ChatGPT

EU AI Act: Overview and Key Provisions

Objectives and Scope

The European Union's Artificial Intelligence Act (AI Act) is the world's first comprehensive legal framework for AI, aiming to ensure **trustworthy, human-centric AI** in Europe. It seeks to safeguard health, safety, and fundamental rights while encouraging AI innovation and investment. The regulation harmonizes rules across all EU member states to prevent fragmented laws and maintain the EU single market for AI.

Scope: The AI Act applies to any **AI system placed on the market, put into service, or used within the EU**, regardless of where the provider is based. This extraterritorial reach means that non-EU companies must comply when offering AI products or services to EU users. The Act defines AI broadly (covering machine-learning, logic- and knowledge-based systems, etc.) and excludes a few domains from its scope: AI systems **developed or used exclusively for military purposes**, those used for pure **research and development**, and most **free and open-source AI components** (with the notable exception that foundational general-purpose models are still covered). In other words, open-source AI and research prototypes are largely exempt to avoid stifling innovation. The AI Act is part of a wider EU digital strategy and is designed to complement existing laws like the GDPR, focusing specifically on AI-related risks and not overlapping with data protection or content moderation rules.

Risk-Based Classification of AI Systems

Central to the AI Act is a **risk-based regulatory approach** that categorizes AI systems into four tiers of risk: **Unacceptable, High, Limited,** and **Minimal or No Risk**. The stringency of requirements and restrictions increases with the level of risk. Table 1 below summarizes these risk levels, examples, and how the Act regulates them:

Risk Level	Description & Examples	Regulatory Treatment
Unacceptable	AI posing a clear threat to safety or fundamental rights (e.g. AI that manipulates vulnerable people, social scoring systems, or certain biometric surveillance tools) [1].	**Prohibited outright** – such systems **cannot be used or sold** in the EU [1]. (See the banned practices listed below.)
High Risk	AI with a serious impact on health, safety, or rights, often in critical domains (e.g. AI in medical devices, transportation, education, employment/hiring, law enforcement, border control, judicial decisions) [2] [3].	**Strictly regulated** – allowed on the market **only if numerous obligations are met**, including risk assessments, high-quality data, human oversight, etc [4]. Subject to conformity assessment and inclusion in an EU-wide **public registry** of high-risk AI.

Risk Level	Description & Examples	Regulatory Treatment
Limited Risk	AI with limited potential for harm but requiring transparency to maintain trust (e.g. **chatbots** or **generative AI** that may be mistaken for human output) [5] .	**Transparency requirements** – e.g. users must be informed they are interacting with AI, AI-generated content must be labeled as such [5] . No stringent pre-approval, but **disclosure obligations** apply.
Minimal or No Risk	AI systems with negligible or routine risk (the vast majority of AI, e.g. spam filters, AI in video games) [6] .	**No specific requirements** under the AI Act [7] – these systems are largely unregulated by this law (general EU product safety laws still apply).

Table 1: EU AI Act risk categories, examples, and regulatory treatment. [8] [5]

Prohibited Practices (Unacceptable Risk AI)

The AI Act **bans a set of AI practices** deemed to pose "**unacceptable risk**" to people's safety or fundamental rights [9] . These represent uses of AI that the EU considers socially harmful or ethically problematic, and they are **outlawed outright** (with only very narrow exceptions). The prohibited AI practices include [1] :

