SPITCH 1 YEARS

White Paper on the Impact of the EU AI Act on Using Conversational AI for Business Purposes

www.spitch.ai

March 2024

Relevance and potential impact of the EU AI Act for businesses

Early adopters among some of the leading enterprises are already capitalizing on vast opportunities presented by GenAI[1] and LLMs[2]. They are keen, however, to balance it with data privacy and security imperatives, which are also perceived as fundamental by end-customers. Conversational AI solutions for customer service, such as virtual assistants (text and voice chatbots), voice biometrics, and speech analytics, among others provided by Spitch, make a real difference to business performance and data security. Balancing business advantages of innovative technologies, effective regulation to control risk, and social responsibility appears critical for sustainable AI adoption.

The introduction of the EU AI Act provides the first comprehensive regulatory framework for AI governance. It is bound to have an impact on both the providers and deployers of some forms of GenAI and face recognition biometrics that can potentially present risks if used in an unregulated and uncontrolled manner. We believe that the impact of the new EU regulations will be largely positive for use cases in different industries and the public sector, with relatively smaller influence on the providers of business domain-specific Conversational AI. Nevertheless, some new requirements will have to be observed by all providers and deployers of AI systems.

These requirements are expected to be further developed by European standardisation bodies. Subsequently national authorities will need to ensure that companies comply with the new AI governance, risk management requirements and standards, while assessing the extent to which more detailed sectoral guidance may be required.

EU AI Act Summary*

The AI Act classifies AI according to its risk:

- Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI).
- Most of the text addresses high-risk AI systems, which are regulated.
- A smaller section handles limited risk AI systems, subject to lighter transparency obligations: developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes).
- Minimal risk is unregulated (including the majority of AI applications currently available on the EU single market, such as AI enabled video games and spam filters – at least in 2021; this is changing with generative AI).

The majority of obligations fall on providers (developers) of high-risk AI systems.

- Those that intend to place on the market or put into service high-risk AI systems in the EU, regardless of whether they are based in the EU or a third country.
- And also third country providers where the high risk AI system's output is used in the EU.

Users are natural or legal persons that deploy an AI system in a professional capacity, not affected end-users.

- **Users (deployers) of high-risk AI systems** have some obligations, though less than providers (developers).
- This applies to users located in the EU, and third country users where the AI system's output is used in the EU.

* High-level summary of the AI Act

Spitch believes that it may be useful for existing and prospective customers and partners to review the answers to some of the frequently asked questions on the new requirements summarised below.

[1] *GenAI – Generative Artificial Intelligence*

[2] *LLMs – Large Language Models*

Q&As

1. What are the prohibited AI systems under the EU AI Act? Do any of the conversational AI systems for customer service fall under the 'prohibited' category?

The following types of AI system are 'Prohibited' according to the AI Act (Title II, Art. 5)[3]:

- Deploying **subliminal, manipulative, or deceptive techniques** to distort behaviour and impair informed decision-making, causing significant harm.
- **Exploiting vulnerabilities** related to age, disability, or socio-economic circumstances to distort behaviour, causing significant harm.
- **Biometric categorisation systems** inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), except labelling or filtering of lawfully acquired biometric datasets or when law enforcement categorises biometric data.
- **Social scoring**, i.e., evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people.
- **Assessing the risk of an individual committing criminal offenses** solely based on profiling or personality traits, except when used to augment human assessments based on objective, verifiable facts directly linked to criminal activity.
- **Compiling facial recognition databases** by untargeted scraping of facial images from the internet or CCTV footage.
- **Inferring emotions in workplaces or educational institutions**, except for medical or safety reasons.

EU AI Act Summary*

General purpose AI (GPAI):

- All GPAI model providers must provide technical documentation, instructions for use, comply with the Copyright Directive, and publish a summary about the content used for training.
- Free and open license GPAI model providers only need to comply with copyright and publish the training data summary, unless they present a systemic risk.
- All providers of GPAI models that present a systemic risk – open or closed – must also conduct model evaluations, adversarial testing, track and report serious incidents and ensure cybersecurity protections.

