

February 24, 2025

European Commission Publishes Guidance on Prohibited AI Practices Under the EU AI Act

One of the first parts of the EU AI Act (the “**Act**”) to enter into force was the prohibition of certain AI practices, which became effective on 2 February 2025. Two days later, on 4 February 2025, the European Commission published its guidelines on the interpretation of these prohibitions as required under Article 96(1)(b) of the Act (the “**Guidance**”)¹. Although the Guidance is non-binding and does not carry judicial authority, it is likely to be persuasive to competent authorities and courts when interpreting the AI practice prohibitions under the Act, and in that sense provides clarity to participants throughout the AI value chain in advance of the Act’s relevant penalties regime for these practices coming into force on 2 August 2025.

This alert summarises the Guidance for businesses considering AI use cases that may fall within the scope of the prohibitions and identifies key practical takeaways.

Practical Takeaways

1. **Prohibitions on certain AI practices apply now.** Although the regulator enforcement powers of the Act do not come into effect (and there will be no market surveillance) until 2 August 2025, the prohibitions apply now and affected parties may enforce them directly in national courts (e.g. by seeking injunctions) although noting that there is no express right to private damages or similar redress under the Act.
2. **Prohibitions are generally interpreted broadly.** The Commission stresses throughout the Guidance that the prohibitions should be interpreted broadly to offer a high level of protection, and that it is important that the necessary elements of each prohibition are not interpreted in such a way as to permit undue circumvention. For example, the Guidance states that the requirement in Article 5(1)(e) that any prohibited scraping be “untargeted” of facial images should not be capable of circumvention by conducting a mix of targeted and untargeted scraping which creates, in the aggregate, an untargeted facial recognition database.
3. **AI practices that avoid prohibition may still be high risk.** Although an exemption may apply, or certain mitigating measures may be implemented (e.g. human oversight, transparency, data governance, impact assessments), such that an AI practice is no longer prohibited under the Act, that does not mean that the relevant AI system is not high risk and therefore subject to the obligations that result from a high-risk designation.
4. **Businesses should consider the real-world capabilities of their AI systems.** Certain of the prohibitions can be engaged regardless of the intention, objective or knowledge of the relevant provider or deployer, provided that the AI system has or may have the relevant proscribed effect. Providers and deployers should therefore take care to assess and monitor the

¹ Available here: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>

likely real-world capabilities and uses of their AI systems in determining whether a prohibition may apply and therefore what mitigation steps to take to seek to avoid the prohibition being engaged (e.g., contractual or technical guardrails).

5. **Most general exclusions from the Act still apply.** Participants in the AI value chain should not forget that the general exclusions to the applicability of the Act are still available to permit the use of otherwise prohibited AI practices (e.g., in the context of research, testing or development activities of AI systems), save that the exclusion for AI systems released under free and open-source licences does not apply to permit prohibited AI practices.
6. **Do not overlook compliance with other relevant laws.** Compliance with other European Union legislation such as the GDPR, the EU’s Digital Services Act, consumer protection and other laws (which may mandate certain transparency, harm mitigation or consent requirements, and/or provide legitimate purposes for observing people) can be a factor in determining whether an AI practice is prohibited. However, the Guidance reiterates that compliance with one piece of EU law does not generally remove the need to comply with overlapping laws unless stated to the contrary.

Summary of Guidance

Article 5 of the Act entirely prohibits eight AI practices (described in more detail in sections 2 and 3 below) on the basis that such practices inherently violate fundamental rights of the European Union (including under the Charter of Fundamental Rights of the EU)². In the Guidance, the Commission initially discusses certain concepts that apply to all such prohibitions and then systematically considers each of the prohibited practices in detail.