1. **Manipulative or Deceptive AI Systems:** AI that deploys subliminal techniques or exploits cognition to **manipulate people's behavior in ways that cause harm** (e.g. an AI toy covertly influencing a child to engage in dangerous actions) [10] .
2. **Exploitation of Vulnerabilities:** AI that **takes advantage of vulnerable groups** (such as children, the elderly, or persons with disabilities) in order to materially distort their behavior and cause harm [10] .
3. **Social Scoring:** Systems that **rank or classify individuals based on social behavior or personal characteristics** (like government-run social credit scores), leading to unjust or disproportionate treatment [11] . Such blanket social scoring by public authorities is banned [11] .
4. **Predictive Policing:** AI tools used for **predicting criminal behavior or risk** (e.g. algorithms that claim to forecast an individual's likelihood of committing a crime) [12] . These "individual risk assessment" systems in law enforcement are prohibited due to their implications for justice and bias [12] .
5. **Mass Biometric Surveillance (Untargeted Use):** The **untargeted scraping of biometric data** from the internet or CCTV footage to create facial recognition databases is banned [13] . In essence, indiscriminate surveillance AI that aggregates people's biometric information without consent is not allowed.
6. **Emotion Recognition in Sensitive Settings:** AI systems that **detect or infer emotions** of people in contexts like workplaces or schools are forbidden (except for very limited health or safety purposes) [13] . For example, using AI on employees or students to gauge mood or emotion (for performance or disciplinary purposes) is not permitted [14] .
7. **Biometric Categorization of Sensitive Traits:** AI that uses biometric data (facial features, etc.) to **classify people into categories by protected characteristics** – such as race, gender, ethnicity, or political/religious orientation – is prohibited [15] . This aims to prevent automated profiling that could lead to discrimination.
8. **Real-Time Remote Biometric Identification in Public by Law Enforcement:** The use of AI-based systems like live facial recognition in public spaces for law enforcement is generally banned [15] . **Exceptions** are very narrowly defined – for instance, it could be allowed in extreme cases like

searching for a missing child or preventing an immediate terrorist threat, and even then **only with judicial or independent authority approval**. (Post-event “retroactive” facial recognition may be allowed for serious crimes, subject to court warrant.)

These prohibitions took effect early (February 2025) and carry the toughest penalties for violations. The European Commission can update the list of banned practices as AI technology evolves.

Obligations for High-Risk AI Systems

“High-risk” AI systems are those that can significantly affect people’s lives (safety, health, or rights) – for example, AI used in **critical infrastructure, healthcare devices, education or exams, employment decisions, credit scoring, law enforcement analysis, border control, or courtroom decisions** [2] [3]. This category also includes AI components of products already subject to EU safety regulations (toys, cars, medical devices, etc.) if the AI impacts the product’s safety functions. High-risk use cases are explicitly listed in the Act (covering areas like **biometrics, critical services, education, employment, essential services, law enforcement, migration/asylum, and justice**). All such systems must meet strict requirements before and after they reach the market [16].

Pre-Market Compliance: Providers (developers or importers) of high-risk AI must **undergo a conformity assessment** to ensure the system meets the AI Act’s requirements **before deployment** [17]. In many cases providers can self-assess, but for some AI (especially those linked to existing product safety regimes) a **notified body** may need to certify compliance. Additionally, all high-risk AI systems will be **logged in an EU-wide public database** so that regulators and the public know what and where these systems are used.

Key Obligations for High-Risk AI: To lawfully offer or use a high-risk AI system, the following core requirements must be met [4]:

- **Risk Management:** Implement a continuous **risk management system** to identify, evaluate, and mitigate risks throughout the AI system’s lifecycle. Providers must proactively assess foreseeable risks of the AI in its intended context and reduce any potential harm.
- **Data Governance:** Ensure the **quality of datasets** used for training, validation, and testing the AI [18]. Data should be relevant, representative, free of errors, and sufficiently free of bias to minimize discriminatory outcomes [18]. Providers need governance measures to prevent problematic bias and to trace the data provenance.
- **Technical Documentation:** Maintain extensive **technical documentation** and records about the AI system [19]. This includes detailing the model’s design, intended purpose, performance characteristics, limitations, and the steps taken to comply with the Act. Such documentation should enable regulators to understand and audit the AI if needed.
- **Traceability (Logging):** Keep logs and records of the AI system’s operation to ensure **traceability of decisions** [20]. Continuous logging helps in investigating how an outcome was produced and is critical for accountability or if an incident needs review.
- **Transparency & User Information:** Provide **clear instructions and information to users (deployers)** of the AI system [21]. Users must be informed about the AI’s capabilities, limitations, and proper use so they can operate it correctly and safely [21]. In high-risk contexts, people affected by the AI’s decisions may also need to be informed of that use (for example, a person rejected for a loan by an algorithm has a right to know an AI was involved).
- **Human Oversight:** Establish appropriate **human oversight measures** [22]. The design should allow humans to monitor the AI’s functioning and intervene or override when necessary. Humans

in charge of the system may need training to properly oversee the AI. The goal is to prevent the AI from operating as a black-box authority in crucial decisions without recourse to human judgment.

