thws Technical University of Applied Sciences Würzburg-Schweinfurt

FAKULTÄT INFORMATIK UND WIRTSCHAFTSINFORMATIK

LECTURE NOTES

The EU AI Act

Risk-based rules for artificial intelligence

MODUL
Introduction to the Ethics and Regulation of AI

SEMESTER
WS 26/27

VERSION
1.0

Prof. Dr. Markus Oermann
Instructor
markus.oermann@thws.de

Technische Hochschule Würzburg-Schweinfurt
Stand: July 23, 2026

Contents

Building on the previous unit, where we weighed self-regulation and internal codes of ethics against their limits, we now turn to the hardest instrument in the toolbox: binding EU law. Voluntary commitments can shape a corporate culture, but they cannot be enforced against a firm that decides to ignore them. The EU AI Act is the European legislator's answer to exactly that gap. It regulates a foundational technology of the digital society and its applications, and it does so with an eye on safety, health and fundamental rights.

01
KAPITEL

Background and legislative history

1.1 A legal framework for trustworthy AI

The EU AI Act, formally Regulation (EU) 2024/1689, is part of a larger Commission package intended to support *trustworthy* artificial intelligence. Alongside the Regulation itself, that package includes the AI innovation package for SMEs and start-ups and the Coordinated Plan on AI. The Commission markets the AI Act confidently as the **first comprehensive legal framework for AI in the world**, even though the Regulation itself leaves several central questions open. This first-mover role is deliberate: the EU wants to set global standards.

One double delimitation matters from the outset. The AI Act contains **no** rules on civil liability for the outcomes of AI, and no substantive intellectual-property provisions either. It is product-safety and fundamental-rights law, not civil liability law. Anyone harmed by an AI system therefore finds their claims not here but in general and specialised liability law, the subject of the next unit.

1.2 From proposal to regulation in force

The process stretched over more than three years. On **21 April 2021** the Commission tabled its proposal; the project counted as one of the most important regulatory undertakings of the 2019-2024 term. After adoption by Parliament (March 2024) and Council (May 2024), the text was published in the Official Journal on **12 July 2024**. The AI Act **entered into force** on **1 August 2024**.

02
KAPITEL

Scope and definitions

2.1 Territorial: a broad marketplace principle

The territorial scope is drawn deliberately wide (Art. 2 AI Act). The Regulation applies to **all providers, deployers, manufacturers, importers and distributors** who place AI models, AI

systems or AI-based services on the Union market or integrate them into their products, **regardless of where they are established and regardless of where their servers sit**. What counts is not where the company is based, but where its output meets the market. The declared aim is the **Brussels effect**, the global reach of EU law. That creates tension in particular with the United States, whose regulatory philosophy, at least at the federal level, is markedly more restrained.

2.2 Substantive limits

As wide as the territorial reach is, so wide are the substantive exemptions (Art. 2 AI Act). Excluded are AI for **military, defence or national-security purposes**, AI used by third-country authorities, AI used exclusively for **scientific research and development**, and pure research, testing and development activities prior to market placement. Also outside scope is the use of AI in the course of a **purely personal, non-professional activity**: the Regulation does not address the private end user. Free and open-source AI is privileged, though not where prohibited, high-risk or generative systems are concerned.

EXERCISE · TERMS

The **marketplace principle** attaches to where the output meets the market, not to where the provider is established; a purely **personal, non-professional** use of AI falls outside the scope.

2.3 What is an “AI system”?

The key definition sits in Art. 3(1) AI Act. An **AI system** is a *machine-based system* designed to operate with *varying levels of autonomy*, that may exhibit *adaptiveness* after deployment, and that, for *explicit or implicit objectives*, infers from the input it receives how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments. Three features carry the definition: **autonomy**, **adaptiveness** and the **inference of outputs** from inputs. Classic, deterministically programmed software therefore falls outside the concept. The wording deliberately tracks the **OECD definition of AI systems** to secure international interoperability. To aid interpretation, the Commission’s AI Office published guidelines on 6 February 2025.