1. **General Concepts relating to Prohibited Practices.** For seven of the eight prohibitions, the relevant AI system must be: (i) placed on the market; (ii) put into service; or (iii) used, in each case, within the EU market (the exception being the Article 5(1)(h) prohibition on the use of AI systems for certain real-time remote biometric identification, which specifically requires “use” of the AI system). The Guidance advocates a broad interpretation of each of these gateway requirements (based on definitions used in the Act where available):

Placing on the market	the first supply of an AI system for distribution or use in the course of commercial activity, whether in return for payment or free of charge. ³ The Guidance states that the means of that supply is irrelevant and will include supply through APIs, via cloud, direct downloads, as physical copies, or embedded in physical products
Putting into service	the supply of an AI system for first use directly to a deployer or for the provider’s own use for its intended purpose, where the intended purpose is the use intended by the <i>provider</i> (including by reference to the information it supplies and the broader context). ⁴ The Guidance states that this therefore covers both supply for first use to third parties, and in-house development and deployment
Use	understood in a broad manner to cover any use which post-dates an AI system being placed on the market or put into service (including integration of the AI system into more complex systems and misuse of the AI system (where reasonably foreseeable or not))

The Guidance also clarifies that each of the Act’s prohibited AI practices should be understood to apply regardless of whether the relevant effect applies to an **individual** or a **group** of individuals, other than under Article 5(1)(g) which specifically only prohibits the use of biometric categorisation system to categorise individuals (not groups).

Finally, the Guidance restates the concepts of “provider” and “deployer” of an AI system, generally quoting the wording of the Act, but clarifies that: (a) **authority** for the purpose of defining a deployer means assuming responsibility over the decision

² Available here: https://www.europarl.europa.eu/charter/pdf/text_en.pdf

³ Art. 3(9) of the Act

⁴ Art. 3(11)-(12) of the Act

to deploy the system and manner of its actual use; (b) **individual employees** that act within company procedures are not the deployer; and (c) operators of AI systems may **fulfil one or more roles concurrently**.

2. **Prohibitions on Influencing: Harmful Manipulation, Deception and Exploitation.** The prohibitions in Articles 5(1)(a) and 5(1)(b) focus on the use of AI systems to influence people. Specifically, the Act prohibits the placing onto the market, putting into service or use of an AI system that: (i) deploys subliminal, purposefully manipulative or deceptive techniques (Article 5(1)(a)); or (ii) exploits any vulnerabilities due to age, disability or specific social or economic situation (Article 5(1)(b)), in each case, with the objective or effect of materially distorting behaviour such that significant harm is reasonably likely.

These two prohibitions are complementary and concern the covert influence of certain practices and their impact on individuals’ cognitive autonomy and the protection of vulnerable persons against exploitation. As such, each prohibition shares certain conditions that the Guidance seeks to clarify:

- The practice must have the **objective or effect of materially distorting behaviour**. The Guidance states that this: (a) does not necessarily require intent (given that an effect is sufficient); (b) implies an “appreciable impairment” (which is characterised in a substantial reduction in the ability to make an informed and autonomous decision, thereby causing behaviour or a decision which would not otherwise have occurred) which goes beyond minor or negligible influence; (c) goes beyond ‘lawful persuasion’; and (d) may, by analogy with CJEU decisions concerning the Unfair Commercial Practices Directive (Directive 2005/29/EC), be satisfied if material distortion is merely likely / capable of resulting. In relation to ‘lawful persuasion’, the Guidance opines that, in contrast with prohibited techniques that often exploit psychological weaknesses or cognitive biases, lawful persuasion “operates within the bounds of transparency and respect for individual autonomy”, including presenting arguments in a way that appeals to reason and emotion, provided that relevant and accurate information is provided to ensure informed decision-making.
- The distortion must be **reasonably likely to cause a harmful effect**. The Guidance explains that there must be a causal link between the distortion and the harm, such that the harm is plausible/reasonably likely as a result of the distortion. Factors to consider include: (a) whether an objective provider or deployer could have reasonably foreseen the harm; (b) whether the provider or deployer implements appropriate preventative and mitigating measures (including user controls and safeguards), is transparent in how the AI system operates and its capabilities, is compliant with other relevant legislation, and adheres to professional due diligence practices and industry standards; and (c) the extent to which external factors outside the control of, or reasonably foreseeable by, the provider/deployer contribute to the harm.
- The harmful effect that is caused must be **significant harm**. The harm can be physical (personal injury or property damage), psychological or financial, can be direct or indirect, and/or can be combination of harms which exacerbate the overall impact. However, there must be a significant adverse impact, to be determined on a case-by-case basis based on the severity, scale, intensity, duration and reversibility of the harm, the context and cumulative effects, and the affected person’s vulnerability.