- **Robustness, Accuracy, and Security:** Ensure a high level of **accuracy, robustness, and cybersecurity** for the AI system [22]. The system should be tested and verified to perform reliably and safely, with measures to withstand attacks or manipulation. If the AI is likely to be updated or learn over time, its performance should remain within safe limits.
- **Quality Management:** The provider should implement a **quality management system** (internal processes to ensure consistent compliance in design and development). This organizational step helps maintain standards and handle any changes or improvements to the AI system in a controlled manner.

Post-Market Monitoring and Incident Reporting: Even after a high-risk AI is deployed, providers must monitor its performance and keep authorities informed of serious issues. They are required to **report serious incidents or malfunctions** to regulators (e.g. if the AI led to a serious health/safety issue or legal rights violation). Keeping logs and allowing audits is mandatory, so that if something goes wrong, it can be investigated. Deployers (operators) of high-risk AI also have duties: for instance, they must use the system as intended, maintain logs, oversee its operation, and in some cases conduct their own **fundamental rights impact assessment** before use (especially if they are public sector or providing essential services). If a high-risk AI system is significantly modified or its purpose changes, it may need re-assessment before continued use.

High-risk AI regulation is therefore an ongoing process, not a one-time certification. Failure to meet the requirements or improperly self-categorizing an AI as low-risk when it is high-risk can lead to substantial fines (in the final Act, up to €20 million or 4% of global turnover for such violations, and even higher for banned uses). To assist compliance, providers may follow harmonized EU standards (to be developed) or other best-practice techniques to meet these obligations.

Transparency Requirements (Limited-Risk AI and General-Purpose AI)

For AI systems that fall into the “**limited risk**” category, the AI Act imposes specific **transparency and disclosure obligations** [5]. These measures are designed to ensure people **know when AI is being used** and can trust AI-driven content. Key transparency requirements include:

- **Human-AI Interaction Disclosure:** Whenever a system **interacts with a human**, the person should be informed that they are dealing with an AI and not another human [5]. For example, if a company uses a chatbot or an AI customer service agent, it must clearly notify the user that it is an AI system (unless it’s obvious from context) [5]. This prevents deception and allows users to make informed decisions about how to engage with the system.
- **Labeling AI-Generated Content:** AI systems that **generate or manipulate content** (text, images, audio, video) must ensure the outputs are appropriately **labeled** as AI-generated [23]. If an image or video has been synthetically altered or created (for instance, a deepfake), it should be clearly marked so that viewers know it is not authentic footage. Similarly, AI-generated text published in contexts where readers might assume a human author (e.g. news articles or public communications) should disclose that AI was involved. This requirement tackles concerns about deepfakes and misinformation by mandating visible disclosure [23].
- **Generative AI (General-Purpose AI) Transparency:** Providers of **general-purpose AI models** (like large language models à la ChatGPT) that are not inherently high-risk must still comply with certain transparency and **content governance rules**. Notably, a generative AI should be designed in a way that **prevents it from producing illegal content** (for example, having measures to avoid

illegal hate speech or child exploitation content). Additionally, providers of such models must **publish summaries of the copyrighted data** used for training the AI. In practice, this means if a foundation model was trained on internet data that includes copyrighted works, the provider needs to release a document listing the sources or datasets (at least in summary form) that were used. This is intended to address intellectual property concerns and give downstream users information about the model's training material.

- **Notification of Biometric or Emotion Recognition:** If an AI system performs **emotion recognition or biometric categorization** on people (in contexts where it's allowed), those being analyzed should be informed about it. For example, if in a retail setting an AI camera system analyzes shoppers' faces to categorize them or detect mood, the shoppers should see a notice that such AI is in operation. While many such uses might be high-risk or even prohibited in sensitive contexts (like employment), the Act ensures that any allowed use comes with transparency to the subjects.

