- ✓ Check (carefully) your vendor agreements with the AI tool providers. Check how the agreement addresses data protection, use restrictions and compliance obligations

The EU's AI Act will likely set the standard for how countries regulate and manage the risks of AI. Make sure you are fully prepared for this wide-ranging and impactful regulation.

Learn more

Making the most of artificial intelligence while keeping your business safe

Artificial intelligence (AI) is rapidly changing the workplace. Generative AI tools like ChatGPT and Dall-E now allow people around the world to accomplish more than humans ever dreamed possible, generating text and images instantly, and crunching huge volumes of data effortlessly. Meanwhile, businesses are struggling to understand the opportunities and threats emerging in the AI revolution. VinciWorks' innovative AI compliance courses include a variety of training courses for organisations looking to keep their team ahead of the curve.

With our in-browser editing tool, you can now tailor any of the courses in real-time. Edits are clearly visible, and you can easily share the updated training materials with your colleagues via a unique link.

Try our collection of AI-at-work courses

With our AI courses you will...

- ✓ Understand the concepts and terms used in discussing AI
- ✓ Get advice on best practices for using AI in the workplace
- ✓ Gain familiarity with the risks associated with AI use
- ✓ Explore AI's moral issues and challenges

Contact us

Learn more

VinciWorks

Contact us

www.vinciworks.com
enquiries@vinciworks.com
+44 (0) 208 815 9308