* High-level summary of the AI Act

* High-level summary of the AI Act

- **'Real-time' remote biometric identification (RBI) in publicly accessible spaces for law enforcement**, except when:
  - Searching for missing persons, abduction victims, and people who have been human trafficked or sexually exploited;
  - Preventing substantial and imminent threat to life, or foreseeable terrorist attack; or
  - Identifying suspects in serious crimes (e.g., murder, rape, armed robbery, narcotic and illegal weapons trafficking, organised crime, and environmental crime, etc.).

None of the generally-accepted conversational AI systems commonly used for business purposes in customer service fall under these categories.

[3] Source: https://artificialintelligenceact.eu/high-level-summary/

Many companies, especially banks, have started deploying the real-time form of voice biometric identity verification for their customers. In most cases, it serves to augment a human-controlled process of identity verification during the live conversation with a contact centre agent. During the call, the voice of the customer is compared with the previously created voiceprint (a mathematical representation of the user's voice characteristics) to ensure that the caller's voice matches.

There are also 'hybrid' voice biometrics verification processes where a randomly generated numbers or words are repeated by the caller and the live voice is matched with the voiceprint. This method helps ensure that the verification happens quicker.

Unless these are used in public spaces by the law enforcement bodies, such remote voice biometrics authentication methods do not fall under the category of prohibited AI systems.

However, if conversational AI systems use emotion detection and inference to inform automated decisions in a workplace environment or categorize callers based on the emotions detected, such systems may fall under the category of prohibited AI systems under the EU AI Act.

But, if sentiment analysis is used as part of a broader customer service AI system to understand customer feedback and improve service quality, without making automated decisions that directly impact individuals, it would not be considered a prohibited use case. It's important to note that the EU AI Act does not prohibit the use of sentiment analysis in general, but rather specific applications that are deemed high-risk or unethical. Businesses should carefully assess how they plan to use sentiment analysis and ensure that it does not fall under the prohibited use cases outlined in the Act.

2. What AI systems are classified as 'high-risk' under the EU AI Act? Is there a possibility that conversational AI systems use will create risks for businesses because of this classification and the need to meet additional requirements?

Some AI systems are considered 'high-risk' under the AI Act (**Title III**)[4]. Providers of those systems will be subject to additional requirements. High-risk AI systems (**Art. 6**) are considered to be those that are used as a safety component or a product covered by EU laws in **Annex II** **AND** required to undergo a third-party conformity assessment under those Annex II laws; **OR** those under **Annex III** use cases (see the box below), except if:

- the AI system performs a narrow procedural task;
- improves the result of a previously completed human activity;
- detects decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review; or
- performs a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III.

Importantly, the AI system is always considered high-risk if it **profiles individuals**, i.e. conducts automated processing of personal data to assess various aspects of a person's life, such as work performance, economic situation, health, preferences, interests, reliability, behaviour, location or movement.

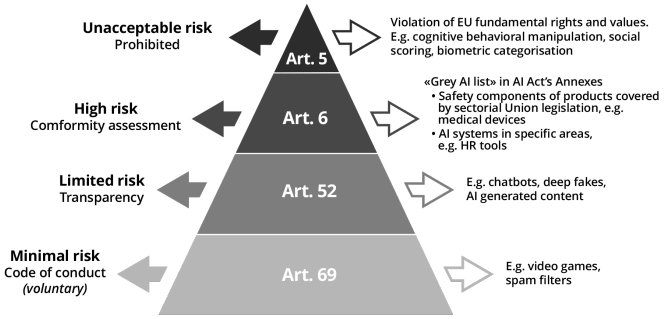
Providers that believe their AI system, which falls under **Annex III**, is not high-risk, **must document** such an assessment before placing it on the market or putting it into service.

There is a number of requirements for providers of high-risk AI systems as specified in Article 8-25 of the Act. These include the obligation to establish a **risk management system** throughout the high-risk AI system's lifecycle; conduct appropriate **data governance**, draw up **technical documentation** to demonstrate compliance and provide authorities with the information to assess that compliance; ensure that the high-risk AI system provides for **record-keeping** automatically recording events relevant for identifying national-level risks and substantial modifications throughout the system's lifecycle.

Other requirements include developing **instructions for use** by downstream deployers to enable the latter's compliance and allowing to implement **human oversight**; guaranteeing the appropriate levels of **accuracy, robustness, and cybersecurity**, as well as establishing a **quality management system** to ensure compliance.