These transparency obligations take effect by August 2026, giving organizations time to implement the necessary disclosures. They are considered minimal compliance steps for limited-risk AI – the idea is that when outright bans or heavy oversight aren't warranted, simply keeping users informed can address potential issues of trust. The **onus is on providers** to incorporate these features (e.g., content watermarking, user notifications) into their AI systems. In addition, the Act has introduced an **"AI literacy" initiative** (Article 4 of the Act) requiring that users and society are educated about AI systems, which underpins the transparency requirements – an informed public is better able to handle AI technology responsibly.

Governance and Enforcement

The EU AI Act establishes a multi-tiered **governance structure** to oversee its implementation and enforcement across the Union:

- **European AI Office:** A new EU-level body will coordinate the Act's implementation. This office helps clarify requirements, issue guidance, and facilitate cooperation among member states. It effectively serves as the central hub of expertise for AI regulation in Europe.
- **National Supervisory Authorities:** Each EU member state must designate one or more **national authorities** to supervise and enforce the AI Act at country level. These authorities (often market surveillance or digital regulators) will handle tasks like evaluating high-risk AI compliance, conducting audits or inspections, and investigating complaints. They have the power to prohibit or recall non-compliant AI systems from the market.
- **European Artificial Intelligence Board:** The Act creates an AI Board (comprising representatives of each national authority and the Commission) to **steer the regulation's consistent application**. This Board, supported by the AI Office, can draft opinions, share best practices, and advise on new issues (similar to how the GDPR's European Data Protection Board functions). Additionally, expert groups like a Scientific Panel and an Advisory Forum will provide input on technical or ethical matters.
- **EU Database of High-Risk AI:** To increase oversight, all certified high-risk AI systems will be listed in a public EU database (managed by the Commission). This database improves transparency by allowing anyone (including consumers or civil society) to see what high-risk AI systems are in use in the EU and to identify the responsible providers. It also aids regulators in monitoring compliance and tracking the spread of certain AI applications.
- **Regulatory Sandboxes:** Recognizing the need to foster innovation, the Act encourages member states to create **AI regulatory sandboxes**. These are controlled environments where companies (especially start-ups and SMEs) can **test new AI systems under supervision** before full deployment. National authorities are required to offer such sandbox programs that simulate real-

world conditions. In a sandbox, certain regulatory requirements might be temporarily relaxed or guided by regulators, helping innovators fine-tune their systems and ensuring compliance by design. This approach should help smaller players and researchers experiment without immediately risking penalties, aligning with the Act's goal to support AI development in Europe.

- **Codes of Conduct:** For AI systems that are **not high-risk**, the Act encourages the development of **voluntary codes of conduct**. The European AI Office and the Commission will facilitate industry stakeholders in creating codes that promote **ethical and trustworthy AI beyond the legal minimum**. For example, a code of conduct might cover environmental sustainability of AI or best practices for transparency in medium-risk applications. While following such codes is optional, companies may adopt them to demonstrate leadership in responsible AI, and regulators view this favorably as it can raise the overall standard of AI systems across the board.

Enforcement and Penalties: The AI Act will be enforced through a combination of proactive oversight (e.g. conformity assessments, market surveillance) and reactive measures (investigating complaints or incidents). **Non-compliance can lead to steep fines**, analogous to the GDPR's penalty structure. The maximum fines depend on the severity of the violation:

- **Banned Practices:** Engaging in an AI use that is prohibited (unacceptable risk) can incur fines up to **€35 million or 7% of global annual turnover** (whichever is higher). This is the harshest tier, reflecting the egregiousness of deploying outlawed AI.
- **High-Risk Requirements Violations:** Failure to meet obligations for a high-risk AI system (e.g. no risk controls, no documentation, misrepresenting the system's compliance) carries slightly lower – but still very significant – fines (e.g. up to €20 million or 4% of turnover, based on earlier drafts). The final law set tiered penalties (for instance, up to €15 million or 3% turnover for certain lesser offenses, according to official summaries).
- **Data and Transparency Violations:** Providing incorrect, incomplete or misleading information to regulators, or failing to fulfill transparency obligations, can result in fines in the range of €7.5 million or 1.5%–2% of turnover.