DEFINITION

AI system (Art. 3(1) AI Act)

A machine-based system that operates with varying autonomy, may adapt after deployment, and infers from inputs how to generate outputs (predictions, content, recommendations, decisions) that influence physical or virtual environments.

03
KAPITEL
The risk-based approach

The core idea of the AI Act is simple: instead of treating every AI alike, the Regulation grades its requirements by the assumed **risk** to safety, life and fundamental rights. The result is a pyramid of four tiers. The higher the risk, the denser the obligations; the vast mass of everyday AI stays unregulated.

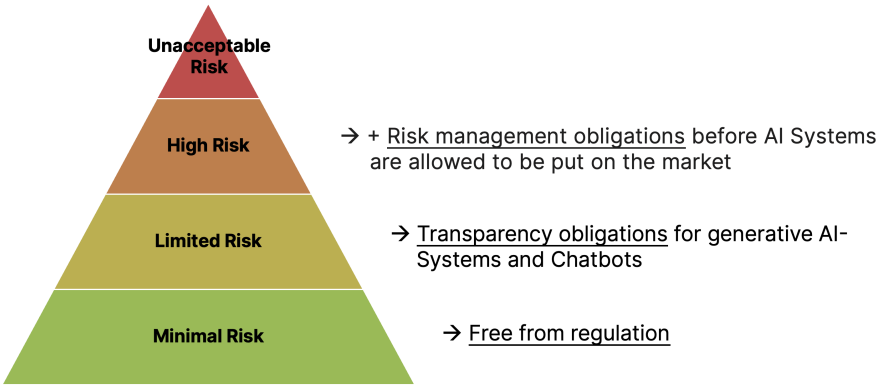

Figure 1: The four-tier risk pyramid of the AI Act: unacceptable risk is prohibited outright; high-risk systems face risk-management obligations before they may be placed on the market; limited-risk systems, such as generative AI and chatbots, face transparency obligations; minimal-risk AI remains free from regulation (own illustration).

Test this logic in the classifier below by sorting concrete use cases into their tiers.

04
KAPITEL
Prohibited practices (Art. 5 AI Act)

At the apex of the pyramid sit the practices sometimes called, a little dramatically, the “deadly sins”. Art. 5(1) AI Act bans AI practices deemed a threat to safety, life and fundamental rights. They include in particular:

- **manipulative AI** that uses subliminal techniques to cause significant harm (lit. a);
- AI that exploits **vulnerabilities** (for instance due to age, disability or economic situation) to distort behaviour (lit. b);
- comprehensive, cross-context **social scoring** (lit. c);

- **predictive policing** that forecasts crime solely on the basis of personal profiling (lit. d);
- **emotion recognition** in the workplace and in education (lit. f);
- certain forms of **biometric categorisation** and the **untargeted scraping** of facial images to build databases.

To interpret these bans, the AI Office likewise issued guidelines on 4 February 2025. The **2026 Digital Omnibus** (see the deep dive further below) adds a further prohibition: AI systems that generate or manipulate **non-consensual intimate images ("nudifiers")** or **child sexual abuse material**, with a transition period until **2 December 2026**.

4.1 Special case: real-time remote biometric identification

The thorniest point was **live facial recognition for law enforcement**. Here the positions clashed head-on: Parliament wanted a blanket ban, the Council was against it. The result is a **compromise** (Art. 5(1)(h), (2)-(5) AI Act). Real-time remote identification in publicly accessible spaces for law-enforcement purposes is **prohibited in principle**, but permitted subject to authorisation in narrow exceptions: to search for a missing person, to avert a concrete and imminent terrorist threat, and to locate perpetrators or suspects of serious crime. Outside these exceptions, remote biometric systems are classified as high-risk and regulated strictly.

QUICK CHECK

A school application is meant to read pupils' emotions during class in order to measure "attention". How is this classified under the AI Act?

