kubrick

Inside the EU AI Act:
What is coming and what does it mean for organisations embracing AI?

AUGUST 2023

kubrick

CONTENTS

kubrick

As the pace of development in AI technology continues to accelerate, organisations must balance innovation with caution in the face of pending regulation – and a sharp deadline to reach compliance. Kubrick Head of Data Management Simon Duncan examines the proposed EU AI Act as it awaits final negotiations to unpack the key takeaways which will help organisations adopt trustworthy and human-centric AI. Simon breaks down the Act's defining principles and classifications regarding High Risk, Low Risk, and Prohibited AI systems, and the lessons learned from GDPR to help leaders prepare for rapid adoption of AI governance.

INTRODUCTION

ORIGINS OF THE EU AI ACT

The European Commission's proposal for the regulation of AI and the framework around its evaluation and application was first announced formally in April 2021, following the 2020 white paper 'On Artificial Intelligence – A European approach to excellence and trust'. The Artificial Intelligence (AI) Act represents the EU Parliament's position prior to review (and negotiation with) the EU Council of Ministers and the EU Member States. As of June 14, 2023, the Act was adopted with an overwhelming 499 votes in favour, 28 against, and 93 abstentions.

A final text will not become law until the trilogue (3-way talks) involving the EU Commission, Council, and Parliament have been completed, which is expected to take 2 years end-to-end. When the same process was enacted for GDPR, it resulted in several exemptions for Member States[1], including several alterations for the UK such as the Age of Consent for use of personal data. Similar adjustments should be expected as a part of these negotiations, with particular concerns regarding individual governments' use of facial recognition technology.

From the Act itself - and the rhetoric surrounding it - we have chosen to explore the elements of the Act that will most likely provide a range of challenges to the Council of Ministers, the EU Member States, and potential tensions with countries outside the EU, particularly the UK and US. We will consider the impact this Act may impose on organisations which not only develop AI but also those who leverage AI technologies and products to improve efficiencies, examining the key risk categories which will determine how viable AI products will be. While there cannot be any concrete recommendations for action until the Act is formally enacted as law, this commentary should support proactive considerations of what is to come. As AI regulation follows closely in the wake of GDPR and the challenges it posed for many organisations, the need to anticipate and prepare for change will become a top priority for technology leaders.

[1] As consolidated in the powers of Article 25

kubrick

CONTEXT

A HISTORY OF DEFINING PRINCIPLES

Before delving into specific elements of the legislation in some detail, we should consider how they fit within the broader principles and consensus of earlier legislation that has been passed. The two defining principles underlining earlier legislation were consumer protection and financial transparency – these lay at the heart of legislation directed at the financial sector, Solvency II (Insurance) and MiFID II (Fund Management). Similarly for GDPR, there was an adjustment to emphasise consumer protection and data transparency (in placed financial transparency) as a core principle. Underlying both principles is the tradition of individual human rights and political freedoms enshrined in the European Convention of Human Rights (ECHR). This tradition has continued to act as a key driver in this new Act and is clearly recognisable in the details and wider rhetoric: protecting EU citizens through a human-centric approach is at the heart of the EU AI Act.

Importantly, the Act provides a fundamental basis for all member states to cohere around a developed (and still developing) political consensus. GDPR came into force in 2018 - the same year a High-Level Expert Group on AI (HLEG) was established, constituted by 52 well-known experts in the field. The HLEG was responsible for swathes of documentation, white papers, and guidelines which preceded the AI act, including various voluntary compliance initiatives such as the Ethics Guidelines for trustworthy AI. Subsequently, the AI Alliance of 4000 stakeholders was established, demonstrating the exhaustive focus and information gathering that led to the formulation of this Act.

ESTABLISHING A FRAMEWORK AND EXPECTATIONS

OVERVIEW

The Commission has proposed 4 specific objectives within their regulatory framework on AI:

- Ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values.
- Ensure legal certainty to facilitate investment and innovation in AI.
- Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems.
- Facilitate the development of a single market for lawful, safe, and trustworthy AI applications and prevent market fragmentation.

This AI Act is intended to be fast-tracked for the necessary negotiations to be completed and in law by the end of 2023 in order to meet the urgency for regulation in industry. However, there is still likely to be a 2-year commencement period to enable organisations enough time to comply, and most importantly give member states enough time to develop the necessary regulators. While some commentators have suggested the expedition of elements of the Act into law to address the riskiest and most concerning facets of AI, it was noted that dates should not be adjusted with to serve the sense of urgency, in keeping with true governance best practice.

The 4 key objectives are complemented by additional legislation that has already been passed by the EU Parliament: The EU Digital Services Act, The EU Data Act, the Open Data Initiative, and the EU Data Governance Act. Before commencement of the EU AI Act, these ancillary pieces of legislation will already be in effect (either in part or wholly) with compliance and potential fines already in force under the broader European strategy for data[2]. Most notably, all of these regulations will have broad territorial scope, requiring compliance by UK companies engaging in the described activities within the EU.

The EU AI Act's framework, as stated in the preamble of the Act itself, '...lays down a coherent, effective, and proportionate framework to ensure AI is developed in ways that respect people's rights and earn their trust, making Europe fit for the digital age and turning the next ten years into the Digital Decade. Furthermore, the promotion of AI-driven innovation is closely linked to the Data Governance Act, the Open Data Directive, and other initiatives under the EU strategy for data, which will establish trusted mechanisms and services for the re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality. The proposal also strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI that is consistent with Union values and interests. It provides the Union with a powerful basis to engage further with its external partners, including third countries, and at international fora on issues relating to AI.'

> **"The proposal...strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI..."**
> EU AI ACT

This proclamation of intent also represents a continuity to the ambitions realised by GDPR – a recognised 'gold standard' of regulation and guidance that the UK continues to be associated with around personal data protections. It also provides a preview to the biometric protections that will be published by the EU later this year visible in the New York State S.H.I.E.L.D. Act and California's C.P.R.A. In the UK, the Centre for Data Ethics and Innovation has also provided and promoted a lot of detail shadowing this EU legislation[3].

[2] Published 19th February 2020, Document 52020DC0066
[3] e.g., the recent white paper from the UK government 'A pro-innovation approach to AI regulation' (March 29th, 2023)

kubrick

The Act is broken down into 12 Titles[4] to encompass legislation for all key areas of AI, from classifying prohibited and high-risk AI systems, to a highly detailed examination of transparency obligations and governance and human oversight (Titles II – IV). The Act also considers risks for information sharing and market surveillance (Title VIII) and lays down the rules for confidentiality and penalties (Title X), which could see fines of up to 7% of an organisation's global turnover. But its not all doom and gloom, as Title V explores 'Measures in support of Innovation'.

KEY ASPECTS

INSIDE THE EU AI ACT: DEFINING AND MANAGING RISK

To protect the maintain the tradition of fundamental rights, the EU proposal recommends a risk-based approach. If the characteristics of an AI model are recognized as 'opacity, complexity, dependency on data, autonomous behaviour' then from this definition, 3 levels of risk have been identified with specific characteristics: Unacceptable Risk, High Risk, and Low or No Risk

Unacceptable Risk

Unacceptable Risk AI systems are systems considered a threat to people and will be banned. They include:

- Cognitive behavioral manipulation of people or specific vulnerable groups: for example, voice-activated toys that encourage dangerous behavior in children.
- Social scoring: classifying people based on behavior, socio-economic status, or personal characteristics.
- Real-time and remote biometric identification systems, such as facial recognition

There may be some exceptions: For instance, "post" remote biometric identification systems where identification occurs after a significant delay will be allowed to prosecute serious crimes but only after court approval. Facial recognition did prove a particularly tendentious point of discussion during the passing of the Act but, as noted in the press conference afterwards, the EU Parliament approved all aspects of the Act resoundingly.

Despite this, the use of facial recognition in public spaces will certainly be one topic that will require careful monitoring during the trilogue negotiations. Likewise, emotional recognition AI products fall into the 'Unacceptable risk' category to be banned, pending the negotiations, but more innovative and emerging HR technologies which use AI to analyse body-language and facial-expression in interviews could be grouped into the below High-Risk category, meaning accepted but strongly restricted use. Striking the balance between improved efficiencies and insights, with ethical considerations might feel like taking one step forwards and two steps backwards, but it will always elicit a nuanced - and controversial - discussion.

[4] https://artificialintelligenceact.com/

A large omission from the Act is the use of AI-driven technologies in the military. The concerns around such use of AI (from which anyone could imagine a Terminator-like capability), which clearly challenge the human-centric, ethical focus of the Act. It is clear that a considerable amount of discussion, negotiation, and international cooperation has been put into the formulation of this Act, but nonetheless unbridled innovation within military research and development could threaten the effectiveness of the Act in practice.