The Guidance also provides commentary on the specific requirements of each prohibition:

Subliminal techniques	techniques that operate beyond (below or above) the threshold for conscious awareness such that the person remains unaware of the influence, operation, or the effects on their decision-making (and can be visual, auditory and/or temporal)
Purposefully manipulative techniques	techniques that are designed or objectively aim to influence, alter or control behaviour in a manner that undermines autonomy and free choice (whether subliminally or otherwise), although it is not necessary for the provider/developer to intend the techniques to cause harm or be manipulative/deceptive. Indeed, where the AI systems themselves learn manipulative/deceptive techniques because of the contents of the data on which they are trained, or because reinforcement learning, these are prohibited if they meet the other relevant conditions even if there was not human intention
Deceptive techniques	techniques that subvert or impair autonomy, decision-making or free choice in ways that a person is not consciously aware of or, where aware, is deceived by or unable to resist,

including via the presentation of false or misleading information. Again, this covers AI systems that implement such techniques without human intention

Vulnerabilities

encompasses a broad spectrum of categories, including cognitive, emotional, physical and other forms of susceptibility that can affect the ability of person to make informed decisions or otherwise influence their behaviour (although the Guidance provides a reminder that, as required under the Act, the vulnerability must be the result of age, disability or specific social or economic situation (examples of the final category being persons living in extreme poverty, and ethnic and religious minority groups)). Interestingly, in relation to age, the Guidance states that children under 18 would be considered in the vulnerable group, which contrasts to the GDPR where member states are permitted to reduce the age that an individual is considered a child to as low as 13.

In each case, compliance with the Act’s transparency obligations can assist a provider or deployer in minimising the likelihood of engaging either prohibition (e.g., in relation to an AI chatbot or deep fake generative system) but that is not guaranteed.

3. **Prohibitions on Observing: Categorisation, Profiling and Evaluation.** The remaining six prohibitions thematically are directed towards the use of AI systems to observe people. The Act prohibits the placing onto the market, putting into service or use of an AI system:

- **Article 5(1)(c)** – for the *evaluation or classification* of people over a *certain period of time* based on their *social behaviour or known, inferred or predicted personality characteristics* (i.e. social scoring), with such social scoring leading to *detrimental or unfavourable treatment in unrelated social contexts* and/or that is *unjustified or disproportionate* to the gravity of the evaluated or classified behaviour. The Guidance goes into significant detail but a few key takeaways are:
 - whereas **evaluation** suggests an involvement of some form of an assessment or judgement, **classification** is broader and need not necessarily lead to an evaluation;
 - the requirement for evaluation/classification over a **certain period of time** suggests that the underlying behaviour must not be limited to one-time behaviour;
 - the behaviour being evaluated/classified can be in a **public, private and/or business** contexts;
 - **social behaviour** is a broad term that includes actions, habits, interactions within society, and which usually covers behaviour related to data points from multiple sources;
 - **personal** and **personality characteristics** are synonymous with each other and with the concepts of “personality traits and characteristics” (Article 5(1)(d)), and which can cover a variety of information relating to a person);
 - personal/personality characteristics that are: (a) **known** are those based on an input to the AI system and in most cases verifiable information, (b) **inferred** are based on information inferred from other information (usually by the AI system), and (c) **predicted** are estimated based on patterns with less than 100% accuracy;
 - the social scoring must **lead to** (not necessarily solely, but it must play a sufficiently important role in causing) detrimental or unfavourable treatment;
 - the organisation producing the social score can **differ** from the organisation using the score; and
 - **unfavourable treatment** means that a person is treated less favourably compared to others (including by not receiving the benefits that another does) and therefore does not necessarily require harm or damage, whereas **detriment** requires harm or damage.
- **Article 5(1)(d)** – for making *risk assessments* of people in order to *assess or predict the risk of criminal offences* based *solely* on profiling or assessing their personality characteristics (save where used to support human assessment which is already based on objective and verifiable facts). The Guidance notes that this includes risk assessments conducted **at any stage of law enforcement** and includes relevant activities conducted by **private entities** acting on behalf of law enforcement authorities that are entrusted with certain specific law enforcement tasks (but does not cover private

entities' ordinary course of business where risk assessments relate to the risk of criminal offences as a purely accidental and secondary circumstance).