High-risk AI with conformity assessment
Prohibited practice (emotion recognition in education)
Limited risk with a transparency duty
Minimal risk with no specific requirements

05

KAPITEL

High-risk AI (Art. 6 et seq. AI Act)

Below the threshold of prohibition sits the true centre of gravity of the Regulation. The rules on high-risk AI have their roots in **product-safety law** and transfer its logic to AI.

5.1 Two gateways into the high-risk class

Art. 6 AI Act knows two points of attachment. Under **paragraph 1**, a system is high-risk if it is used as a **safety component** of a product that is already subject to EU safety law and requires **third-party conformity assessment** (Annex I, for instance vehicles, lifts, toys, medical devices). Under **paragraph 2**, it is high-risk if its intended purpose is listed in **Annex III** and, on a case-by-case assessment, it poses risks to health, safety or fundamental rights. Annex III names, among others,

critical infrastructure, education (such as exam scoring), employment (such as **CV screening** in recruitment), essential services (such as **credit scoring**), law enforcement, migration, and the administration of justice and democratic processes.

5.2 The corrective in Art. 6(3)

So that not every trivial application in an Annex III field becomes high-risk AI, **Art. 6(3) AI Act** provides a corrective: a listed system is *not* high-risk if it poses **no significant risk** to health, safety or fundamental rights, in particular because it **does not materially influence the outcome of decision-making**. Systems that merely *support* human decisions, without taking them over or shaping them decisively, are thus excluded. Providers must, however, **document** their assessment and register the system (Art. 6(4)).

5.3 Providers and deployers

Who is regulated, and how, depends on the role in the value chain. The **provider** is the person who develops an AI system or model, or has it developed, and places it on the market (in everyday terms, the manufacturer). The **deployer**, by contrast, uses the system under its own authority (beyond purely private use). The provider originally determines the intended purpose and can exclude unwanted uses. Decisive is the **reassignment rule** of Art. 25 AI Act: anyone who markets a system under their own name, substantially modifies it, or through actual use puts it to a high-risk purpose **steps legally into the provider's position** and takes on its obligations. **Clarifying one's own role is therefore the very first compliance step.** Test in the simulator below which circumstances turn a deployer into a provider.

DEFINITION

Provider vs. deployer

The provider develops an AI system or model and places it on the market; the deployer uses it under its own authority. Under Art. 25 a deployer can become a provider by rebranding, substantial modification, or a change of intended purpose.

5.4 The obligations of providers (Art. 8-15 AI Act)

For high-risk systems a dense catalogue of obligations applies across the whole lifecycle: a **risk-management system** with risk analysis and mitigation before and after market entry (Art. 9); **data governance**, meaning training only on qualified, representative datasets (Art. 10); **technical documentation** (Art. 11); **record-keeping** for traceability (Art. 12); **transparency** and clear instructions for downstream actors (Art. 13); **human oversight**, for instance through a stop function (Art. 14); and **robustness, accuracy and cybersecurity** (Art. 15).

5.5 Conformity assessment and post-market obligations

Before a high-risk system may reach the market, it goes through **conformity assessment** (Art. 43-49 AI Act): checking against the requirements, registering stand-alone Annex III systems in

an EU database, signing a declaration of conformity and, finally, the **CE marking**. Substantial changes trigger the procedure afresh.