Indeed, much of the development pursued by national military infrastructures has also overlapped with the private sector, Palantir Technologies being an example of this private/public nexus. However, this is also an area that we shall be observing closely to see how much (and which) Member States seek exemptions for their military research and development on AI systems.

High Risk

'Title III contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons. In line with a risk-based approach, those high-risk AI systems are permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment. The classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation. Therefore, the classification as high-risk does not only depend on the function performed by the AI system, but also on the specific purpose and modalities for which that system is used.'

The above statement has been copied exactly from the Act as this is likely the most contentious area of regulatory focus. High-risk systems will require a level of transparency to ensure that the 'intended purpose' (the use case) can be described, documented, and catalogued by the regulator comprehensively. There is a clear similarity with GDPR: the 'purpose limitation' requires an understanding of why the processing of data is being done and documented justification. This requirement points to how organisations might need to prepare for implementation: it is very probable that risk assessments, documented justification, uses, testing of any AI products are going to be necessary.

Within this category of AI systems that can negatively affect safety or fundamental rights, such systems defined as high-risk will be divided into two groups:

1) AI systems that are used in products falling under the EU's product safety legislation. This includes toys, aviation, cars, medical devices, and lifts.

2) AI systems falling into eight specific areas that will have to be registered in an EU database:

- Biometric identification and categorization of natural persons
- 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 throughout their lifecycle.

There are a range of supporting documents and tools that highlight distinct aspects of risk management assessment and identification. One of these is capAI[5], described by the authors as 'A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act.' The purpose of capAI is to provide '...organisations with practical guidance on how to translate high-level ethics principles into verifiable criteria that help shape the design, development, deployment, and use of ethical AI. This tool can be used to demonstrate that the development and operation of an AI system are trustworthy.'

This emphasis on the 'trustworthiness' of an AI system links back directly to the work of HLEG and has 3 defined criteria: AI systems should be lawful, ethical, and technically robust. This is a truly relevant document for anyone wanting to understand in more detail how an AI System can be categorised within this risk management classification and be compliant with the Act. It is also of note that the Act references specific industries such as Pharmaceuticals and Financial Services as requiring to co-ordinate their regulation with this new legislation. There is also specific reference to stand-alone high-risk AI systems (also identified in Annex III of the Act) that will require the establishing of a new compliance and enforcement system.

Low or No Risk

AI systems defined and classified as low or no risk should comply with minimal transparency requirements that would allow users to make informed decisions. After interacting with the applications, users 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.

An additional mention is provided for Generative AI: In March 2023 the European Parliamentary Research Service (EPRS) published a paper 'General-purpose artificial intelligence'. The preamble stated the following:

'General-purpose artificial intelligence (AI) technologies, such as ChatGPT, are quickly transforming the way AI systems are built and deployed. While these technologies are expected to bring huge benefits in the coming years, spurring innovation in many sectors, their disruptive nature raises policy

[5] https://artificialintelligenceact.eu/assessment/

questions around privacy and intellectual property rights, liability and accountability, and concerns about their potential to spread disinformation and misinformation. EU lawmakers need to strike a delicate balance between fostering the deployment of these technologies while making sure adequate safeguards are in place.'

The subsequent press conference highlighted the concerns around this technology and the potential risks it presents. Within the category of Artificial General Intelligence (AGI), also referred to as 'strong AI', it was observed that this new wave of foundation models has benefitted from the technological development around such tools as Large Language Models (LLMs). The data for these models is scraped, collected, consolidated from a multitude of sources that can then be directed into the models via an Application Programming Interface (API).

In this same press conference it was noted that, even if the AI Act can keep to the aspirational timelines of completing the trilogue before the end of this year, the Digital Services Act will be in place and this provides a number of safeguards, specifically around the content moderation rules e.g. notice and action mechanisms and trusted flaggers that could be applied to AI Systems prior to the AI Act becoming an enforceable regulation.

Additional issues around copyright and source data would also require Generative AI, like ChatGPT, to comply with the following 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.

While the creators of ChatGPT may have their work cut out for them in preparing this information, organisations who use ChatGPT don't need to adjust according to this act. A central concept is that the responsibilities for sensible, ethical, well-governed AI should sit in the hands of the developers and not in the hands of institutions that use AI.

Despite all this obvious effort in creating this risk management classification, one need only consider the 2008 financial crash to appreciate how those unknown unknowns (or Black Swans) could impact these extensive attempts to build in some futureproofing to AI systems that will provide sustained protection to EU citizens. The financial regulations, Solvency II and MIFiD II, were applied after 2008 (as was BCBS239) so at least in this Act there has been a comprehensive attempt to bring a cohesive approach to managing the risks in the use of AI within the EU. The issue of education, awareness, and benefit are clear inducements to protect individual rights but, as we experience so often in Data Governance, the benefits of oversight and adoption can be a really challenging process of persuasion and negotiation.

kubrick

HARMONISING STANDARDS

APPROACHING THE EU AI ACT IN PRACTICE

While the EU AI Act awaits final negotiations, it has been pro-active in providing guidance on how the regulatory requirements should be adopted by all parties concerned. In the press conference that followed the passing of the act in June, this focus on guidance was referred to as a 'lesson-learnt' from the implementation of GDPR, which could only be described as costly and chaotic. Instead, the EU AI Act has moved away the complex legal jargon of GDRP which hindered so many organisations and provides standards, guardrails, and policies from the 'bottom up' to support compliance with the regulation.

The full title of the Act proposal is:

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACT

The section on Standards within the Act's supporting detail describes "Harmonised Standards" as 'an important role in EU legislation by making what are at times vague essential requirements into concrete technical requirements. It is these standards that will specify, for example, what the "suitable risk management measures" mentioned in the AIA include. They are standards specifically designed to support EU legislation, and adhering to them carries a "presumption of conformity" with the essential requirements.'

Within the Act itself the section on Standards, Chapter 5, begins with Article 40:

'High-risk AI systems which are in conformity with harmonised standards or parts thereof the references of which have been published in the Official Journal of the European Union shall be presumed to be in conformity with the requirements set out in Chapter 2 of this Title, to the extent those standards cover those requirements.'

There is a considerable amount of guidance and support on the topic of high-risk AI systems as this is the most challenging area of classification, compliance, and transparency. As noted earlier, the supporting legislation around the EU's Open Data Initiative provides detailed guidance on the oversight of all data related aspects of digital transfers, personal data protections, and the sustained focus on transparency and protecting the EU data citizen. These combined cluster around the term 'trustworthiness', which summarises the overarching focus of these data-centric regulations: How do we (the EU) ensure that the legislation around AI systems and the management of data provide confidence and transparency to our citizens?

[6] https://artificialintelligenceact.eu/standard-setting/

Within the EU's supporting documentation, a process flow diagram has been provided as a template for the development of harmonised standards[6], as shown in Figure i.