[4] *Source: https://artificialintelligenceact.eu/high-level-summary/*

Some non-banned biometrics systems such as voice biometrics systems provided by Spitch may fall under those listed in Annex III. Other conversational AI systems, such as chatbots, speech analytics, knowledge bases etc. may fall under 'Limited risk' or Minimal risk' categories with minimal or no additional requirements as illustrated by the graph[5] below:



3. What is the timeline for ensuring compliance with the EU AI Act?

Spitch offers consulting services to its customers to help ensure full and timely regulatory compliance, in accordance with EU AI Act and other regulatory frameworks in a specific country or region, including data protection regulations.

Our lawyers help clients formulate customer agreements based on the existing legal requirements in each country that regulate the deployment of AI systems. These should be approved by lawyers at the client side before the project roll-out.

Below is the graph[5] illustrating the timeline for implementation, according to the official journal of the EU. For most conversational AI solutions, e.g. non-banned biometrics and chatbots, the additional requirements will have to be met within 12-36 months starting May 2024, when the EU AI Act enters into force.

Annex III use cases

Non-banned biometrics: Remote biometric identification systems, excluding biometric verification that confirm a person is who they claim to be. Biometric categorisation systems inferring sensitive or protected attributes or characteristics. Emotion recognition systems.

Critical infrastructure: Safety components in the management and operation of critical digital infrastructure, road traffic and the supply of water, gas, heating and electricity.

Education and vocational training: AI systems determining access, admission or assignment to educational and vocational training institutions at all levels. Evaluating learning outcomes, including those used to steer the student's learning process. Assessing the appropriate level of education for an individual. Monitoring and detecting prohibited student behaviour during tests.

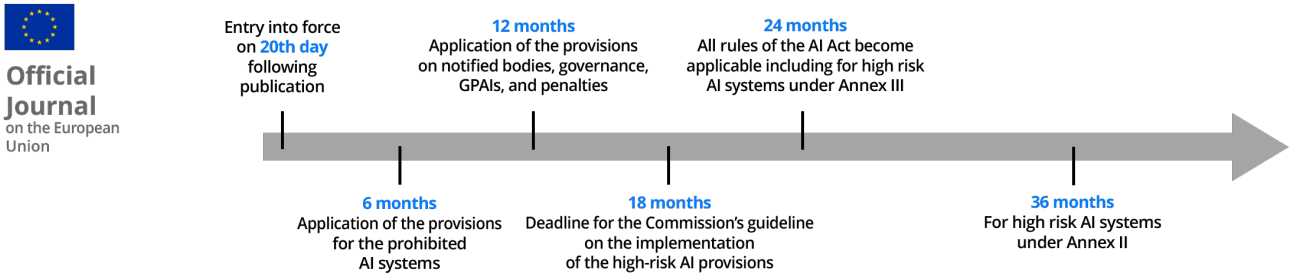
Employment, workers management and access to self-employment: AI systems used for recruitment or selection, particularly targeted job ads, analysing and filtering applications, and evaluating candidates. Promotion and termination of contracts, allocating tasks based on personality traits or characteristics and behaviour, and monitoring and evaluating performance.

Access to and enjoyment of essential public and private services: AI systems used by public authorities for assessing eligibility to benefits and services, including their allocation, reduction, revocation, or recovery. Evaluating creditworthiness, except when detecting financial fraud. Evaluating and classifying emergency calls, including dispatch prioritising of police, firefighters, medical aid and urgent patient triage services. Risk assessments and pricing in health and life insurance.

Law enforcement: AI systems used to assess an individual's risk of becoming a crime victim. Polygraphs. Evaluating evidence reliability during criminal investigations or prosecutions. Assessing an individual's risk of offending or re-offending not solely based on profiling or assessing personality traits or past criminal behaviour. Profiling during criminal detections, investigations or prosecutions.

Migration, asylum and border control management: Polygraphs. Assessments of irregular migration or health risks. Examination of applications for asylum, visa and residence permits, and associated complaints related to eligibility. Detecting, recognising or identifying individuals, except verifying travel documents.

Administration of justice and democratic processes: AI systems used in researching and interpreting facts and applying the law to concrete facts or used in alternative dispute resolution. Influencing elections and referenda outcomes or voting behaviour, excluding outputs that do not directly interact with people, like tools used to organise, optimise and structure political campaigns.



[5] Source: https://www.engage.hoganlovells.com/knowledgeservices/news/the-eu-ai-act-an-impact-analysis-part-1

4. Are there any provisions of the EU AI Act that would make it impossible or too risky to use conversational AI solutions provided by Spitch?

In short, there are none.

Certain provisions would require additional requirements to be met by the provider of the AI systems, i.e., Spitch, but not the customers. Spitch ensures that all the solutions using real-time voice biometrics, sentiment analysis and emotion detection, those referring to GPT-4 or other General purpose AI models for summarization and categorization, as well as the systems classified under the EU AI Act as 'limited-risk', such as virtual assistants (AI chatbots) are fully compliant within the prescribed time-limits.

5. What are new obligations of the customers deploying the AI systems provided by Spitch under the EU AI act?

According to the Article 52 of the Act, Deployers of an emotion recognition system or a biometric categorisation system must inform of the operation of the system the natural persons exposed thereto and process the personal data in accordance with Regulation (EU) 2016/679, Regulation (EU) 2016/1725 and Directive (EU) 2016/280, as applicable.

Besides, Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated.

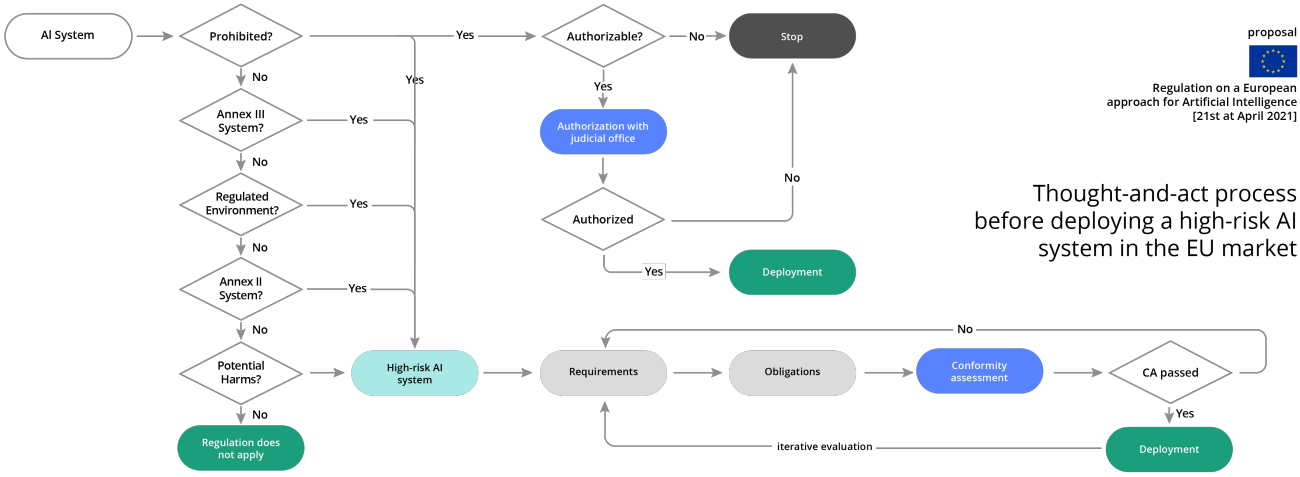
The information referred to above must be provided to the concerned natural persons in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. The information must respect the applicable accessibility requirements.

How to implement the requirements may vary depending on the Spitch solution, our Professional Services Team will assist you during the implementation to ensure you are compliant with the latest regulations.

6. Is extraction of performance data in the workplace environment for the purposes of quality management permitted under the Act? What actions ensure legality?

Extracting such data by AI systems is permitted unless it is used for making automated decisions without the involvement of human beings, inferring emotions in workplaces, inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), or social scoring, i.e., evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people.

In order to determine whether the provisions of the new regulation apply in your specific case, or in case any of the AI systems may fall under 'high-risk' category, according to the EU AI Act, the following decision-making procedure[6] is recommended to ensure compliance:



[6] Source: https://futurium.ec.europa.eu/en/european-ai-alliance/open-discussion/more-visual-guide-proposed-eu-artificial-intelligence-act?language=da