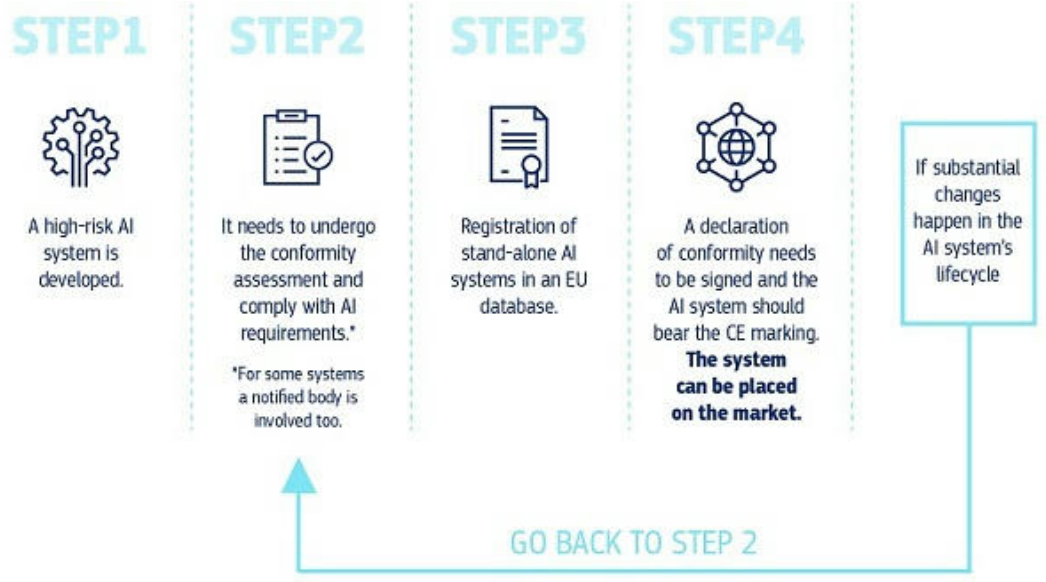

Figure 2: The path of a high-risk AI system to market: from development through conformity assessment, registration in the EU database and the declaration of conformity to placement on the market with the CE marking. Substantial changes lead back to step 2 (source: EU Commission (2025), CC BY 4.0).

After market entry, national market-surveillance authorities take over supervision. **Deployers** must ensure use in line with the instructions, keep logs (Art. 26 AI Act) and, as public bodies and certain private deployers of Annex III systems, carry out a **fundamental-rights impact assessment** before putting the system into use (Art. 27 AI Act). **Providers** set up post-market monitoring; both report serious incidents to the authorities.

CASE STUDY

CV screening in recruitment: high-risk or not?

A medium-sized company introduces an AI that automatically scores incoming applications and produces a ranking of candidates. HR invites only the top ten to interview. How is the system classified, and what follows?
Solution. Applicant management in hiring is an intended purpose listed in **Annex III** (employment). The system is therefore **high-risk AI in principle** under Art. 6(2) AI Act. The corrective of **Art. 6(3)** does not apply here, because the AI *materially influences* the outcome: anyone outside the top ten is effectively filtered out, which is more than mere support. Consequently, the obligations of Art. 8-15 apply (among them data governance against discrimination under Art. 10 and human oversight under Art. 14), as do conformity assessment with CE marking and the deployer obligations under Art. 26. The company is initially a deployer here; were it to substantially modify the system or market it under its own name, it would become a provider under Art. 25 and would bear the full provider obligations.

06
KAPITEL

Limited risk: transparency duties (Art. 50 AI Act)

One tier lower on the pyramid, the AI Act regulates not the *whether* but only the *disclosure*. Art. 50 AI Act requires **transparency** wherever people might otherwise be deceived about the nature of an interaction or a piece of content. The content itself is not governed; what is regulated is solely its recognisability. Providers of large generative systems had already reacted to synthetic content with ad hoc technical restrictions, especially before elections, and platforms such as YouTube and Meta had introduced their own labelling requirements; the European legislator's structural answer is instead a binding, harmonised duty that stops short of an outright ban, which would collide with communication freedoms.

Paragraph	Who	Duty
Art. 50(1)	Provider	Disclose that a system **interacts** with a person (for example a chatbot).
Art. 50(2)	Provider	Mark **synthetic content** in a machine-readable way (watermarking).
Art. 50(3)	Deployer	Inform affected persons of **emotion recognition** or **biometric categorisation**.
Art. 50(4)	Deployer	Label **deepfakes** and AI-generated text on matters of public interest as artificial.
Art. 50(5)	Both	Mark clearly and at first exposure; techniques effective, interoperable, robust, reliable.

Across all of this, the marking must be clear, distinguishable and applied at the latest at the first interaction or exposure, and the technical solutions must be effective, interoperable, robust and reliable so far as technically feasible (Art. 50(5)). Narrow exceptions cover, among others, law-enforcement uses and manifestly artistic, satirical or fictional works. The decisive move in Art. 50 is that the duty attaches to a *role* in the value chain, provider or deployer, rather than to the content alone. Work through the scenarios below to see who owes which disclosure, under which paragraph, and where a narrow exception lets the obligation fall away.

QUICK CHECK

Under Art. 50 AI Act, an AI-generated deep-fake video published by a company must be:

removed from the platform within 24 hours
labelled by the deployer as artificially generated
submitted to a conformity assessment before publication
banned outright as a prohibited practice

6.1 Operationalising Art. 50: the Code of Practice

Art. 50 states obligations but not techniques, and the gap is filled by the **Code of Practice on Transparency of AI-Generated Content** (Bontcheva, Bechmann et al. 2026), drawn up under the Commission's AI Office as a soft-law interpretive aid. It works on two levels. On the **provider** side it requires two layers of marking: digitally signed provenance metadata following the C2PA standard, and imperceptible **watermarking**, together with a detection tool made available free of charge to regulators, media and researchers, with an interoperability deadline of 2 February 2027. On the **deployer** side it introduces a visible EU icon reading "AI GENERATED" or "AI MODIFIED", to be applied for deep fakes and published AI text at the latest on first exposure, with carve-outs for law enforcement, genuine editorial control in journalism, and art or satire.

The Code is a real step toward closing the **provenance gap**, the absence of a reliable chain of custody for digital media discussed in the unit on discourse and disinformation, but it remains structurally limited. Watermarks can be fragile under compression, re-encoding and cross-model remixing; the "analog hole" (re-recording a screen) defeats embedded signals; and there is as yet no settled interoperability standard across providers. Transparency, in short, mitigates the harm of synthetic reality without dissolving it.

DEEP DIVE Deep dive: is Art. 50 a genuine "risk category"?

It is common to draw the AI Act as a four-level pyramid and to slot Art. 50 in as the "limited-risk" tier. That picture can mislead. Art. 50 does not turn on the *size* of a risk at all; it turns on a *capability* of the system, namely its use in communication with people, and high-risk duties continue to apply alongside it (Art. 50(6)). Commentary has also criticised Art. 50(4) for text as poorly drafted: tying the duty to content that "informs the public" on "matters of public interest" produces a paradox in which precisely the contributions most strongly protected by communication freedom are regulated, while gossip is not, and the legislator seems to have had algorithmic journalism in mind rather than the wider field of non-journalistic information. The point for this course is that even a transparency-only instrument raises hard questions of scope and fit, which the AI Office's practice guidelines under Art. 50(7) will have to resolve.

07
KAPITEL

General-purpose AI and AI literacy

7.1 GPAI models (Art. 53-55 AI Act)

Large language and foundation models fit poorly into the application-based risk grid, because their purposes are open. The AI Act therefore addresses them separately as **general-purpose AI (GPAI) models**. Under Art. 3(63) AI Act, a GPAI model is one that (often trained on large amounts of data with broad self-supervision) displays significant **generality**, competently performs a **wide range** of distinct tasks, and can be integrated into many downstream systems. Providers of such models must supply technical documentation focused on training procedures and data and inform downstream actors (Art. 53); providers established in third countries must additionally appoint an authorised representative in the EU (Art. 54).

Tightened obligations apply to GPAI models with **systemic risk** (Art. 55 AI Act): their own risk analysis and mitigation, model evaluation, serious-incident reporting and cybersecurity safeguards. Systemic risk is presumed principally from the **compute used for training**, currently more than 10^{25} FLOP (Art. 51 AI Act), supplemented by further criteria.

DEFINITION

GPAI vs. high-risk AI

High-risk obligations attach to a concrete, risky *intended purpose*; GPAI obligations attach to the *model* itself, whose purposes are open.