- **Article 5(1)(e)** – that *creates or expands facial recognition databases* through *untargeted scraping* of facial images from the internet or CCTV footage. The Guidance notes that: (a) **untargeted scraping** requires scraping (usually via the use of web crawlers, bots or other context-extraction methods) without a specific focus on a given individual or group, further explaining that merely abiding by automatic no-scraping opt outs does not constitute “targeted” scraping; (b) **existing facial databases** (even if populated by untargeted AI scraping) are not prohibited so long as they remain static; and (c) it is sufficient that the database **can be used** for facial recognition (and that need not be its sole purpose).
 - **Article 5(1)(f)** – to *infer emotions* of natural persons in the *workplace or education institution* (save where intended for medical or safety reasons). The Guidance notes that: (a) although the prohibition does not reference “emotion recognition systems” and only references “inference”, it should nevertheless be understood to encompass **emotion recognition systems** and capture AI systems that **identify** emotions (by processing biometric data and directly comparing with previous programming); (b) the inference or identification must be based on **biometric data**; (c) the inference can be based on data not directly collected from the individuals (including machine learning approaches), but must infer emotions of specific people rather than general moods in a workplace; (d) **emotion** does not include physical states (e.g. pain, fatigue) or simply the deduction of readily apparent expressions, gestures or movements; (e) both **workplace** and **education institution** should be interpreted broadly; and (f) the carve-out for **medical and safety reasons** must be narrowly interpreted, applying only to CE-marked systems for therapeutic use or the protection of life and health (not general wellbeing or the protection of other interests) and only where strictly necessary and proportionate.
 - **Article 5(1)(g)** – that is a *biometric categorisation system* to *categorise* individuals based on their *biometric data* to *deduce or infer* their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation (save that the labelling or filtering of lawfully acquired biometric datasets and the categorising of biometric data for law enforcement are not prohibited). The Guidance notes that: (a) a **biometric categorisation system** is a system which assigns individuals to categories (as opposed to merely identifying them or verifying their identity), unless that categorisation is merely ancillary to another commercial service and strictly necessary for objective technical reasons; (b) the purpose or outcome of the system must be to categorise people **individually** (rather than to categorise groups); and (c) the objective must be to deduce or infer **one of the listed characteristics**.
 - **Article 5(1)(h)** – that is a *‘real-time’ remote biometric identification systems*, in *publicly accessible spaces*, for *law enforcement* (unless such use is strictly necessary for the targeted search for specific victims, the prevention of a specific, substantial and imminent threat, or the localisation or identification of a person suspected of certain crimes where registration and authorisation requirements are met). The Guidance notes that: (a) there must be a **reference database** for there to be identification as needed for this prohibition; (b) the concept of “real-time” necessarily covers processing both **instantaneously** and where there is **no significant delay** (to avoid circumvention by retrospective use of such systems); (c) public spaces are agnostic as to the owner of the space and the use of the space, but do not include online spaces, spaces with entry control, or prisons and border control; and (d) this prohibition also covers **private entities** entrusted with specific law enforcement tasks.
4. **Exclusions and High-Risk Systems.** The Guidance notes that, despite otherwise meeting the criteria for prohibition, the exclusions in Article 2 of the Act are still generally available to exclude the system from the scope of the Act, i.e. the exclusions for national security, defence and military purposes, judicial and law enforcement cooperation with other countries, research and development (including specifically scientific research and development), and personal non-professional use. Although the Guidance generally repeats the scope of such exclusions as set out in the Act, it helpfully clarifies that: (i) the national security etc. exclusion requires that to be the **exclusive purpose** of the AI system and so dual use systems do not benefit from the exclusion; (ii) the exclusion for research and development does not cover testing in **real-world conditions**; (iii) the exclusion for personal non-professional **only applies to deployers** and not providers, importers or distributors; and (iv) the exclusion for free and open source licensed systems does not apply to allow prohibited practices.