These penalties create a strong incentive for companies to comply. In addition to fines, authorities can order **withdrawals or recalls** of AI systems, or impose corrective actions. Individuals and companies will also have avenues to **file complaints or seek remedies** if they are harmed by an AI system – the Act empowers national regulators to handle such complaints about AI systems posing risks or violating the rules [17].

Timeline: The AI Act was **formally adopted in mid-2024** and entered into force on 1 August 2024. However, its provisions apply in a staggered timeline to give stakeholders time to adapt. The ban on unacceptable-risk AI and certain user awareness measures began applying on **2 February 2025**. The rules for **general-purpose AI (foundation models)** and the establishment of the new governance structures took effect on **2 August 2025**. The bulk of obligations (e.g. for high-risk AI systems) will become **fully applicable by 2 August 2026** (24 months after entry into force), with some specific high-risk use cases embedded in other regulated products given an extension to 2027. This phased rollout means businesses should use the lead time to achieve compliance, as enforcement will ramp up in stages rather than all at once.

Implications for Businesses and Developers

The EU AI Act has **far-reaching implications** for companies that develop or deploy AI systems, as well as for software developers and data scientists in the AI field. It essentially mandates a shift toward "**responsible AI**" practices. Key implications include:

- **Compliance as a Core Requirement:** Businesses must now **treat AI compliance as part of their operational risk management**. Companies should **inventory the AI systems** they use or offer, determine which risk category each falls into, and ensure that each system meets the corresponding requirements. For some firms this may mean instituting new internal processes, like AI ethics committees or model audit procedures, to regularly assess AI systems' impact on rights and safety. For high-risk AI, the development lifecycle will need to incorporate regulatory checkpoints (e.g. producing documentation, engaging external auditors if needed, etc.), potentially extending time-to-market.

- **Global Reach – Affecting Non-EU Companies:** Given the Act's scope, any AI provider outside Europe that wants access to the EU market must **comply with EU requirements** or risk being shut out. For example, an American AI software company selling to EU clients might have to implement new transparency features and provide EU-conforming documentation for its product. This effectively sets a **global benchmark** – we can expect the EU AI Act to influence AI product standards internationally, similar to how GDPR influenced data privacy practices worldwide. Developers should be mindful that an AI system that is legal and unregulated elsewhere might still be illegal to supply or use in Europe (e.g. an emotion-detection HR tool would be forbidden in the EU context) and adjust their offerings accordingly.

- **Product Design Changes:** Developers will need to build compliance into AI systems from the design phase. For instance, a company creating a generative AI tool must now include features to **label AI-generated outputs** and filters to block illegal content generation. Chatbot interfaces must clearly signal that they are automated [5]. These requirements may require additional R&D and could slightly alter user experience (for example, watermarks on AI images, pop-up notices in chatbots, etc.). Nonetheless, they can also be viewed as trust-building features that make users more comfortable engaging with AI.

- **High-Risk AI = Higher Compliance Burden:** Firms working on high-impact AI applications (like AI in healthcare, finance, or public services) will face **significant compliance tasks**. This includes hiring or training personnel for **quality management, legal compliance, and documentation**. Some companies might need to bring in domain experts (e.g. ethicists, safety engineers) to perform risk assessments or bias testing on AI models. Providers of high-risk AI should prepare for possible **third-party audits and certification** processes prior to product launch, which could incur costs. They'll also have ongoing duties like **maintaining logs, performing post-market surveillance, and updating risk measures** as their AI is used in the field. While burdensome, these practices align with good governance and may reduce liability by catching issues early.

- **Penalties for Non-Compliance:** The substantial fines and enforcement powers under the Act mean that AI compliance is not just a theoretical checkbox – it's backed by real financial and reputational risk. Companies could face multi-million euro fines (up to 7% of global turnover in worst cases) if they ignore the rules. This threat of penalties will likely drive more cautious behavior: businesses might proactively drop or avoid certain AI functionalities that are risky or hard to make compliant. For example, an analytics firm might disable any emotion recognition feature in its software for EU clients, rather than risk falling foul of the ban. In strategic terms,

legal compliance in AI becomes a competitive factor – those who manage it will have access to the EU market, those who don't may be excluded or punished.