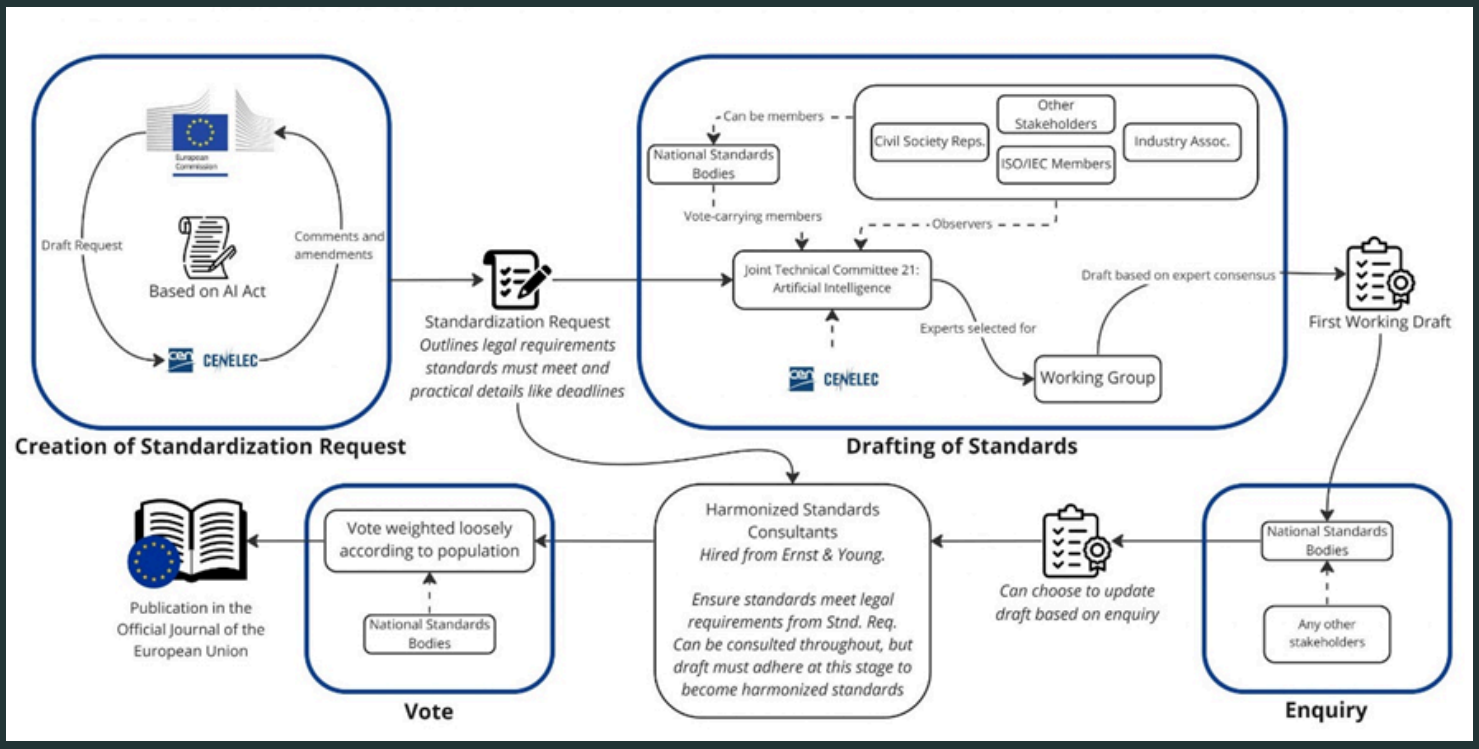

Figure i: The Development Process for Harmonised Standards

Until these standards are properly developed, it's hard to say the impact these will have on businesses. We must sit in wait of confirmed standards to guide us, whilst appreciating the more inclusive approach to regulation harmonised standards provides.

CONCLUDING THOUGHTS

DELIVERING THE PROMISE OF THE EU AI ACT

After exploring some of the most important aspects of the EU AI Act, it is clear that the level of detail, thought, and coverage matches the stated ambition above of the EU legislators to provide a gold standard for the regulation of AI and AI systems. Once the political process has formed the required ratifying level of consensus on this Act, it is anticipated it will provide a level of leadership – or at the very least guidance - for other nation states on the oversight of AI, including the US. At the very least, the global impact of the EU AI Act for businesses working with EU Member States is unavoidable.

While there is a degree of patience that we must exercise while we await the outcomes of the trilogue, the guidance provided by the EU AI Act as it stands, particularly once unpacked, and the learnings from GDPR are more than sufficient for leaders to begin making steps in the right direction for preparedness. While the difficulty of defining and regulating AI systems will remain a clear challenge for the upcoming negotiations, data and technology leaders should be excited for the opportunity for innovation that is also supported by the Act. The increased focus on 'harmonised standards' from preceding regulation is also a promising step towards better support for organisations to reach compliance. However, the nuance and risk of meeting regulation still looms large, heightened by the intent to accelerate negotiations so as to enact the Act to law by the end of 2023.

Once the countdown to compliance begins, organisations must be equipped with the governance domain expertise to communicate between the technology and business teams and embed impactful processes which are cohesively adopted. In the meantime, as the pace of AI development continues to accelerate, organisations must recognise how widespread and wide-ranging AI use will become across their teams by the time regulation comes into effect. The danger of developing or using High Risk or even Prohibited AI systems within siloed teams which are not correctly governed goes beyond hefty financial repercussions. Technology leaders cannot risk being left behind the competition without AI, but poorly governed AI presents a threat to public trust in organisations, industry, and technology at large.

About the author

Simon Duncan,
HEAD OF DATA MANAGEMENT

Simon joined Kubrick in April 2018 to design and deliver their Data Management practice. He utilises his extensive experience from ground-breaking regulatory initiatives and implementations within the financial sector which shaped current understanding of data management and governance. His advisory work includes Basel II, Solvency II, MiFID II, and GDPR, with organisations including FCA, Schroders, and HSBC.

Simon underpins Kubrick's Data Management training with the core principles of governance, ensuring Kubrick's consultants can harness theoretical understanding in order to enhance their comprehension of tools and their purpose with longevity. Simon is an advocate for personal and professional development and runs mindfulness sessions to support the wellbeing of consultants and our HQ staff alike.

@kubrickgroup