Critically, the Guidance adds that the fact that an AI system does not meet the conditions to be prohibited under Article 5 of the Act solely because of the specific circumstances in which it is used, it is still likely to be classified as a **high-risk system** and

therefore be subject to the increased obligations on high-risk systems under the Act (e.g. emotion recognition systems that are not used in the workplace or educational institutions). There are **mitigating steps** that providers and deployers can take to increase the chance of an AI system not being categorised as a prohibited system (e.g. providers designing safeguards to prevent reasonably foreseeable harmful behaviour and misuse, implementing contractual obligations with deployers, providing appropriate instructions on use to deployers, and monitoring use (provided this is specific monitoring and in compliance with other EU laws; and deployers complying with such instructions and obligations).

- 5. **Interplay with other EU laws.** The Guidance mentions in a number of places that compliance with the Act does not relieve providers or deployers of AI systems from their obligations under other EU laws (i.e. the fact that an AI system is not prohibited under the Act does not mean it is permitted under all relevant laws). In particular, AI systems must still comply with privacy laws (including the GDPR), the Law Enforcement Directive, the Digital Services Act and consumer protection laws. This is unsurprising given the lengths that have been taken within the EU's suite of new digital and cyber laws to make clear that, unless otherwise stated, these laws are intended to be cumulative for those entities and practices covered.

Conclusion

Although the Guidance is cumbersome (running to almost 140 pages), it provides a helpful steer on a number of concepts under Article 5 of the Act which had until now been unaddressed or lacked clarity. It represents the first of what are expected to be many sets of guidance and codes of conduct that will be released by the Commission in the coming months (including the requirements for high-risk AI systems, responsibilities across the AI value chain, the provisions relating to substantial modification of AI systems and transparency obligations and a general-purpose AI code of practice), and has already been followed by guidance on the definition of 'AI Systems'. These are intended to help businesses that develop, provide or use AI systems within the EU grapple with the requirements of the Act, understand when their systems (or specific use of their systems) may be prohibited, and identify potential steps can be taken to reduce the chance of an outright regulatory prohibition on their system in the EU.

However, the Guidance is also stated to be non-binding and not comprehensive. As such, it is not yet known whether competent authorities will agree with the Commission's views on these core concepts and how they will be interpreted and enforced in practice. We will also not see the start of any competent authority actions until at least 2 August 2025, when those provisions of the Act become applicable, although the Guidance notes that this does not prevent private actions for non-compliance prior to that date. There are also likely to be further nuances or explanations that arise as the Guidance is tested and additional guidance is released. For now, businesses can look to the Guidance as a baseline barometer of the mood of the Commission and start to take steps to adjust accordingly.

* * *

This memorandum is not intended to provide legal advice, and no legal or business decision should be based on its content. Questions concerning issues addressed in this memorandum should be directed to:

Jonathan H. Ashtor
+1-212-373-3823
jashtor@paulweiss.com

John P. Carlin
+1-202-223-7372
jcarlin@paulweiss.com

Katherine B. Forrest
+1-212-373-3195
kforrest@paulweiss.com

Anna R. Gressel
+1-212-373-3388
agressel@paulweiss.com

Nicole Kar
+44-20-7601-8657
nkar@paulweiss.com

Henrik Morch
+32-2-884-0802
hmorch@paulweiss.com

John Patten
+44-20-7367-1684
jpatten@paulweiss.com

Audrey M. Paquet
+1-212-373-2397
apaquet@paulweiss.com

Alex Zapalowski
+44-20-7367-1697
azapalowski@paulweiss.com

Associates Edmund Berney, Scott Caravello, Ali Fazeli-Nia and Jason Gerson contributed to this Client Alert.