7.2 AI literacy (Art. 4 AI Act)

Cutting across all risk tiers is one duty that touches every use. Under **Art. 4 AI Act**, providers *and* deployers must ensure that their staff and the persons they engage have sufficient **AI literacy**: the skills, knowledge and understanding for the competent use of AI, and an awareness of its opportunities, risks and possible harms. Prior technical knowledge, experience, education and the deployment context are to be taken into account. This duty has applied since **2 February 2025** and concerns practically every organisation that uses AI. The Digital Omnibus softens it in 2026 in drafting terms: providers and deployers must now only **promote** the AI literacy of their staff, rather than **ensure** it.

08
KAPITEL

Governance and timeline

8.1 Who watches over the AI Act?

Enforcement rests on a **multi-level governance structure** (Novelli et al. 2024). At the supra-national level sits the **AI Office (DG CNECT)** within the Commission, which supervises GPAI providers in particular; it is flanked by the **AI Board**, an advisory **Forum** and a **Scientific Panel**. At the national level, Member States designate **notifying** and **market-surveillance authorities**; conformity is checked by independent **notified bodies**.

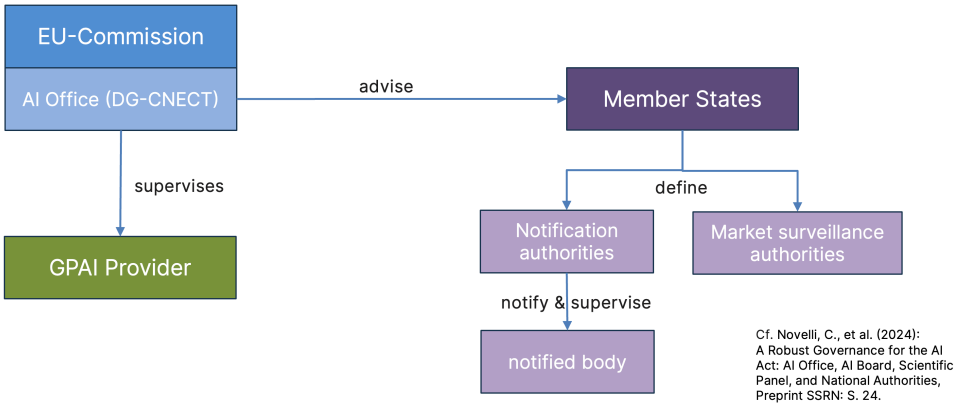

Figure 3: Governance structure of the AI Act (simplified): the Commission's AI Office supervises GPAI providers and advises the Member States, which in turn designate notifying and market-surveillance authorities; the notifying authorities notify and supervise the independent notified bodies (own illustration based on Novelli et al. 2024).

8.2 The staggered applicability (Art. 113 AI Act)

The AI Act does not apply in one stroke but in stages (Art. 113). Four stages were originally envisaged; the **Digital Omnibus** (see the deep dive below) pushed two of them back by more than a year. The timeline that governs after the political agreement of May 2026:

- **2 February 2025**: the **prohibitions** (Chapters I and II) and the AI-literacy duty become applicable (the latter softened by the Omnibus; the new ban on "nudifiers" and AI-generated abuse material follows on **2 December 2026**);
- **2 August 2025**: the rules on **GPAI models**, the authorities, **governance** and **penalties**, which the Omnibus leaves untouched;
- **2 August 2026**: the **transparency duties** of Art. 50 (chatbot disclosure, deepfake labelling); for the marking duty under Art. 50(2), systems already on the market have a transition period until **2 December 2026**;
- **2 August 2027**: the Member States' deadline to set up **regulatory sandboxes** (Art. 57), originally 2026;

- **2 December 2027**: the stand-alone **Annex III high-risk systems** (Art. 6(2)), originally 2 August 2026;
- **2 August 2028**: the **product-related** high-risk systems (Art. 6(1)), originally 2 August 2027.

DEEP DIVE Deep dive: why 2028 for Art. 6(1)?

The product-related high-risk systems (gateway 1) still get the longest grace period, because they must be embedded into the existing conformity procedures of sectoral product-safety law (Annex I, for instance machinery, medical devices). This interlocking with established testing regimes and notified bodies evidently needed more lead time than first estimated: the Digital Omnibus postpones both deadlines by more than a year without changing the underlying logic. The stand-alone Annex III use cases (gateway 2) now start on 2 December 2027, roughly eight months before the product-integrated systems.

09
KAPITEL

Individual rights and penalties

9.1 Two individual rights

At Parliament's insistence, the AI Act goes beyond what is usual in product-safety law and grants two **individual rights**. The **right to lodge a complaint** with a market-surveillance authority arises on suspicion of an infringement and does not require the complainant to be personally affected (Art. 85). The **right to an explanation** belongs to persons whose decisions are taken with the involvement of an Annex III high-risk system (Art. 86).

9.2 Penalties (Art. 99 AI Act)

Member States must provide for effective penalties. The frame is graded: infringements of the **prohibited practices** can be fined up to **EUR 35 million or 7%** of worldwide annual turnover; infringements of other obligations (including for GPAI) up to EUR 15 million or 3%; and the supply of incorrect information up to EUR 7.5 million or 1.5% (whichever is higher).

QUICK CHECK

Which statement about the enforcement of the AI Act is correct?

The AI Act contains its own claim for damages against AI providers.
Infringements of the prohibited practices can be fined up to EUR 35 million or 7% of annual turnover.
Real-time remote biometric identification is banned without exception.
GPAI models are subject to the same obligations as high-risk systems.

DEEP DIVE **Deep dive: the Digital Omnibus, from announcement to agreed postponement**

Shortly after the rules began to apply, the Commission announced adjustments. On **19 November 2025** it tabled the **Digital Omnibus on AI**: simplified documentation duties (also for mid-caps), a clearer legal basis for processing special categories of data for **bias detection**, a strengthening of the AI Office, and above all a postponement of certain implementation deadlines. The trilogue negotiations were tough: a first attempt on 28 April 2026 failed over the conformity assessment for Annex I systems. On **7 May 2026** Council and Parliament reached a provisional agreement, the Council confirmed on 13 May, Parliament approved on **16 June 2026** by 423 votes to 57 (with 174 abstentions), and the Council gave its final assent on **29 June 2026**. Only signature and publication in the Official Journal remained outstanding at the editorial close of this unit; **only with publication do the new deadlines become legally binding**. The episode is instructive: the AI Act is not a finished work but is already being adjusted during its implementation phase, here by extending the lead time rather than lowering the substantive requirements.

010

KAPITEL

From ethics to law, and on to liability

With this, a large arc of the course closes. In the units on ethics we grounded values such as fairness, transparency and human oversight, probed the limits of machine responsibility, and, with Collingridge and Dignum, demanded that ethical requirements be built into the development process **from the start**. The AI Act is the legal answer to precisely that demand: it translates abstract principles into enforceable obligations across the entire lifecycle of a system. What it deliberately leaves out, however, is the question of who pays when an AI causes harm, and who owns what an AI produces. Those two questions, intellectual property and civil liability, are the subject of the final unit.

011
KAPITEL

References

11.1 Literature

- Bontcheva, K., Bechmann, A., Pedreschi, D., De Gregorio, G., Riess, C. & Botan, M. (2026): *Code of Practice on Transparency of AI-Generated Content*. AI Office / European Commission.
- Novelli, C., Hacker, P., Morley, J., Trondal, J. & Floridi, L. (2024): *A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities*. Digital Society 3, 13. https://doi.org/10.1007/s44206-024-00107-0

11.2 Norms & Standards

- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689
- Proposal for a Regulation amending the AI Act as regards certain application dates for high-risk AI systems and amending further legal acts (“Digital Omnibus on AI”), COM(2025) 897, tabled on 19 November 2025; provisional trilogue agreement of 7 May 2026, approved by the European Parliament on 16 June 2026 and the Council on 29 June 2026 (publication in the Official Journal pending at editorial close).
- OECD (2024): *Explanatory memorandum on the updated OECD definition of an AI system*. OECD Artificial Intelligence Papers No. 8. OECD Publishing, Paris.
- AI Office guidelines on the definition of an AI system (6 February 2025) and on prohibited practices (4 February 2025).