- **Impact on AI Supply Chain and Contracts:** Organizations that deploy AI solutions (e.g. a bank using a third-party AI for credit scoring) will demand assurances from their vendors. We can expect AI contracts to include clauses requiring the provider to **warrant compliance with the EU AI Act**. Businesses might also require suppliers to share necessary documentation or audit rights so that the business itself can prove compliance if scrutinized. In turn, software providers will likely create **compliance documentation "kits"** for their enterprise AI products to satisfy client and regulator expectations (similar to how cloud service providers share GDPR compliance info). This dynamic pushes compliance responsibilities across the supply chain: everyone involved in bringing an AI system to users is accountable to some degree.

- **Opportunities for Trust and Innovation:** On the positive side, the Act's emphasis on **trustworthy AI** could become a market differentiator. Companies that achieve compliance early can advertise their AI as **"EU AI Act compliant"**, giving customers and investors confidence in its safety and ethics. This might be especially important in sensitive sectors like healthcare or finance where trust is paramount. Furthermore, the availability of **regulatory sandboxes** provides an opportunity for innovators: startups can collaborate with regulators in sandbox programs to shape and test cutting-edge AI (like autonomous vehicles or advanced robotics) in a safe setting. Insights gained there can not only ensure compliance but also improve the product. The Act also explicitly promotes **research and innovation exemptions** (for example, allowing use of otherwise high-risk AI in controlled trial settings) and **funding for AI testing**, which businesses and developers can leverage to advance technology within the new rules.

- **SMEs and Open Source Developers:** Small and medium enterprises may have fewer resources for compliance, so the Act's phased timeline and sandbox support are particularly beneficial to them. SMEs are encouraged to participate in sandboxes to get guidance. The Act's exemption for **open source AI** components means that independent developers and researchers can continue to publish AI tools and code without immediately facing compliance costs, as long as those are not placed into commercial service. This carve-out was included to sustain the open innovation ecosystem. However, if an open-source model becomes widely used or integrated into products, compliance obligations might kick in for whoever deploys it. Developers of large general-purpose AI models (even in open source) should note that **foundation models are treated specially** – highly capable models might be classified as having "systemic risk" and thus attract extra scrutiny and obligations (like performing evaluations for societal impact and reporting issues). In practice, this means even research labs and big open-source projects need to watch the evolution of rules around **high-impact foundation models**.

In summary, the EU AI Act imposes a new regulatory discipline on AI similar to the impact of the GDPR on data – companies must build compliance into their AI products and processes. While this presents challenges and may increase short-term costs, it also lays a stable groundwork for **AI adoption in the long run by building public trust**. Businesses and developers that align early with the Act's principles (safety, transparency, accountability) are likely to gain a competitive edge in the European market. The Act is not only about restrictions – it also explicitly aims to **"guarantee safety and fundamental rights ... while strengthening uptake, investment and innovation in AI across the EU"**. By setting clear rules, it attempts to create a level playing field and avoid a race to the bottom in AI practices. Developers and companies should follow official guidance from the European Commission (which is continuously being developed to clarify technical details) and engage with industry groups to stay updated. The EU AI Act marks the beginning of a new era of AI governance, and understanding its framework will be crucial for anyone building or selling AI systems in the years ahead.

Sources:

- European Commission, Shaping Europe’s Digital Future – AI Act [9] [24] [5] .
- European Parliament, EU AI Act: First regulation on artificial intelligence.
- Center for Security and Emerging Technology (Georgetown CSET), The EU AI Act: A Primer.
- BSR (Business for Social Responsibility), The EU AI Act: Where Do We Stand in 2025? [14] .
- IBM, What is the EU AI Act?.

---

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [15] [16] [18] [19] [20] [21] [22] [23] [24] AI Act | Shaping Europe’s digital future
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[14] The EU AI Act: Where Do We Stand in 2025? | Blog | Sustainable Business Network and Advisory Services | BSR
https://www.bsr.org/en/blog/the-eu-ai-act-where-do-we-stand-in-2025

[17] EU AI Act: first regulation on artificial intelligence | Topics | European Parliament
https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence