UNDERSTANDING EUROPE'S AI REGULATION

The EU AI Act: An Introduction

From prohibited practices to high-risk obligations: a plain-language guide to Europe's AI regulation

Robert Barcik

First edition, July 2026

The EU AI Act: An Introduction · First edition, July 2026

This book is the companion textbook to the online course “EU AI Act Compliance Introduction”.

This book is educational material, not legal advice. The EU AI Act and its guidelines continue to evolve; for decisions with legal exposure, consult your legal representative.

Introduction: How to Read This Book

Somewhere in Europe right now, a bank is deciding whether its loan-approval model counts as artificial intelligence. A marketing team is wondering whether the chatbot on their website needs to announce that it is a chatbot. An HR director is being pitched a tool that ranks job applicants, and something about it makes her uneasy, though she cannot yet say what.

The EU AI Act is the law that answers all three of them. This book is a guided tour through it.

Before anything else, the disclaimer I give at the start of every course, every talk, and now every book: **take everything here with a grain of salt.** I am a data scientist and educator, not your lawyer. This book is careful research turned into plain language, and I stand behind that care, but for any decision with real legal exposure, consult your legal representative. I cannot be held liable for what you build or decide. There, we can relax now.

What this book is

This is the companion textbook to my online course *EU AI Act Compliance Introduction*. It serves the same purpose and the same audience: business people, lawyers, product managers, consultants, and curious professionals who need a working understanding of Europe's AI regulation without wading through 180 recitals of legislative prose.

You do not need the course to read the book, and you do not need the book to follow the course. They cover the same ground in the same order, so you can use either one alone, or use the book as the thing you return to when you need to check what Article 6(3) actually said. The videos are better at building intuition step by step. The book is better at being searchable, quotable, and readable in bed.

What you will not find here is the AI Act reordered with commentary. My ambition is different: I want you to *understand* this regulation the way you understand a story, so

that when a new situation lands on your desk, you can reason about it instead of ctrl-F-ing through the official text in mild panic.

What you will be able to do by the end

By the last chapter, you should be able to look at an AI system, real or proposed, and walk it through the questions that matter: Is it even AI in the Act's sense? Is what it does prohibited outright? Is it high-risk, and if so, what does that actually demand? What must it disclose to the people it touches? And who, in the chain from developer to end user, carries which obligation?

You will also, I hope, come away with something rarer: a calibrated sense of how worried to be. Spoiler from mid-2026: less panicked than the headlines suggested, more attentive than doing nothing.

How the book is organized

We follow the course structure, which follows the logic of the Act itself. Chapters 1 and 2 build the foundations: what AI actually is, and why Europe decided it needed its own law. Chapter 3 is the whole Act at a glance, deadlines included, for readers in a hurry. Chapters 4 and 5 build your vocabulary: the definition of an AI system, and the terms (provider, deployer, profiling, intended purpose) that everything else stands on.

Then the risk ladder, from the top down. Chapter 6 covers the prohibited practices, all eight of them. Chapters 7 and 8 take on high-risk AI, first how a system ends up classified as high-risk, then what obligations follow. Chapter 9 covers the transparency duties, chapter 10 the special regime for general-purpose AI models, and chapter 11 the machinery around it all: regulators, sandboxes, penalties, and the honest state of enforcement. Chapter 12 closes with practice: what to actually do on Monday, plus a bonus tour of the messy, fascinating overlap between AI and copyright.

Every chapter ends with a short quiz. Nobody grades you. But if you can answer five questions after your coffee, the chapter did its job.

You will also meet four kinds of boxes along the way. **The Act says** boxes hold the law's own words, for the moments when exact wording matters. **My take** boxes are me stepping out of the tour-guide role to tell you what I actually think. **Watch out** boxes flag the misreadings I run into most often. And **As of** boxes mark facts with an expiry date: enforcement status, pending guidelines, anything worth re-checking before you rely on it. Everything outside a box is still me talking; the boxes just tell you what kind of claim you are holding.

A note on dates, written in July 2026

The AI Act is a living thing. It entered into force in August 2024, its first obligations began applying in 2025, and in 2026 it was already amended once: the so-called Digital Omnibus moved several deadlines, most notably pushing the high-risk obligations to December 2027.

> AS OF JULY 2026
>
> The Omnibus amendment has been approved by the European Parliament and the Council but is still waiting for its final publication in the Official Journal. This book uses the post-Omnibus dates throughout, because those are the ones that will govern your planning.

There, you have met your first box. It also happens to carry your first real lesson about EU regulation, before chapter 1 even starts: the dates are law, and the law can change. Whenever a deadline truly matters to you, spend two minutes checking the current state before you act on it. The book will teach you where to look.

Pour the coffee. Let's begin.

Next up: what artificial intelligence actually is, told through a chess program that cheated by memorizing.

CHAPTER 1

AI, in Plain Terms

Picture yourself in a bank branch on a Tuesday morning. You want a mortgage. The representative asks the usual questions (your income, your other loans, your employment situation) and types your answers into a computer. Then she says: “Give me twenty seconds and I can tell you whether you qualify.”

Twenty seconds. Nobody walked to a back office. No committee met. Somewhere inside that computer, a piece of software looked at your numbers and made a call about one of the biggest financial decisions of your life.

That is the artificial intelligence this book is about. Not a robot uprising, not some artificial general intelligence from a sci-fi movie, but a quiet algorithm in a bank representative’s computer, deciding things about you while you sip the complimentary coffee. Before we can talk about how Europe regulates such systems, you and I need a shared, honest picture of what they actually are. That is this chapter’s whole job. No law yet, I promise. Just the machinery.

Two ways to build a chess player

Let’s say you and I want to build a computer program that plays chess. We have two fundamentally different ways to go about it.

The first way is to program it explicitly. We sit down and write a long list of rules: prioritize going after the bishops, control the centre, and so on. Or we go even further and encode every position the program might encounter, turning it into one enormous lookup machine. It always checks where it is and reads off what to play. This can work, but you can imagine the cost: we could spend years writing rules and still miss most of what happens on a real board.

The second way is to let the machine learn. We collect data, say 10,000 games played historically by decent players, and we feed them to a learning algorithm. We call this the **training data**. The algorithm studies those games and extracts patterns from them.

Nobody tells it “protect your queen”; it works that out from what winning players actually did. If the data is large enough and good enough, and the learning algorithm is decent, we end up with a machine that plays chess without anyone ever programming the chess into it.

This second approach, building systems that derive their behaviour from data rather than from hand-written instructions, is how most modern AI is built, from the bank’s loan model to the chatbot on your phone. It was already happening with chess in the 1950s. Keep this picture in mind, because almost everything the AI Act worries about flows from it: if the behaviour comes from the data, then whoever shapes the data shapes the behaviour. We will see exactly how sharp that edge is by the end of this chapter.

The word the law actually hangs on: inference

Now, a small but important calibration before we go further. It is tempting to say “AI means self-learning algorithms” and leave it at that. I have said it myself, and as an intuition it serves you well. But it is not what the law hangs its definition on, and the difference matters.

The Act’s definition of an AI system (we will dissect it properly in chapter 4) pivots on one word: **infers**. Here is the core of it.

> THE ACT SAYS · ARTICLE 3(1)
>
> ‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and 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;

In plain terms: what makes something an AI system, in the law’s eyes, is that it works out its outputs from its inputs. It infers, rather than mechanically executing a fixed recipe a human wrote. Notice what the definition says about learning: the system “*may*

exhibit adaptiveness after deployment". May. A system that learned once, was frozen, and never learns again is still squarely an AI system. Self-learning is how most of these systems get built, but it is optional in the definition itself: the term "self-learning" appears exactly once in the whole regulation, in a recital, and only to explain that optional adaptiveness.

> WATCH OUT
>
> A tempting line of reasoning goes: "Our model was trained years ago and doesn't learn anymore, so it's not really AI, so the Act isn't our problem." That reasoning fails on the very first word of the analysis. The test is inference, not ongoing learning.

Why am I being pedantic about this? Because I have watched intelligent people walk straight into exactly that trap. Park the word *inference* somewhere safe; chapter 4 will unpack it fully. For now, back to the machinery.

Training, deployment, utilization

Whatever the system, its life splits into two large phases, and the boundary between them has a name you will meet constantly in this book.

First comes the **training phase**. This is where we introduce the 10,000 historical chess games and the system learns from them. It is a period of change: the model's internal decision-making is still being shaped by the data flowing in.

At some point we stop. We tell the system: that's enough studying, here is a new game, now play it. From that moment on, we are in the **utilization phase**. The system is no longer learning; it is doing its job, over and over, on cases it has never seen.

The boundary between the two, the moment we move the system from the workshop into the world, is what we call **deployment**.

> MY TAKE
>
> Is this two-phase split a simplification? Yes, and I want to be honest about it. There are online learning algorithms that never stop learning: they study the 10,000 games and then keep learning from every new game they play, so the boundary blurs. But those are niche cases. For reading this book, and frankly for reading the regulation, the two-phase picture will serve you well: training, then deployment, then utilization.

One vocabulary note while we are here. In the Act, the organisation that *uses* an AI system in its operations is called a **deployer**: the bank running the loan model is a deployer, distinct from the **provider** who built it. We will make these roles precise later; I mention it now only so the word doesn't surprise you.

Generalization: the student and the exam

Here is the property that separates a good AI system from a bad one, and it is best explained with a student.

Imagine someone preparing for an exam. Option one: study the theory properly, attend the lectures, understand the subject. Option two, the tempting one: get hold of the questions from previous runs of the exam and memorize the answers. If the same questions come up, the memorizer sails through. If the questions change even slightly, they are lost, because they never learned the underlying patterns, only the specific examples.

AI systems face exactly the same fork. During training, a system only ever sees a **sample**: our 10,000 games, the bank's file of past customers. What we hope it learns are patterns that hold for the whole **population**: every game that will ever be played, every customer who will ever walk in. If you have ever studied statistics, you will recognize the sample-versus-population idea; this is it, wearing an AI costume.

When a system learns transferable patterns from its sample, we say it **generalizes**. A chess machine that generalizes plays well in games that resemble nothing in its training

data. A machine that merely memorized its 10,000 games plays brilliantly, right up until the opponent does something it has never seen.

Where it goes wrong, part one: overfitting

Memorizing instead of generalizing has a technical name: **overfitting**. The student cramming old exam questions was overfitting to their training data. Our hypothetical chess machine that leans too hard on the exact historical games is overfitting too.

Let me make it more practical, because with modern AI this stops being a quality problem and becomes a legal one. Suppose we train an image-generating model on paintings by famous artists. If the model overfits, then in the utilization phase it will not create new art inspired by what it studied; it will reproduce the actual artworks it saw during training. Our “creative” product is now recreating real paintings, and we have a copyright problem on our hands. What we wanted was a model that absorbed styles and patterns and can generate something genuinely new. What an overfitted model gives us is an expensive photocopier.

So overfitting is failure mode number one: the system learned its training data too literally. Failure mode number two is subtler, and it is the one the AI Act worries about most.

Where it goes wrong, part two: bias

We like to think of machines as impartial. Humans are the biased ones: we carry cognitive biases, prejudices, bad days. Surely an algorithm, having no psychology, is neutral?

No. And the chess example shows why with almost embarrassing clarity.

Suppose those 10,000 training games came from only four players. Our machine does not learn to play chess in general; it learns to imitate those four people. It is biased toward their style. And if one of the four had a particular recurring weakness, say a habit of mishandling endgames, the machine inherits that weakness faithfully. Nobody programmed the flaw in. It arrived through the data.

Now scale this up. Our society's historical record, the text we have written and digitized over decades, contains plenty of discrimination and prejudice. Train a text-generating model on that record and the model inherits those patterns exactly the way the chess machine inherited a bad endgame. Train a loan model on decades of past lending decisions and it inherits whatever unfairness those decisions contained. The machine is not biased because it is malicious; it is biased because it is a mirror. This is why bias runs as a thread through the entire AI Act, and why we will keep returning to it.

Back to the bank: how data shapes decisions, and how it can be poisoned

Let's close the loop and return to your mortgage. This time, switch sides: you are no longer the customer, you are the bank, and you want to build an automated engine that approves loans. In reality such a model would use many inputs; let's simplify to two, so we can see the mechanics. Each customer has an **income** (low, medium or high) and a **debt ratio** (low, medium or high).

We did not sit down and write approval rules. True to everything above, we trained the model on historical data: past customers, what their income and debt ratio were, and whether they repaid. From that sample the model learned a **decision boundary**: a line through the space of possible customers. On one side of the line sit the profiles that historically repaid: they get approved. On the other side sit the profiles that were rejected or defaulted: they get declined. High income? Approved regardless of debt ratio. Medium income with low debt? Approved. Medium income with medium or high debt? Declined. Low income? Declined.

You, waiting your twenty seconds in the branch, are simply a point being placed on one side of that line or the other.

Now let me show you the attack, because it is wonderfully instructive. Suppose someone wants to cheat our bank, someone whose profile sits just on the wrong side of the line. They cannot move their own point. But what if they could move the *line*?

The line came from the training data. So the attacker recruits thirty or so people from their surroundings and sends them to our branches to register as ordinary customers.

Registration only requires basics, name and address. But each of them, following instructions, volunteers a little extra: “By the way, I’d like to disclose my income, my debt ratio, and mention that I was granted a loan at another bank.” Medium income, medium debt ratio, loan granted and repaid elsewhere. All fabricated. And here is the uncomfortable part: the bank has no particular reason to verify any of it. These people are not applying for anything. It is just voluntary data from friendly new customers, so it goes into the database as given.

Then comes the day the bank retrains its model on the accumulated data. The model, doing exactly what it was built to do, notices a cluster of customers with medium income and medium debt ratio who were good loan recipients elsewhere, and it obligingly moves its decision boundary to let such profiles through. It has learned a lie, diligently. This is called **data poisoning**.

Now the attacker walks in for real. Medium income, medium debt ratio: a profile the honest model would have declined. The poisoned line has shifted, and the loan is approved. No server was hacked, no password stolen. The attacker never touched the model at all; they fed it, and that was enough.

I love this example because it compresses the whole chapter into one scene. The model’s behaviour came from data, so controlling the data meant controlling the decision. The bias arrived silently, through records nobody thought to check. And notice that no individual step looked alarming: thirty registrations, some volunteered details, a routine retraining. This is why, when the AI Act later demands things like data governance and robustness against attacks for certain systems, it is not bureaucratic imagination at work. It is this bank, this line, this red mark on the wrong side of it.

You now have the full toolkit for the rest of the book: systems that infer their outputs rather than follow hand-written rules, usually trained on data and then deployed into a utilization phase; the hope of generalization; and the twin failure modes of overfitting and bias, with data poisoning as a live demonstration of how fragile the whole arrangement can be. Everything the regulation does is a reaction to this machinery.

Check yourself

1. A hospital buys a diagnostic model that was trained once by its vendor and never learns from new patients. In the lifecycle terms of this chapter, the hospital's daily use of it is: A) The training phase B) The utilization phase C) The deployment D) Online learning

2. Under Article 3(1) of the AI Act, the key capability that makes a system an "AI system" is that it: A) Continuously learns and updates itself after deployment B) Infers, from the input it receives, how to generate outputs C) Was trained on large amounts of data D) Exceeds human performance at its task

3. An image-generating model trained on famous paintings starts reproducing those exact artworks for its users. This is a classic case of: A) Bias B) Data poisoning C) Overfitting D) Generalization

4. Our chess machine was trained only on games by four players, one of whom routinely mishandled endgames, and the machine now mishandles endgames the same way. The lesson is that: A) Machines are inherently more objective than humans B) A model inherits the flaws and patterns of its training data C) More training always removes errors D) Chess is too complex for AI systems

5. In the bank attack, the thirty recruits never applied for loans, yet the model started approving profiles it used to decline. Why did the attack work? A) The attacker hacked the model's code during utilization B) The model was never deployed properly C) False voluntary data entered the training set and shifted the learned decision boundary D) The bank's employees were bribed to change the approval rules

Answers: 1. B: the system is past its deployment boundary and is simply doing its job on new cases, which is the utilization phase. 2. B: the definition pivots on inference, while adaptiveness after deployment is only something the system "may exhibit", so self-learning is optional. 3. C: the model memorized its training examples instead of generalizing to new outputs. 4. B: nobody programmed the flaw in; it arrived through the sample the model learned from. 5. C: the attacker poisoned the training data with unverified fabricated records, and the retrained model dutifully moved its decision boundary to accommodate the lie.

Now that you know what these systems are and how quietly they can go wrong, the next question is the obvious one: why did Europe, of all places, decide this needed a law?

CHAPTER 2

Why Europe Regulates AI

A while ago I asked an image-generating AI to draw me a picture connecting two topics: artificial intelligence and the European Union. At first glance it did a stellar job. The EU flag, the circle of stars, something that looks like a glowing neuron, and behind it all a beautiful stretch of European countryside. If I had flashed it on a slide for three seconds, you would have nodded and moved on.

Now take a second, skeptical look. Why are the stars there twice? Why is the count of stars wrong? Why does the river in the background begin nowhere and end nowhere? The fields in the front are ready for harvest; the ones just behind them are bare, as if the seasons gave up halfway across the picture.

Here is the thing that matters: these are mistakes no human illustrator would ever make. A tired human might draw eleven stars instead of twelve. No human draws the flag twice and a river from nothing, blends it all in beautifully, and hands it over with confidence. The mistakes are *novel*, and they are subtle enough that you miss them unless you were already looking.

That one picture, honestly, carries most of the argument of this chapter. Let me unpack it into four reasons why I believe AI needed its own law, and then let's try to wriggle out of that law together, just to see what happens.

Four reasons AI needs its own law

First: AI makes new kinds of mistakes. Existing product rules and quality practices are built around human failure modes: fatigue, sloppiness, fraud. AI systems fail differently. They fail confidently, subtly, and at scale, in ways we have no instinct for. Recall the loan model from the previous chapter that learned its behaviour from data: poison the data, and you poison every decision downstream, with nothing visibly "broken" anywhere. A sensible response is exactly what the AI Act attempts: write the novel failure modes down and ask both the builders and the auditors to go looking for them specifically. You cannot search for a mistake you have not named.

Second: AI is a technology, not an application. The flawed image cost me one sentence of prompting and a few seconds of waiting. In a different setting, that same sentence produces a cheap, convincing lie. But don't stop at any single misuse. If you want to size up AI's future impact, compare it to the internet, a layer that everything else gets built on top of, rather than to any one popular app. We regulate technologies with that kind of reach: aviation, pharmaceuticals, telecoms. It would have been strange if this one stayed exempt.

Third: the automation is aimed at new targets. Automation is as old as the plough, and we have made peace with it. What is genuinely new is that white-collar work, and even creative work, is now being automated. That kind of work has never faced this before, and it has not prepared for it through reskilling or, frankly, through admitting it is happening. I think our reluctance to admit it is itself a risk. A regulation can at least shape how fast and through what means that wave arrives.

Fourth: there is a gap in the rulebook. You might reasonably ask: don't we already have GDPR? Copyright law? Fundamental rights? We do, and they all touch AI. GDPR governs personal data; copyright gives authors tools when their works are swallowed into training sets. But none of them regulates the algorithm itself. Take GDPR, simplified: obtain a valid consent to use a customer's data for certain purposes, stay within those purposes, and nobody questions you much further. Say you collected that consent years ago with simple marketing campaigns in mind. Since then your capabilities grew, and now, within the same consent, you can build profiling machinery that genuinely steers what your customers buy. The data handling is compliant. The *system* has never been looked at. That is the gap.

This is what the legislators wrote into the very first sentence of the Act, Regulation (EU) 2024/1689, in force since 1 August 2024.

> THE ACT SAYS · ARTICLE 1(1)
>
> The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation.

In plain terms: the EU wants AI to be adopted, not banned. Adopted, though, in a way people can trust, with the harms named and managed. Keep both halves of that sentence in mind. The Act is routinely described as hostile to AI; its own statement of purpose says "promote the uptake."

"Fine, then we just won't use AI"

Now let me tell you about a tempting escape route, because I meet it in almost every company I train.

Picture the bank from the previous chapter. It built a self-learning loan model, heard about data poisoning, read a few headlines about the AI Act, and someone in a meeting says: *let's forget all these learning algorithms. Let's build a good old expert rule-based system. Those have worked for ages, and no "AI Act" will apply to us.*

An expert rule-based system is exactly what it sounds like. You gather your most experienced colleagues, extract their rules of thumb (*if the customer is active, and hasn't been contacted in three months, and is likely to want this product, send the email*) and encode those rules into a single engine. These systems were the pride of the 1980s and 1990s, and plenty of them still quietly run banks, insurers, and hospitals today. No learning from data, no neural networks. Just human knowledge, written down.

So our bank builds its loan decisions this way: if credit score is medium and income is low, reject; if both are medium, approve; and so on. Clean, staircase-shaped decision

boundaries drawn by humans. And because the bank wants to be customer-centric, it exposes the system on its website: type in your income and credit score, and see whether you would be approved. A friendly free service.

Also a wide-open door.

An attacker starts *probing* that public interface. They send in artificial combinations of income and credit score, hundreds of them, sweeping in lines across the input space, and they record every answer. Rejected, rejected, approved… approved, rejected, approved. If the website has no limit on requests, the attacker gradually maps out the exact decision boundaries of the system. This is sometimes called model stealing, and notice: it worked perfectly well even though there is no machine learning anywhere in sight.

Then comes phase two. Knowing precisely where the boundaries sit, the attacker submits applications engineered to land just inside "approved": edge cases that should never have been waved through by a machine, cases where a human should have looked. The bank's own transparency became the attack surface.

The same family of tricks has a more famous cousin. Self-driving cars rely on visual recognition models to read traffic signs. Researchers showed years ago that you could construct a special *noise*, a pattern meaningless to human eyes, print a small sticker of it, and stick it on a stop sign overnight. Come morning, the cars see a speed limit sign where you and I see STOP. The model does not share our cognition; it examines pixels, and the pixels were chosen to fool it. (Developers have since hardened systems against this specific stunt, but the lesson stands.)

So, did our bank escape the AI Act by avoiding machine learning?

Here I have to be more careful than the folklore usually is, in both directions. The Act does not regulate "machine learning." Article 3(1) defines an AI system by what it does: a machine-based system that, for explicit or implicit objectives, *infers* from its inputs how to generate outputs such as predictions, content, recommendations or decisions. Whether a rule-based system falls under that definition is genuinely a case-by-case question. Some expert systems reason over encoded knowledge in ways that count as

inference and are in scope; others merely execute rules written entirely by humans, and the Act's own recitals say those are not meant to be covered.

> WATCH OUT
>
> "We built it rule-based, so the AI Act doesn't apply to us" is not a safe conclusion. Neither is the opposite claim that every rule engine is captured. What decides the question is whether the system *infers* how to generate its outputs, case by case. Chapter 4 walks through exactly where that line runs, and it is one of the most practically valuable lines in the whole regulation.

For now, take the deeper point, which I find more important than the classification game: **the risks never asked which technique you used.** The probing attack, the manipulated edge-case approvals, the customers wrongly judged by a machine: all of it happened to a "safe," old-fashioned rule engine.

> MY TAKE
>
> If your instinct is to dodge the regulation by relabelling your technology, you may or may not succeed legally, but the failure modes the regulation worries about will follow you anyway. I would rather build the safeguards first and argue about definitions second, not the other way round.

Who actually does this work?

A regulation is just text until people carry it. So who handles AI Act compliance? I see four personas forming around this law, and if you are reading this book to position yourself, this section is for you.

The scrutinizers, meaning auditors. Every regulation breeds auditors; that is not new. What is new is that these auditors must be at least partially *technical*, not purely legal. Compare a classic GDPR check with an AI Act check. Under GDPR, does the marketing opt-out work? Easy: create a test customer, subscribe, click unsubscribe,

wait, observe. Anyone diligent can run that audit. Under the AI Act, a high-risk system must have, among other things, fail-safe behaviour, so that when a component dies, the system degrades gracefully instead of taking your customers down with it. How do you audit that? You go *into* the system. You examine components, backup paths, the code. We are talking about someone who can read Python and understands what happens inside a model. With several of my corporate clients we are already training exactly these hybrids (part data scientist, part auditor), and I expect "AI Act compliance officer" to become an ordinary job title.

The builders, meaning developers and data scientists. The Act pulls developers onto the compliance team, the way aviation long ago pulled in aircraft designers: the person shaping the component must know the regulation that component has to satisfy. Data scientists are not used to this. Until now, most have worked as if regulation were somebody else's department. Upskilling them so their systems come out compliant *by construction*, rather than patched after the fact, is the other request I keep getting from corporate clients.

The rule makers and enforcers. When I first taught this material, these roles were mostly on paper. They are not anymore. The EU-level AI Office exists, the governance machinery and the penalty regime have applied since 2 August 2025, and every Member State must stand up at least one national *regulatory sandbox*: a supervised environment where you can develop and test an innovative AI system in cooperation with the authority before placing it on the market (Article 57). The deadline for those national sandboxes was moved to 2 August 2027 by the 2026 Omnibus amendment, so expect this persona to keep growing for years.

> AS OF MID-2026
>
> The fine ceilings are dramatic, up to EUR 35 million or 7 percent of worldwide turnover for the worst violations, and they have been live since 2 August 2025. Yet no confirmed AI Act fine exists anywhere as I write this. Enforcement is a machine still being assembled, not a hammer already falling.

The watchdogs. AI is a deeply societal topic, so I expect civil society organizations, ethical-AI groups and NGOs to grow into a real force here: filing complaints, testing systems from the outside, keeping the other three personas honest. This one is the least formed of the four, and I will not pretend to predict its shape.

If you want my practical advice: the first two personas are where individuals can position themselves today. Either become the scrutinizer with a mixed legal-technical skill set, or the builder who ships compliant systems out of the box. Both are scarce, and both will stay scarce for a while.

The optimistic case

We have spent this chapter on attacks, gaps, and obligations, so let me close with a deliberate change of register, because I genuinely view the future with this regulation rather positively.

Step back and ask what the AI Act actually calls for, at the highest level: trust, professionalism, resilience, transparency, accountability. As a data scientist, I'll be honest with you: our field needs this. Our code practices, our testing discipline, our documentation are not yet where mature engineering fields sit. The Act's requirements largely align with software development norms we should have adopted anyway. Teams that go through this are not just becoming compliant; they are becoming better.

Trust is the cornerstone. You rely on your car every morning without inspecting its brakes, because a whole regulatory ecosystem made that trust rational. The Act's insistence on testing, risk management and transparency is laying the same groundwork for AI. And resilience is not charity to the regulator either: ask yourself what one damaging breach (one bank whose loan engine was publicly gamed) costs the entire industry in eroded trust.

There is a strategic prize here too. Imagine, twenty or thirty years from now, a small device in your living room that runs half your life: your work, your calendar, advice on your personal development. Would you want to trust it? Now imagine it carries a little label, **AI from EU**, shorthand for *built under the strictest AI safety and ethics rules in the world*. That label does not exist today, and I am speculating with you. But "made in

EU" earned exactly that meaning in food and machinery, and I see no reason AI cannot follow.

Every great technology goes through a maturing phase. The early internet was experimentation with barely any rules; today's internet runs on sophisticated protocols and security machinery nobody now calls optional. AI is charting the same course, just faster. The AI Act is Europe's bet that the maturing can be steered rather than merely survived. Some hurdles ahead are real, and I will show you the critiques as we go. But the bet itself, I think, is a reasonable one.

Check yourself

1. The AI-generated image of the EU (double flags, impossible river) was used to illustrate which argument for regulating AI? A) AI systems are too expensive for small companies B) AI makes novel mistakes, blended in subtly, that no human author would make C) Image generators infringe copyright D) AI systems always require personal data to work

2. Why doesn't GDPR alone close the regulatory gap around AI? A) GDPR only applies to companies outside the EU B) GDPR was repealed when the AI Act entered into force C) GDPR governs the handling of personal data, but does not examine the algorithm's behaviour; a compliant consent can still feed a harmful profiling system D) GDPR contains no penalties, so it cannot be enforced

3. The bank replaced its self-learning loan model with a hand-built expert rule-based system and exposed it on its website. What happened, and what does it show? A) Nothing, because rule-based systems cannot be attacked B) An attacker probed the public interface, mapped the decision boundaries, and exploited the edge cases; the risks did not depend on machine learning being used C) The system was immediately banned under the AI Act D) The system stopped working because rules cannot make loan decisions

4. Is that rule-based loan system covered by the AI Act? A) Yes, every software system used in banking is automatically an AI system B) No, the Act only covers self-learning systems C) It depends on whether the system *infers* how to generate its

outputs; rules merely executing human-written logic are not meant to be covered, while expert systems that reason may be D) Only if the bank voluntarily registers it

5. Which statement about AI Act compliance roles and enforcement is accurate as of mid-2026? A) Auditors will need mixed legal-technical skills; the penalty regime (up to EUR 35 million or 7 percent of turnover) has applied since August 2025, though no confirmed fine exists yet anywhere B) Only lawyers may audit AI systems, and hundreds of fines have already been issued C) Developers are exempt from compliance duties; only managers are responsible D) National sandboxes have been mandatory since 2024 and every Member State already runs one

Answers. 1 is B: the image's errors (duplicated stars, a river from nowhere) are failure modes specific to AI, which is why the Act asks builders and auditors to look for AI-specific mistakes. 2 is C: GDPR regulates data handling and consent, not the system built on top of them, which is the gap the AI Act fills. 3 is B: the probing (model-stealing) attack mapped the hand-drawn decision boundaries through the public API, proving the dangers survive a change of technique. 4 is C: Article 3(1) defines an AI system by its capability to infer, so rule-based systems are neither automatically captured nor automatically exempt; Chapter 4 draws the line in detail. 5 is A: the fine ceilings have been live since 2 August 2025 with no confirmed fine as of July 2026, and the national-sandbox obligation runs to 2 August 2027, moved there by the 2026 Omnibus amendment.

You now know why the law exists. Next, let's spread the whole Act out on the table and see how its chapters, risk tiers and dates fit together at a glance.

CHAPTER 3

The AI Act at a Glance

Picture a Monday morning. Your CEO forwards you a news article, *"EU AI law: fines up to 35 million euro"*, with a one-line message: "Are we affected?"

This chapter exists so that you can answer before lunch. Not in detail (the rest of the book is the detail), but correctly, in the right order, without the two or three misunderstandings that I see in almost every news article about the AI Act. Think of it as the view from the aeroplane before we land and start walking the streets.

Colleagues of mine like to say the AI Act is a long and complex regulation, and they are right. What I do not want is for you to build your understanding brick by brick across many chapters and only see the shape of the building at the end. You should see the shape now. So here it is: five ideas, one map, one timeline.

First question: is it even AI?

Here is a confession. I spent years as a data scientist, mostly in banking. If you sat me down in front of a simple model (say, one that scores which customers should receive a marketing email) and asked me, "Is this artificial intelligence under the AI Act?", I could not give you an instant yes or no. And I built things like that for a living.

That is not false modesty. The Act's definition of an AI system is genuinely blurry at the edges, and the edges are exactly where most companies live. Everyone agrees that ChatGPT, Claude and Gemini are AI. But what about a logistic regression that helps decide loan applications? An old expert system full of hand-written rules? A chess program that plays by looking up memorised games? Some of these are in, some are out, and the line runs through places your intuition would not put it.

So the first habit I want you to build is this: **do not trust your intuition about what counts as AI.** The next chapter is devoted entirely to the definition, and I promise you will be surprised in both directions: things you thought were "just

statistics" may be covered, and things that feel like AI may fall outside. For now, hold the question open.

Model versus system

The second distinction sounds pedantic and turns out to matter a lot.

An **AI model** is a piece of technology: the trained thing itself, the weights, the engine. On its own it does nothing for a customer. An **AI system** is what you build around the model: the application with its user interface, its integrations, its data feeds, its purpose. The model is the engine; the system is the car.

The AI Act regulates both, but separately. Most of the Act is about systems, actual products being used in the world. A smaller part (Chapter V, as we will see) is about general-purpose models themselves. This matters even if you never sell a product. If you are, say, an open-source contributor publishing a large model on a platform like Hugging Face, you are not shipping a system to anyone, and you can still be within scope of the model rules.

Any AI that touches the EU

The scope question has a mercifully short answer. If you develop AI inside the EU, you are covered. If you import AI into the EU, you are covered. And if you sit entirely outside the EU but your system's output is used in the EU (customers in Berlin, employees in Madrid), you are covered too. Article 2 dresses this in longer sentences, but the summary holds: **any AI that touches the EU is regulated.** Geography of your headquarters is not an escape route, which is precisely the lesson companies learned from the GDPR.

Provider versus deployer

One more distinction, and it is the one your compliance work will actually start from. A **provider** develops an AI system or model and places it on the market or puts it into service. A **deployer** uses an AI system under its own authority; it did not build the

thing, it runs it. (If you read older commentary you will sometimes see “user”; the final Act says deployer, and so will this book.)

Providers carry most of the obligations. Deployers carry fewer, but not zero: being “just” a deployer does not put you outside the Act.

Now the practical twist: your company is probably both. Take a multinational bank, one of our running examples in this book. Its credit-scoring model, built in-house by its own data science team: the bank is the provider. The HR department’s off-the-shelf CV-screening tool, bought from a vendor: the bank is a deployer. A marketing model tweaked from a vendor’s product: it depends, and the answer can shift. The first real step of AI Act compliance in almost every organisation I have advised is exactly this exercise: list your AI use cases, and for each one decide, provider or deployer?

The four risk lanes

The news usually presents the AI Act as a “risk pyramid”. That picture is not wrong, but I want you to see it as four lanes that a use case can fall into, because you will check them in this order.

Lane one: prohibited practices. Chapter II contains a short list of things you simply must not do with AI: manipulative techniques that materially distort behaviour, exploiting vulnerable groups, certain uses of biometrics, and a few more. Note the word *practices*. The Act does not ban use cases or whole projects; it bans practices. Your recruitment tool can be perfectly legal as a project while one feature buried inside it (say, inferring candidates’ emotions from video) crosses a prohibited line. When you audit, you are not asking “is this project banned?” but “does anything inside this project constitute a banned practice?”

One of these prohibitions deserves its exact wording, because the press so often gets it wrong. Social scoring is usually described as a ban on *governments* rating citizens. Here is what the Act actually prohibits:

> THE ACT SAYS · ARTICLE 5(1)(C)
>
> the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to [...] detrimental or unfavourable treatment [...]

In plain terms: nobody may do this. The text names no actor at all; a government, a bank, an insurer and a platform are all equally caught if they score people across contexts in a way that leads to unjustified or unrelated detrimental treatment. An early draft limited this to public authorities; the final Act does not. If your company builds a cross-context "customer trustworthiness score", the ban is your problem too.

Lane two: high-risk AI. If you avoid the prohibitions, the next question is whether your system is high-risk under Chapter III. There are two routes in. The main one is Annex III, a list of sensitive use-case areas: hiring, credit-worthiness, education, essential services, and so on. Our bank's loan-scoring model is the textbook case; creditworthiness evaluation of natural persons sits right there in Annex III. The second route covers AI acting as a safety component of products already regulated by EU law (machinery, medical devices; this is the Annex I route). High-risk is where the Act's real weight lands: risk management, data quality, documentation, logging, human oversight, conformity assessment. In my view this is the core of the Act, and it gets the longest chapters of this book.

> WATCH OUT
>
> For most Annex III systems, the conformity assessment is not some external inspection. Article 43(2) prescribes a procedure "based on internal control": self-assessment by the provider, with no notified body involved. Independent third-party assessment is mainly reserved for certain biometric systems and the Annex I product routes.

Whether self-assessment makes the regime lighter or just lonelier is a fair question; we will return to it.

Lane three: transparency duties. Chapter IV (essentially one article, Article 50) covers systems that are not high-risk but can mislead people about what they are dealing with. Chatbots must not pass as humans; AI-generated content must be marked; deepfakes must be labelled; people must be told when emotion recognition or biometric categorisation is applied to them. Small chapter, wide reach: this lane touches almost everyone who runs a customer-facing chatbot.

Lane four: general-purpose AI models. Chapter V regulates the big foundation models (the engines behind ChatGPT, Claude, Midjourney) at the model level, with an extra tier of duties for models classified as posing systemic risk. If you are not training such models, this lane mostly reaches you indirectly, through the models you build on.

A single use case can sit in more than one lane at once, which is why you check all four.

The map: Chapters I to XIII

The enacted Act, Regulation (EU) 2024/1689, in force since 1 August 2024, is organised into thirteen chapters. (Older drafts used “Titles” and a different order; if a blog post walks you through “Title 8”, it is describing a text that no longer exists.) Here is the map you will actually navigate by:

CHAPTER	WHAT LIVES THERE
I	Scope, definitions, AI literacy
II	Prohibited AI practices
III	High-risk AI systems: classification and obligations (the bulk of the Act)
IV	Transparency obligations (Article 50)
V	General-purpose AI models
VI	Innovation support, regulatory sandboxes

CHAPTER	WHAT LIVES THERE
VII	Governance: the AI Office and national authorities
VIII	EU database for high-risk systems
IX	Post-market monitoring and market surveillance
X	Codes of conduct
XI	Delegation of powers
XII	Penalties
XIII	Final provisions, including Article 113, the timeline

If you memorise one navigation fact, make it this: **GPAI is Chapter V, immediately after transparency.** The four lanes you just learned (prohibitions, high-risk, transparency, GPAI) sit in Chapters II, III, IV and V, in that order. The rest is machinery around them. Chapters I through V are where you will spend most of your reading life; VI through XIII tell you who enforces all this, where high-risk systems get registered, and what non-compliance costs.

The timeline: three waves behind us, three ahead

Every EU regulation phases in, and the AI Act's phase-in has already had a plot twist, so let me give you the honest, current calendar rather than the one from 2024's headlines.

Three waves have already washed over us:

2 February 2025: the prohibited practices became applicable, together with a duty most calendars forgot: Article 4 on AI literacy, which asks providers *and* deployers to take measures supporting the AI literacy of their staff. If your organisation uses AI at all, this one has applied to you for over a year.

2 August 2025: the GPAI model obligations, the governance structure, and the penalties.

> WATCH OUT
>
> The fine ceilings of 35 million euro or 7 percent of worldwide turnover for prohibited practices have been legally live since August 2025, not 2026. Enforcement exposure did not wait for the “main” deadline.

2 August 2026: a month after I write this, the Article 50 transparency duties arrive: chatbot disclosure, synthetic-content marking, deepfake labelling. This date did *not* move.

And here is the plot twist. The original Article 113 said the high-risk obligations would apply from August 2026. In 2026 the EU passed the Digital Omnibus amendment (agreed by Parliament on 16 June and by the Council on 29 June 2026, awaiting Official Journal publication as I write), which pushed the high-risk dates back. These are fixed calendar dates, not conditional on standards being ready:

DATE	WHAT STARTS TO APPLY
2 Feb 2025	Prohibited practices (Article 5) and AI literacy (Article 4): **live**
2 Aug 2025	GPAI model obligations, governance, penalties up to €35m / 7%: **live**
2 Aug 2026	Article 50 transparency duties (not deferred)
2 Dec 2026	New prohibition, added by the Omnibus: AI-generated non-consensual intimate imagery and CSAM
2 Aug 2027	Every Member State must have a national regulatory sandbox operational
2 Dec 2027	High-risk obligations, Annex III route (moved from 2 Aug 2026 by the 2026 Omnibus amendment)
2 Aug 2028	High-risk obligations, Annex I product route (moved from 2 Aug 2027)

For most readers of this book, the date that matters most is **2 December 2027**: that is when the high-risk machinery bites for the Annex III use cases, which is where the vast majority of high-risk AI will be classified.

The honest status report, mid-2026

When I first taught this material, I predicted a GDPR-style frenzy around the high-risk deadline: consultants booked out, panicked hiring, shady overnight experts. I lived through the GDPR version of that in 2018 from inside a bank, and the pattern seemed obvious.

I have to report that, so far, the frenzy never came.

> AS OF MID-2026
>
> There is no confirmed AI Act fine anywhere in the EU, even though the penalty regime has been live since August 2025. Not a single harmonised technical standard has been cited in the Official Journal yet, which also means an ISO/IEC 42001 certificate is a fine management-system credential but is *not* AI Act compliance. And the Digital Omnibus amendment, though agreed by Parliament and Council, still awaits Official Journal publication. All of this is the kind of state of play that changes; re-verify it before you rely on it.

And the Omnibus itself taught everyone a lesson I would not have put in the 2024 syllabus: EU deadlines, which I described to my clients as fixed the moment the Act entered into force, can legally move. One amendment later, “August 2026” became “December 2027”.

Does that mean you can relax?

> MY TAKE
>
> The prohibitions and the penalty ceilings are live today, so lane one is not optional and never was. The transparency duties arrive within weeks of my writing this. And the high-risk work (inventorying your use cases, sorting provider from deployer, checking Annex III) takes many months in any real organisation. The deadline moved once; I would not build a compliance plan on the hope that it moves twice. I would rather count with the worse possibility, at least for now: December 2027 is real, and the calm of 2026 is the quiet part of the curve, not proof that the curve is flat.

That is the view from the plane. Now we land, and we start with the question that everything else depends on: what, exactly, counts as AI?

Check yourself

1. The AI Act's ban on social scoring applies to… A) Public authorities only B) Public authorities and companies acting on their behalf C) Any actor, public or private, placing such a system on the market or using it D) Only companies above a certain turnover threshold

2. A hospital buys a vendor's AI triage tool and uses it on its patients. Under the AI Act the hospital is a… A) Provider B) Deployer C) Distributor D) Not covered, because it did not build the tool

3. Where do the rules on general-purpose AI models sit in the enacted Act? A) Chapter V, right after the transparency chapter B) Chapter VIII, after governance C) Title 8, near the end of the Act D) Annex III

4. After the 2026 Digital Omnibus amendment, when do the high-risk obligations for Annex III systems (hiring, credit scoring and similar use cases) apply? A) 2 August 2026 B) 2 December 2026 C) 2 December 2027 D) Whenever harmonised standards become available

5. Which statement about penalties and enforcement is correct as of mid-2026? A) Fines cannot be imposed until the high-risk rules apply in December 2027 B) The €35m / 7% fine ceilings have applied since August 2025, but no confirmed fine exists anywhere yet C) Several large fines have already been issued for prohibited practices D) Penalties apply only to providers, never to deployers

Answers: 1: C. Article 5(1)(c) names no actor; the government-only limitation existed only in an early draft, so private social scoring is equally prohibited. 2: B. The hospital uses a system it did not develop under its own authority, which is the definition of a deployer (and deployers still carry obligations). 3: A. The enacted Regulation 2024/1689 uses Chapters, and GPAI is Chapter V, immediately after transparency in Chapter IV. 4: C. The Omnibus moved the Annex III date from 2 August 2026 to 2 December 2027 as a fixed calendar date, not one conditional on standards. 5: B. The penalties chapter has applied since 2 August 2025, yet as of July 2026 no confirmed AI Act fine exists anywhere in the EU.

Next, we open the question this chapter told you not to answer by intuition: what actually counts as AI under the Act?

CHAPTER 4

What Counts as AI

Somewhere in your company there is a system nobody thinks of as artificial intelligence. Maybe it is a scoring engine the risk department built fifteen years ago, thirty hand-written rules deciding which invoices get flagged. Maybe it is a spreadsheet that averages last quarter's sales. And somewhere else there is a system everybody proudly calls AI, because the vendor's brochure said so.

The AI Act does not care about brochures. It cares about one definition, Article 3(1), and that definition is the gate to the entire regulation. If a system is not an "AI system" under Article 3(1), you can close the Act and go home, at least for that system. If it is, you continue down the road: is it prohibited, high-risk, minimal risk? Everything starts here.

So this chapter answers a deceptively simple question: what counts as AI? We will build the answer from three ingredients (autonomy, adaptiveness and inference) and then assemble them into the official definition. I promise concrete examples throughout, including a few that will surprise you in both directions: systems you would never call AI that are covered, and systems you would swear are covered that are not.

Autonomy: who is really deciding?

Let's say you run a bank, and you want an engine that decides who gets a mortgage. You have two ways to build it.

In the first version, the AI decides alone. A customer applies; let's call him Robert. The system pulls his transaction history, spending patterns, income, risk rating, and says yes or no. Nobody would argue with calling this autonomous. It independently decides something important about a person's life.

In the second version, the AI only recommends. A human advisor reads the recommendation, checks the underlying data, and makes the final call. Surely this one is not autonomous; the machine cannot actually do anything?

Do not be so quick. When I worked as a data scientist in banking, we had a name for what happens next: the human factor. The model is good. For six months, the advisor checks its recommendations and agrees every single time. Slowly, without any decision being made, the checking stops. The advisor learns to trust the machine implicitly and starts waving its recommendations through. The system was designed to have no autonomy, and it acquired autonomy anyway, through nothing more exotic than human nature. (It cuts the other way too: an advisor with a monthly credit-card quota may override the model for reasons that have nothing to do with Robert's creditworthiness. Putting a human after an AI is always trickier than it looks.)

The Act's answer to this puzzle sits in the phrase "varying levels of autonomy" in the definition, unpacked by Recital 12: some degree of independence of actions from human involvement. The word *varying* does a lot of work. Even a modest degree of independence (the system produces its recommendation without a human co-authoring it) satisfies this element. You will rarely escape the definition by pointing at a human in the loop.

Rarely, but not never. The Commission's guidelines on the AI system definition (C(2025) 924, adopted in February 2025) confirm that autonomy is a genuine, necessary condition: a system that requires full, continuous manual involvement, where the human is effectively doing the work and the machine is just a tool in their hand, falls outside the definition. The bar is low, but it exists.

Adaptiveness: five systems, one question

The second ingredient is adaptiveness: roughly, the ability of a system to change to suit different conditions. To get a feel for it, walk through five systems with me and ask of each: does it adapt?

A calculator. Fully deterministic. Type the same formula, get the same answer, forever. Not adaptive; the easiest verdict we will get all chapter.

A website that remembers your dark theme in a cookie. You pick dark mode; next visit, it is dark. The system has, in a pedantic sense, adapted to you. But let's not

stretch our imagination; nobody drafted this regulation with theme cookies in mind. Probably not adaptive.

An expert system with hand-written rules. Remember the chess machine from earlier in the book, the version where we explicitly programmed thousands of board positions and the right move for each, and playing meant looking up the current position in that library. These rule sets are called expert rules because, hopefully, an expert wrote them; the approach had its heyday in the 1980s, in medical diagnostics among other fields. Is such a system adaptive? At any moment in time, no: it is frozen exactly as its makers left it. Humans may revise the rules every week, but the system itself does not change through use.

A machine learning system. This is the other chess machine, the one we did not program move by move, but trained on ten thousand historical games, letting it work out for itself how to play. During training, this system is adapting constantly: it tries an opening, discovers from the data that better ones exist, and changes its strategy. Once deployed, a typical ML system stops learning, but it was built by adaptation.

A machine learning system with online, continuous learning. The same chess machine, except it keeps learning after deployment, updating itself after every game it plays. This is a real technique (weather forecasting uses it, because yesterday's data matters), but a niche one, and for good reason: it is fragile. If an attacker knows your chess machine learns from every opponent, they can deliberately play terrible games and poison it. You may remember the poisoned-training-data story from earlier in the book; with online learning, the poisoning never has to stop.

Scale of adaptiveness, from the calculator's zero to the online learner's constant self-modification. Now the punchline: for the Act's definition, almost none of this matters. The definition says an AI system "may exhibit adaptiveness after deployment". *May*. I will be honest with you: when I first read that, I spent weeks on it, wrote articles about it, even contacted the European Commission, because read strictly with the surrounding "and"s it seemed no fixed system could qualify. The settled answer, confirmed by the February 2025 guidelines, is that adaptiveness is optional. If your system keeps learning after deployment, fine; if it never changes again, it can still be an

AI system. ChatGPT, which is essentially frozen after training, is not saved by its non-adaptiveness, and neither is your model.

> WATCH OUT
>
> You will sometimes hear AI defined as "self-learning algorithms". Be careful with that shortcut: the Act does not define AI by self-learning. The phrase appears exactly once in the whole regulation, in Recital 12, and only to explain this optional adaptiveness element. Anchor on self-learning and you will wrongly de-scope every deployed model that no longer learns, which is most of them.

Inference: the heart of the definition

If autonomy is a low bar and adaptiveness is optional, what actually separates AI from ordinary software? Recital 12 tells you plainly: "A key characteristic of AI systems is their capability to infer." Inference is where the definition earns its keep, so let's slow down.

Inference, in the Act's sense, is the system's capability to derive how to generate its outputs: to work out models, patterns or rules from data or encoded knowledge, rather than merely executing a recipe a human wrote out in full. Our trained chess machine infers: nobody told it how to respond to a board position it has never seen, yet it produces a sensible move, because during training it derived its own model of the game. The mortgage model infers: it learned from historical customers who did or did not repay, and applies those learned patterns to Robert, whom it has never met. ChatGPT infers: it has almost certainly never seen your exact prompt, yet it generates an answer from patterns learned across its training data. A Netflix or Spotify recommender infers: you are unique, but people similar to you have been on the platform for years, and the system has learned from them what you might like.

Now, the crucial boundary, and this is where careless readings of the Act go wrong. Recital 12 does not stop at "capability to infer". It adds a limiting sentence that deserves its own frame on the wall.

> THE ACT SAYS · RECITAL 12
>
> The capacity of an AI system to infer transcends basic data processing by enabling learning, reasoning or modelling.

In plain terms: not every computation that produces an output from an input is inference. If a statistician surveys 500 people about their age and computes an average to estimate the city-wide mean, that is a perfectly respectable piece of inferential statistics, and it is *not* AI under the Act. The February 2025 guidelines say so almost verbatim: software that calculates a population average from a survey is their own example of basic data processing, outside the definition. The same recital sentence rules out your spreadsheet, your database queries, your monthly KPI dashboard. Statistical inference and Article 3(1) inference share a word, not a scope. The Act is after systems that *learn, reason or model*, systems that derive their own way of producing outputs, not systems that apply a fixed formula, however statistical the formula sounds.

There is a second route into inference besides machine learning, and Recital 12 names it: “logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved”. Hold that sentence: it decides the fate of our expert systems in a moment, and it is more selective than it first appears.

The definition, assembled

You now have all three ingredients, so here is the most consequential sentence in the regulation, Article 3(1).

> THE ACT SAYS · ARTICLE 3(1)
>
> ‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and 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.

In plain terms: a machine-based system (humans making decisions are not covered; your receptionist recommending tourist attractions is not a regulated system), with at least some independence from human involvement (a low bar, but a real one), which may or may not keep learning after deployment (genuinely optional), and which infers how to generate its outputs (the heart of the matter, and remember: inference must transcend basic data processing).

The definition is broad, and deliberately so. The technical people I talk to say it is future-proofing: the Commission did not want to amend the Act every time a new architecture appears. My legal colleagues add a second motive: the breadth forces companies to actually inventory their systems and think about each one, instead of assuming none of it applies.

Broad, though, is not boundless. Exactly where the boundary runs, and which systems the Commission has since said fall outside it, is the business of the next section.

So which systems are covered?

Let's give a verdict on our five systems, honestly this time, including the messy middle.

The **calculator**: out. No inference, no debate. The **dark-theme cookie**: out. It remembers a preference; it learns, reasons and models nothing.

The **machine learning system** (our trained chess machine, the mortgage model): in. It derived its own rules from data; this is exactly the system the definition was written for. The **online continuous learner**: emphatically in. It ticks every box, including the optional one.

And the **expert system with hand-written rules**: the case I most need you to get right, because European industries are full of these. Banking, insurance, manufacturing all run decades-old rule engines, and their owners need a correct answer, not a slogan. The answer is: *it depends on whether the system reasons*.

Recital 12 draws the line itself. On one side, it explicitly excludes "systems that are based on the rules defined solely by natural persons to automatically execute

operations”. A fixed rule set that a human wrote and the machine merely executes (if the amount exceeds 10,000 and the country is on the list, flag the invoice) is not an AI system, no matter how many rules there are. Our lookup-table chess machine falls here too: every move was pre-programmed by humans; the machine executes, it does not infer. The guidelines even use a rules-plus-heuristics chess program as their own example of a classical heuristic outside the definition.

On the other side, Recital 12 includes logic- and knowledge-based approaches that *infer from encoded knowledge*. A genuine expert system in the 1980s medical-diagnostics tradition, meaning encoded expert knowledge plus a reasoning engine that chains rules together, deduces conclusions and derives answers to cases nobody explicitly programmed, does infer, and is an AI system. The guidelines give medical-diagnosis expert systems as exactly this kind of in-scope case.

So if you are inventorying a legacy rule estate, the question for each system is: does it merely execute human-written rules, or does it reason over encoded knowledge to reach conclusions its authors never spelled out? Execution is out; reasoning is in.

The expert-rule line is one instance of a wider correction, and here I have updated my own advice over time. When the Act was new, my honest recommendation was to assume everything with inference was in scope and not to hunt for exits, because no official narrowing existed. Since February 2025, one does. The Commission’s definition guidelines name whole categories of systems that sit outside the definition even though they compute, and in a narrow sense even infer: systems for improving mathematical optimisation (including long-established methods like linear and logistic regression used in that settled way), basic data processing, classical heuristics, and simple prediction systems whose performance a basic statistical rule could match (predicting tomorrow’s demand with a historical average, say). You are no longer obliged to classify every formula in the building as AI.

AS OF MID-2026

The February 2025 definition guidelines (C(2025) 924) remain the operative Commission reading of Article 3(1). They interpret the Act rather than amend it, and no court has yet ruled on the definition's edges, so a judge could see a borderline case differently. Before you rely on one of the out-of-scope categories for an important system, check whether the guidelines have been updated.

MY TAKE

Some systems will sit uncomfortably on the execution-versus-reasoning line, and this is a case-by-case assessment. Where a system genuinely straddles it, I would rather count with the worse possibility, at least for now, and document why. Treating a borderline system as an AI system costs you little; wrongly de-scoping one can cost you a lot.

With the boundary in hand, run through a typical company and you will still find plenty in scope: the marketing engine predicting which product to offer which customer and when; the retail credit and insurance-premium models; the website that rearranges itself based on learned visitor behaviour; fraud detection on card transactions; the solar-output forecaster; the car system reading stop signs; Face ID on your phone; the ops model predicting which hard disk fails next; and of course every chatbot from ChatGPT to Claude. All machine-based, all inferring.

One deep breath before you panic-email your inventory to the board: being an AI system does not mean being heavily regulated. It means you keep reading the Act. Most of the systems in that list will land in minimal-risk territory, with no obligations worth losing sleep over. Some will pick up transparency duties, a few will be high-risk, and a handful of practices are banned outright. That sorting is where the rest of this book goes. The definition is just the gate, but you now know, with more precision than most compliance departments, who has to walk through it.

Check yourself

1. Which element of the Article 3(1) definition is genuinely optional? A) Being machine-based B) Autonomy C) Adaptiveness after deployment D) The capability to infer

2. Your team builds software that surveys 500 employees and computes their average commute time to estimate the company-wide mean. Under the Act and the Commission's February 2025 definition guidelines, this is: A) An AI system, because sample-to-population inference is the Act's core concept B) Not an AI system, because it is basic data processing, which the capacity to infer must transcend C) An AI system, but exempt because it is used internally D) Prohibited, because it processes employee data

3. Which of these rule-based systems is most likely an AI system under the Act? A) A chess program that looks up pre-programmed moves for each board position B) An invoice filter executing thirty fixed human-written rules C) A medical-diagnostics expert system whose reasoning engine chains encoded expert knowledge to reach conclusions its authors never explicitly programmed D) A thermostat that switches heating on below 19 °C

4. A bank's mortgage model only issues recommendations; a human advisor always makes the final decision. Regarding the autonomy element of the definition, the safest reading is: A) The system fails the autonomy element, so it is not an AI system B) "Varying levels of autonomy" means even limited independence counts, so the human in the loop does not take the system out of the definition C) Autonomy only matters for high-risk systems D) The system is autonomous only if the advisor stops checking its outputs

5. What did the Commission's guidelines on the AI system definition (C(2025) 924, February 2025) do? A) Legally amended Article 3(1) to narrow the definition B) Confirmed that all rule-based systems are AI systems C) Described categories of systems outside the definition (such as basic data processing, classical heuristics and long-established statistical methods) as the Commission's non-binding reading D) Created a certification scheme for AI system classification

Answers: 1: C, because the definition says a system "may exhibit adaptiveness after deployment", so a system that never changes after deployment still qualifies. 2: B, because Recital 12 requires inference to transcend basic data processing, and the guidelines list survey-average software as their own out-of-scope example. 3: C, because systems that reason over encoded knowledge infer under Recital 12, while fixed rule-execution, lookup tables and simple heuristics fall outside. 4: B, because the autonomy bar is low and only systems under full continuous manual control fall out on this element, so plan as if a recommender system qualifies. 5: C, because the guidelines interpret rather than amend the Act, naming out-of-scope categories, and being the Commission's view they guide but do not bind courts.

You now know which systems walk through the Act's front gate; next we meet the cast of characters the Act assigns to them: providers, deployers, intended purpose and the rest of the Act's vocabulary.

CHAPTER 5

The Act's Vocabulary

Good news first: the hardest part is behind you. Last chapter we wrestled with the definition of an AI system itself: inference, autonomy, the chess machine that turned out not to be AI after all. From here on, the concepts get shorter and more concrete.

Before we can talk about what the Act prohibits and treats as high-risk, we need a shared vocabulary. The Act front-loads its definitions into Article 3, dozens of them, and a handful will follow us through the rest of this book like recurring characters. This chapter picks out the ones you will actually need. Learn them now, and the prohibited-practices and high-risk chapters will read twice as fast.

Publicly accessible space

Imagine two cameras. One hangs in your company's back office, where only badge-holding employees ever set foot. The other hangs in the customer hall of a bank branch, where anyone can walk in off the street. Same camera, same software, but for the AI Act they live in two legally different worlds, because one of them watches a *publicly accessible space*.

Article 3(44) defines it as any physical place accessible to an undetermined number of people. Two things do the work. First, "undetermined number": you don't know in advance whether ten people or a thousand will show up. Second, ownership is irrelevant: a privately owned shopping mall is just as publicly accessible as a municipal park. Recital 19 gives a generous list: shops, restaurants, cafés, banks, gyms, stadiums, stations, airports, cinemas, museums, parks, playgrounds.

Recital 19 also draws the exclusions. Offices, factory floors and workplaces meant only for employees and service providers are *not* publicly accessible. A merely unlocked door does not make a space public; if there are signs restricting access, the signs win. Interestingly, online spaces are not covered at all. Your website or forum is not a "publicly accessible space", because the term means physical places only. The recital closes with a very European hedge: accessibility is determined case by case.

Now, why does this term matter? Its real weight sits in one specific prohibition: Article 5(1)(h) bans *real-time remote biometric identification* in publicly accessible spaces, but only *for the purposes of law enforcement*, and even then with narrow exceptions we will meet in the next chapter.

> WATCH OUT
>
> It is easy to over-read this term. A private bank running camera analytics on customer behaviour in its own branch is not committing a prohibited practice under this ban. It may well be building a high-risk system, and GDPR certainly has opinions, but the prohibition is aimed at police-style identification, not at every camera in every lobby.

Keep the term in your pocket. You will need it precisely once.

Emotion recognition system

Picture a receptionist whose employer has installed a camera that scores her mood all day, flagging when she looks irritated and rewarding when she smiles. Hold that image; it returns in the next chapter with consequences. For now, the definition.

Article 3(39): an emotion recognition system is an AI system that identifies or infers the *emotions or intentions* of natural persons on the basis of their *biometric data*. Biometric data is, roughly, data about the physical you: your face in an image, your voice, your gait, your eyes. Run an AI over that data to conclude "this person is happy, angry, ashamed" and you have built an emotion recognition system. Recital 18 lists the emotions it has in mind, from happiness and sadness through disgust and embarrassment to contempt and satisfaction.

When the Apple Vision Pro was announced, there was lively discussion about exactly this. Sensors millimetres from your eye are a superb source of biometric data, and your eye betrays what you like even when your face stays still. A system inferring which advertisements your pupils respond to sits squarely inside this definition.

Just as important is what the definition *excludes*: you can build genuinely useful systems right up to the boundary. Recital 18 carves out two things. First, physical states such as pain or fatigue. Detecting that a driver or pilot is drowsy is not reading an emotion; it is reading a physical state, and the automotive industry has shipped such systems for years. Second, the mere detection of readily apparent expressions, gestures or movements, *unless* used to infer emotions. My favourite example: a camera in a bank branch with a single job, spotting when people raise their hands in the air, a decent early signal of a robbery in progress. It detects a gesture and stops there. No emotion inferred, no emotion recognition system. A smile, a frown, a raised voice, whispering: detecting these as expressions is fine. The line is crossed the moment you conclude what someone *feels*.

Why does the boundary matter? Because inferring emotions in certain settings (the workplace among them) is not merely high-risk; it is banned. Our receptionist will be back.

Provider and deployer

Here is the distinction the whole Act hangs on: two roles, two rulebooks, and the provider's is by far the heavier.

A *provider* (Article 3(3)) is anyone who develops an AI system or general-purpose AI model (or has one developed) and places it on the market or puts it into service under their own name or trademark, whether for payment or free of charge. It does not matter whether you wrote the code or paid someone else to, nor whether you charge for it. Develop-and-release under your name, and you are a provider.

A *deployer* (Article 3(4)) is anyone using an AI system under their authority, except purely personal, non-professional use. The hobbyist tinkering at home falls outside the Act entirely; an organisation using an AI system professionally is a deployer. (Older commentary says "user". That was the draft-era word; the enacted Regulation says deployer.)

One clarification the definitions force on us: *system* versus *model*. Think back to our chess machine. The finished product, interface, website plumbing and all, is the AI

system; buried inside sits the component that actually plays chess, the model. A model is the smart core; a system is the product wrapped around it. The distinction matters most for general-purpose AI, where one company builds the model and hundreds build systems on top.

Two practical realities from my consulting work. First, far more companies are providers than expect to be. It is tempting to think "providers means OpenAI, Google, Anthropic; we just use their stuff." But a bank that trains its own loan-approval model on its own customer data, even in an off-the-shelf framework, has developed an AI system. It is a provider. Second, one company is usually *both*: large organisations map their use cases and discover that for five they are the provider and for three the deployer. The role attaches to the use case, not the company. (Importers, distributors and authorised representatives exist too; provider versus deployer is the distinction to internalise.)

When a deployer becomes a provider

Now the trap. You bought a high-risk AI system from a vendor; you are a deployer, with the lighter rulebook. Can you stay that way? Article 25(1) says: only if you keep your hands off it. A deployer, distributor or importer *becomes* the provider, inheriting the full provider obligations, in three situations:

- you put your own name or trademark on a high-risk system already on the market;
- you make a substantial modification to it, and it remains high-risk; or
- you change the intended purpose of a system that was not high-risk, in a way that makes it high-risk.

Rebrand it, rebuild it, or repurpose it into high-risk territory, and congratulations: it is legally yours now; the original provider steps out of that role.

What about fine-tuning? For years this was the grey zone: take a pre-trained model, fine-tune it on your own data, and ask whether you are now the provider of a new model. Since late 2025 there is an official answer for general-purpose AI models. The Commission's GPAI guidelines (C(2025) 7719) say that modifying an existing GPAI

model makes you a GPAI provider only if the modification is *substantial*, with an indicative rule of thumb: the compute you spend exceeds roughly one third of the compute used to train the original model. That is an enormous amount; the Commission itself expects few fine-tuners to ever cross the line. So the typical company fine-tuning an open model on internal documents does not become a GPAI model provider, though it will very likely still be the provider of the AI *system* it builds around the model. Those are separate questions, and Article 25 answers the system-side one.

Intended purpose and reasonably foreseeable misuse

Every AI system under the Act comes with a declared job description. Article 3(12) calls it the *intended purpose*: the use the provider intends, including the specific context and conditions of use, as stated in the instructions for use, the promotional and sales materials, and the technical documentation. Note that last part: your marketing counts. If your sales deck promises things your instructions disclaim, you have widened your own intended purpose.

Concrete examples: “designed to assist doctors in diagnosing diseases using patient data” means doctors, not patients, are the intended operators. “Intended to handle customer queries through chatbots” means chatbots, not email triage. Each sentence draws a boundary.

Now suppose someone uses your system outside its intended purpose and harm follows. Your instinct: not my problem, the deployer ignored the instructions. The Act anticipates that move.

> THE ACT SAYS · ARTICLE 3(13)
>
> ‘reasonably foreseeable misuse’ means the use of an AI system in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behaviour or interaction with other systems, including other AI systems;

In plain terms: if you could reasonably have seen the misuse coming, you were expected to plan for it. The doctor-assistance tool *will* end up in the hands of untrained people self-diagnosing. Everyone can foresee that, so the provider should take preventive steps and be able to show them. The customer-service chatbot *will* be poked into generating inappropriate content, which is exactly why serious providers run moderation layers that screen queries before the model answers. “But they misused it” only works as a defence if the misuse was genuinely unforeseeable. And given human creativity, I would keep your list of foreseeable misuses long rather than short.

Special categories of personal data

Some data is simply more dangerous to compute with. Article 3(37) defines *special categories of personal data* by pointing at GDPR (Article 9(1)) and the Law Enforcement Directive: racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data used to uniquely identify a person, health data, and data concerning sex life or sexual orientation.

One nuance: biometric data appears here only in a specific flavour, *for uniquely identifying* a natural person. A fingerprint identifies you uniquely; that is special-category data. Our hands-in-the-air robbery camera processes images of bodies but identifies nobody. The purpose, not the pixels, decides.

The rule of thumb for this book: whenever an AI use case processes special categories of personal data, your risk level rises. Sometimes it rises into high-risk; combined with certain practices, even into prohibited territory. When you audit a use case, “does it touch this list?” should be one of your first three questions.

Profiling

Here is a term that will matter enormously two chapters from now; I want to plant it early.

Article 3(52) defines *profiling* by pointing at GDPR Article 4(4): any automated processing of personal data that evaluates personal aspects of a natural person, in particular analysing or predicting things like their performance at work, economic situation, health, preferences, reliability, behaviour, location or movements.

That reads abstractly, so let me make it concrete with our running bank. Suppose marketing wants to predict which customers will soon take a mortgage. Age between 30 and 40, decent income, living in a neighbourhood people tend to move out of: the model scores your *economic situation* and life stage. That is profiling, textbook case. Or your employer runs software that does not just count keystrokes but models how *productive* you are: profiling of performance at work. Or a branch camera scores visitors' movements as "suspicious": profiling of behaviour.

Notice how ordinary these examples are. Banks, insurers and HR departments have done this for decades; the Act simply gives the practice one umbrella term and attaches consequences. The consequence to remember, the seed I am planting, is this: when we reach high-risk classification, you will meet escape routes that let borderline systems argue their way out of the high-risk category. Profiling slams those routes shut. An Annex III system that profiles natural persons is *always* high-risk, full stop. Because profiling is such a broad term, this single rule sweeps in many more use cases than intuition suggests. Remember the mortgage model.

The open-source exceptions

Finally, open source. The EU faced a dilemma here: much of the AI ecosystem runs on freely shared models and tools, often maintained by people nobody pays, and regulating them like commercial vendors risks strangling that openness. So the Act carves out relief at two levels, and it pays to keep them apart.

Model level. Providers of general-purpose AI models carry documentation duties under Article 53(1): technical documentation for the AI Office and national authorities, and information packages for downstream system-builders. Article 53(2) lifts exactly those duties for open models.

THE ACT SAYS · ARTICLE 53(2)

The obligations set out in paragraph 1, points (a) and (b), shall not apply to providers of AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available. This exception shall not apply to general-purpose AI models with systemic risks.

In plain terms: the open release *is* the price of admission. Publish under a genuinely free licence (one allowing access, use, modification and distribution), with the weights, architecture information and usage information public, and the formal documentation duties fall away. Fair enough, since everything is already in the open.

Two duties survive even for open-source providers, and this catches people out. Article 53(1)(c) and (d), the copyright policy and the *publicly available summary of training content*, apply to all GPAI providers, open or closed. This is a summary (there is a mandatory AI Office template), not a demand to hand over the training data itself. An open-source provider skips the documentation paperwork but still publishes a training-content summary, same as everyone.

And the exception has its own exception, in that last quoted sentence: models with *systemic risk* get no relief. Systemic risk attaches to the most capable frontier models. The Act presumes it above 10^{25} floating-point operations of training compute, and providers crossing that line must notify the AI Office within two weeks. The principle is what matters here: openness does not buy the biggest models out of the heaviest duties.

AS OF MID-2026

There is no public register of GPAI models designated as carrying systemic risk, so I will resist the temptation to name names, and I would treat any list you see online with suspicion. Check the AI Office's publications for the current state before you rely on one.

System level. Separately, and easy to miss, Article 2(12) says the Regulation *does not apply at all* to AI systems released under free and open-source licences, unless they are placed on the market or put into service as high-risk systems, or fall under Article 5 or Article 50. So open-source relief exists for systems too, not only models. But it lasts only until the system wanders into high-risk, prohibited, or transparency-relevant territory. Release an open-source CV-screening tool as a product and it is high-risk like any other; the licence is no shield.

> MY TAKE
>
> A specific release strategy is a question for your legal counsel, so take this section with the usual grain of salt. But the shape is friendly: the Act genuinely tries not to kill open source, while refusing to let "open" become a loophole for the riskiest uses.

Check yourself

1. A privately owned cinema installs an AI system in its lobby. For the AI Act, the lobby is: A) Not a publicly accessible space, because the cinema is privately owned B) A publicly accessible space, because an undetermined number of people can enter C) Not a publicly accessible space, because tickets are required D) Only a publicly accessible space during opening hours

2. Which of these is an emotion recognition system under Article 3(39)? A) A camera detecting that a driver is fatigued B) A camera detecting raised hands in a bank branch to flag a possible robbery C) A system analysing eye-tracking data to infer which advertisements a person likes D) A microphone detecting that a caller's voice is raised, without drawing conclusions from it

3. A logistics company buys a general-purpose AI system from a vendor, rebrands it under its own trademark, and sells it on as a high-risk fleet-management tool. Under Article 25(1), the company is now: A) Still a deployer, because it did not train the model B) A distributor with no additional duties

C) The provider of the system, with full provider obligations D) Exempt, because the original vendor keeps all obligations forever

4. An open-source GPAI model (no systemic risk) is released under a free licence with public weights, architecture and usage information. Which duty still applies to its provider? A) None, because open-source models are fully outside the Act B) The Article 53(1)(a) technical documentation for the AI Office C) The Article 53(1)(b) information package for downstream providers D) The Article 53(1)(d) publicly available summary of training content

5. A bank's model predicts which customers are likely to take a mortgage, based on age, income and neighbourhood. Under the Act's vocabulary this is: A) Profiling, because it evaluates a natural person's economic situation by automated processing of personal data B) Not profiling, because marketing is a legitimate interest C) Emotion recognition, because it infers customer intentions D) Only profiling if the customers explicitly consented

Answers: 1: B. Ownership is irrelevant; what counts is access by an undetermined number of people, and conditions like buying a ticket do not change that. 2: C. Fatigue is a physical state and a raised hand or voice is mere detection of a gesture or expression, but inferring likes and intentions from eye data is exactly what the definition covers. 3: C. Putting your name or trademark on a high-risk system already on the market makes you the provider under Article 25(1)(a). 4: D. Article 53(2) exempts open-source models without systemic risk from points (a) and (b), but the training-content summary and copyright policy apply to all GPAI providers. 5: A. Automated evaluation of personal aspects such as economic situation is profiling regardless of purpose.

Now that we share the Act's vocabulary, we can open its darkest drawer: the practices Europe decided no one may do at all.

CHAPTER 6

Prohibited Practices

Take a look at your phone. Is there a game on it you can't quite stop playing? One where, just as you decide to quit, the game hands you a bonus, a lucky streak, a "one more level" nudge that lands at exactly the right moment? Behind that beautiful frontend there may be a system watching how you play and individualising the experience, deciding that *this* player, right now, needs *this* boost to keep spending. When I examined the market, I can tell you: such systems exist.

That, roughly, is where the EU decided to draw a red line. Most of the AI Act is about managing risk: documentation, oversight, transparency. Chapter II is different. It lists the practices that are simply banned in the European Union, no paperwork accepted. And this is not a future problem: the prohibitions have applied since 2 February 2025, they carry the Act's highest fine tier, up to 35 million euros or 7 percent of worldwide turnover, and (we will come back to this) as of July 2026 nobody, anywhere, has actually been fined yet.

Article 5(1) contains eight prohibitions, with a ninth on the way. Eight legal paragraphs make for a dry read, so I group them into four themes: manipulation and deception, exploiting vulnerabilities, discriminatory scoring, and biometric practices. Let's walk through them the way I would with a client: what is banned, what only *looks* banned, and where the edges genuinely blur.

Manipulation and deception

In classic detective stories, the villain is a con artist with charm and carefully woven lies. Today's version doesn't need charm. It needs your behavioural data and a metric to optimise. The principle behind this first theme: AI must not be designed in a way that tricks or secretly influences people into decisions harmful to themselves or others.

Article 5(1)(a) targets AI that uses subliminal techniques beyond a person's consciousness, or manipulative or deceptive techniques, to distort someone's behaviour

in a way that causes, or is reasonably likely to cause, significant harm. And "significant harm" is broad: economic loss counts, psychological distress counts.

One phrase in the Act deserves your full attention:

> THE ACT SAYS · ARTICLE 5(1)(A)
>
> "…with the objective, or the effect of materially distorting the behaviour of a person…"

In plain terms: intent is not required. You might read the word "purposefully" in the Act and think the ban only catches developers who *know* they are manipulating people. The Commission's guidelines on prohibited practices (February 2025) close that door explicitly: "purposefully" is read objectively, based on what the technique is designed to do or objectively aims at. Nobody needs to intend the harm; the guidelines even say the ban covers AI systems that manipulate people *without any human intending it*, for instance manipulation the system learned by itself from training data because it happened to boost engagement. If your recommender discovered on its own that anger keeps people scrolling, "we never meant that" is not a defence.

So my advice to corporate clients is simple: go through your deployed use cases and ask two questions. Does anything here individualise the experience in a way that nudges people toward decisions? And if a customer follows the nudge, could it cause significant harm? Think of a bank whose campaigning system learns that a mortgage offer worded *just so*, sent at *just* the right moment, maximises the chance that a financially inexperienced customer bites. That is not marketing anymore. That is the algorithm behind the marketing, and the algorithm is exactly what this article is aimed at.

Exploiting vulnerabilities

The second theme is the first one's sharper sibling. Under Article 5(1)(b), AI must not exploit the vulnerabilities of a person or group due to their age, disability, or a specific social or economic situation, again with the objective or effect of distorting behaviour toward significant harm.

Children are the obvious case. A child playing an online game is still developing impulse control; an AI that learns a child's psychology from play patterns and then relentlessly pushes in-app purchases is precisely what this article describes. But notice the third category: a specific social or economic situation. Picture someone after a medical emergency or a job loss, desperately searching for money, meeting a loan chatbot that doesn't empathise but calculates, targeting distressed people because they are statistically more likely to accept exorbitant rates. The Act's view is that this is not a customer interaction between equals. It is a power imbalance, weaponised.

Here is the insidious part: the developers often never designed for this. A system optimising conversion will find the vulnerable segment on its own, because the vulnerable convert. That is why "we didn't intend it" fails here too: the *effect* is enough.

Discriminatory scoring

The third theme is about AI that reduces a person to a computed score and lets the score dictate their opportunities. The Act attacks this from two angles: scoring your social life, and scoring your criminal future.

Social scoring

If you have seen the Black Mirror episode *Nosedive*, you already understand this prohibition. A society where everyone carries an invisible score, assembled from social behaviour, and the score dictates whether you get the flat, the job, the seat on the plane.

Article 5(1)(c) bans AI that evaluates or classifies people over time based on their social behaviour or known, inferred or predicted personal characteristics, where the resulting score leads to either of two things: detrimental treatment in contexts *unrelated* to where the data was collected, or treatment that is *unjustified or disproportionate* to the behaviour itself. Your vacation photos costing you a job interview is the first branch. One harsh online comment from years ago still inflating your insurance premium is the second.

Two things people routinely get wrong about this one. First, the ban applies to **any actor**, not just governments. The Chinese-style state system is the famous image, but a

private insurer building a “responsibility score” from your takeout orders and music taste is squarely in scope too. Second, it does *not* ban scoring as such. My typical clients are banks and insurers, and they have calculated credit-risk scores for decades. A creditworthiness model using financial data, in a financial context, for a financial decision is not social scoring; it is high-risk AI (wait for the next chapter). The trouble starts when the inputs drift into social behaviour and the outputs leak into unrelated contexts. If your risk model quietly punishes an impulse purchase from years ago, or clusters customers on patterns that turn out to be proxies for who they associate with, you are drifting toward the line.

Predictive policing

The other side of this theme scores something even more sensitive: whether you will commit a crime. You know the movie *Minority Report*, arresting people for crimes they haven’t committed yet. Article 5(1)(d) prohibits AI that assesses or predicts the risk of a person committing a criminal offence, but read the qualifier carefully: **based solely on profiling or on assessing their personality traits and characteristics**.

That word “solely” is load-bearing, and so is the carve-out that follows it: the ban does not apply to AI that supports a human assessment which is *already based on objective and verifiable facts directly linked to a criminal activity*. So an AI flagging you as a future criminal because of your postcode, personality profile and acquaintances: banned. An AI helping investigators organise evidence in an existing case built on actual facts: not banned.

> WATCH OUT
>
> “Predictive policing is banned outright” is a misreading I hear often, and I would not want you repeating it to colleagues. The prohibition is narrower than the movie poster suggests: it catches predictions based *solely* on profiling or personality traits, while AI supporting fact-based human assessment is expressly carved out.

Biometric practices

The final theme is the one that consistently triggers alarm, from science fiction to unsettling news headlines: your face, your body and your behaviour in public becoming a constantly trackable digital signature. In public spaces, some level of blending into the crowd was always expected; mass biometric analysis dissolves that. The Act gathers four bans here, and they are strict.

Untargeted face scraping

This one is short and unusually concrete. Under Article 5(1)(e), you may not build or expand facial-recognition databases by untargeted scraping of facial images from the internet or CCTV footage. If a company hoovering up billions of face photos from social media to sell search-by-face services sounds familiar, yes: that business model is the target. (The famous fines against exactly such a company were GDPR fines, by the way, not AI Act ones. Keep the two regimes straight.)

Emotion inference at work and school

Picture a receptionist who spends the whole day under a mood camera: software scoring her smile, flagging when she seems irritated, reporting "engagement" to her manager. Under Article 5(1)(f), that system is prohibited. AI may not be used to infer the emotions of a person in the workplace or in education institutions.

There is one exception: systems put in place for medical or safety reasons. Fatigue detection for long-haul drivers or machine operators can stand on the safety leg. But an HR dashboard of employee sentiment, a call-centre tool scoring agents' moods, a classroom camera measuring student attention: these are not edge cases. They are the ban's centre of mass, and in my experience they are also the prohibition that corporate audits most often miss, because emotion-analytics vendors sell them as harmless productivity tools.

Note the boundary: emotion inference *outside* work and education (say, on customers) is not prohibited, but it lands in high-risk territory instead. Banned nowhere near means "fine everywhere else".

Biometric categorisation of sensitive traits

Under Article 5(1)(g), AI must not 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. In plain terms: no “this face looks gay”, no “this gait suggests a protestor”, no inferring religion from appearance. The prohibition has technical carve-outs (labelling or filtering of lawfully acquired biometric datasets, and certain law-enforcement contexts), but the core idea is that your body is not a legitimate input for guessing the most protected facts about you.

Real-time remote biometric identification

And now the prohibition everyone has heard of, usually in garbled form: facial recognition in public spaces.

Here is what Article 5(1)(h) actually says. The ban covers **real-time** remote biometric identification systems, in **publicly accessible spaces**, used **for the purposes of law enforcement**. All three elements matter. Real-time means live scanning of crowds, not a detective later comparing a photo. Publicly accessible spaces means streets, parks, transport hubs. And “for the purposes of law enforcement” means this prohibition is about the police, not about companies.

Even for police, there are three exception categories: targeted searches for victims of abduction, trafficking or sexual exploitation, and for missing persons; prevention of a specific, substantial and imminent threat to life or a genuine, foreseeable terrorist attack; and locating suspects of serious crimes (offences listed in Annex II carrying at least four years’ custody). Even then, each use needs prior authorisation from a judicial or independent administrative authority, a fundamental-rights impact assessment, and registration in the EU database. This is not a loophole you drive a surveillance state through; it is a narrow door with three locks, and each Member State decides in national law whether to open it at all.

Now the correction I find myself making most often. A shopping mall that tags visitors’ faces against social-media profiles to build sellable consumer dossiers feels like it must be covered here. It is not.

WATCH OUT

The Commission's guidelines say expressly that private-actor use of facial recognition, real-time or not, falls outside this prohibition, because the ban is scoped to law enforcement. That does not make the mall scenario legal or safe: it collides head-on with the GDPR's rules on biometric data, and biometric identification systems are high-risk AI under Annex III. Wrong cage, same zoo. But if you cite Article 5 against the mall, its lawyers will win that exchange.

The ninth ban: AI-generated intimate imagery and CSAM

The list is about to grow. The 2026 Digital Omnibus amendment (agreed by Parliament on 16 June and Council on 29 June 2026) adds a new prohibition to Article 5: AI systems for generating non-consensual intimate imagery (the so-called "nudifier" apps) and child sexual abuse material. It becomes applicable after a transition running to 2 December 2026, which makes it the only genuinely short-term deadline among the prohibitions.

The liability split is worth remembering: providers are on the hook if such output is the intended purpose *or* a reasonably foreseeable and reproducible outcome absent safeguards; deployers are liable only for deliberate misuse.

AS OF MID-2026

The Omnibus text was still awaiting Official Journal publication when this book went to press. The political agreement is done and the direction is unmistakable, but verify the exact wording of the new prohibition once the final text is published.

Already the law, still no fine

Here is the fact that surprises every audience I put it in front of. These prohibitions have been enforceable since 2 February 2025. The penalty regime, that top tier of 35

million euros or 7 percent, has been live since 2 August 2025. And as of July 2026, there is no confirmed AI Act fine, warning or formal proceeding anywhere in the Union.

Why? Partly because national enforcement machinery is still being bolted together; many Member States were late designating their market surveillance authorities. Partly because first cases in any new regime take time. The nearest thing to a test case is Hungary's expansion of facial-recognition surveillance to public assemblies, which dozens of NGOs argue violates Article 5(1)(h), but that is advocacy pressure, not an opened proceeding.

> MY TAKE
>
> I would not read the silence as safety. The rules bind you today; the enforcement is ahead of us, not behind us. I would rather count with the worse possibility, at least for now.

Banned, or merely high-risk?

Before we move on, let me give you the intuition that separates this chapter from the next, because the same technology keeps appearing on both sides of the line. Ask three questions: **who** uses it, **on whom**, and **for what purpose**.

Real-time face recognition in a public square, run by police to scan a crowd: prohibited, unless one of the three narrow exceptions applies. The same cameras run by a supermarket: outside the ban, but high-risk and under the GDPR's thumb. Emotion inference on your employees: prohibited. The same emotion model watching for driver fatigue: allowed under the safety exception. Scoring citizens' trustworthiness from their social lives: prohibited. Scoring loan applicants' repayment risk from financial data: permitted, but high-risk.

The prohibitions catch uses where the purpose itself is considered incompatible with fundamental rights: no safeguards can fix them. Everything one notch below lands in the Act's high-risk regime, where the answer is not "no" but "yes, with heavy obligations". That regime is the next chapter.

Check yourself

1. A mobile game's AI individualises bonuses to keep specific players spending, though the developers never intended harm. Under Article 5(1)(a), this is: A) Legal, because the developers had no intent to harm B) Potentially prohibited, because the *effect* of materially distorting behaviour counts, not just the objective C) Legal, because games are entertainment and out of scope D) Only a transparency issue under Article 50

2. Which statement about the social-scoring ban in Article 5(1)(c) is correct? A) It applies only to public authorities B) It bans all scoring of individuals, including bank credit-risk models C) It applies to any actor, public or private, when scores cause detrimental treatment in unrelated contexts or disproportionate treatment D) It was deferred to December 2027 by the Omnibus

3. A police force wants AI to predict who will commit crimes based purely on personality profiling. A separate tool helps officers assess suspects in an existing case built on verified evidence. Under Article 5(1)(d): A) Both are prohibited B) Neither is prohibited C) The first is prohibited; the second falls under the carve-out for supporting human assessment based on objective, verifiable facts D) Only the second is prohibited

4. A company installs cameras that infer employees' emotions for performance reviews; a truck fleet uses cameras detecting driver fatigue. Under Article 5(1)(f): A) Both are prohibited B) The performance-review use is prohibited; the fatigue system can rely on the safety exception C) Both are allowed if employees consent D) Only education institutions are covered, so neither is prohibited

5. A shopping mall runs real-time facial recognition to identify returning customers. Which is the most accurate assessment? A) Prohibited under Article 5(1)(h), which bans all real-time biometric identification in public B) Fully legal, since only governments are regulated C) Outside the Article 5(1)(h) ban (which is scoped to law enforcement), but caught by the GDPR and classified as high-risk AI D) Prohibited unless the mall obtains judicial authorisation

Answers. 1: B. The prohibition covers techniques with the objective *or the effect* of materially distorting behaviour, and the Commission's guidelines read "purposefully" objectively, so no intent to harm is needed. 2: C. The ban is actor-neutral and bites on the two harm branches, while ordinary credit scoring in its own context is high-risk rather than banned. 3: C. The ban applies only to predictions based *solely* on profiling or personality traits, with an express carve-out for AI supporting fact-based human assessment. 4: B. Emotion inference in the workplace is prohibited, but systems intended for medical or safety reasons, like fatigue detection, are excepted. 5: C. Article 5(1)(h) covers real-time remote biometric identification for law-enforcement purposes only, so private commercial use falls outside the ban but into GDPR territory and the high-risk regime.

The mall scenario already gave the game away: one notch below "banned" sits a much larger category, high-risk AI, and deciding what falls into it is the next chapter's whole job.

CHAPTER 7

High-Risk AI: Classification

Come back to our bank loan model for a moment. You feed it a customer's income, payment history and a few dozen other variables, and it predicts whether that person will repay a loan. Nobody gets hurt. Nothing explodes. And yet this unassuming model sits squarely in the AI Act's most heavily regulated category: high-risk AI.

Why? Because the Act does not measure risk by how impressive the technology is. It measures risk by what the system can do to a person's life. A loan refusal can decide whether a family buys a home. An automated CV screener can quietly end a career before it starts. The question is never "how smart is the AI?" but "how much is at stake for the human on the receiving end?"

So this chapter answers the Act's second big question. First: is it AI at all? Second: is it high-risk AI? If yes, a long list of obligations follows; that is the next chapter. The classification lives in Article 6 and Annex III: general rules first, then the escape hatch, then the list itself, area by area.

One date to anchor everything: the high-risk obligations for the Annex III use cases below apply from 2 December 2027, moved there from August 2026 by the 2026 Omnibus amendment. You have time, but classification work is exactly the slow, organisational kind you want to start early.

Two doors into high-risk

Article 6 gives you two separate doors into the high-risk category. It is an OR, not an AND: walking through either one is enough.

Door one: the regulated-product route (Annex I). Suppose you manufacture lifts. Or medical devices, toys, machinery, pressure equipment. You are already drowning in EU product legislation, and Annex I is simply a list of those existing laws. Article 6(1) says: if your AI system is a safety component of such a product, or if the AI

system *is itself* the product, and that product must undergo a third-party conformity assessment under its own legislation, then the AI system is high-risk.

Notice both halves of that sentence. The safety-component half is intuitive: an AI module deciding when a lift's emergency brake engages. But the second half is easy to miss: the AI can *be* the product. Think of diagnostic software that analyses medical images. It is nobody's safety component; it is the medical device itself, regulated under the medical devices framework, and it enters high-risk through this same door. If you only ask "is my AI a safety component?", you will miss an entire class of standalone AI products.

Two Omnibus updates on this route. First, the definition of "safety component" was narrowed: AI used purely for user assistance, optimisation or quality control is out, unless its failure could endanger health or safety. Second, the date: obligations for this Annex I route apply from 2 August 2028. If you work in one of these industries, my advice is simple: open Annex I, find the regulation that already governs you, and work from there. These industries have decades of practice absorbing EU product rules; this is one more layer, not a new world.

Door two: the list (Annex III). This is the door that matters for most readers of this book, because it catches industries never product-regulated before: banks, HR departments, universities, insurers. Annex III is literally a list of use cases across eight areas. If your AI system is intended to be used for one of them, it is high-risk. No safety component required, no product legislation required. Grading student essays with AI? High-risk. The bank loan model? High-risk. Just by being on the list. Most of this chapter walks that list, because in my experience this is where nearly every real classification question lands.

The way out, and the way back in

One more piece of Article 6 before the list: the escape hatch. Landing in an Annex III area does not automatically make your system high-risk. That is one of the Act's more sensible design choices. Article 6(3) provides a derogation: your Annex III system is *not* high-risk if it poses no significant risk to health, safety or fundamental rights, including

by not materially influencing the outcome of decision-making, and one of four conditions holds.

The four conditions, with the flavour of each: the system performs only a **narrow procedural task** (an AI that downloads student submissions from a repository and neatly renames them for the teacher: in the education area, but grading nothing); it **improves the result of a previously completed human activity** (an AI suggesting stylistic polish on essays *after* the teacher has decided the grade); it **detects decision-making patterns or deviations** without replacing human assessment; or it performs a **preparatory task** for an assessment (scheduling patient follow-ups in a hospital, touching no diagnosis).

The common thread: the human decision stays with the human, and the AI stays at the edges.

> MY TAKE
>
> Positioning a system at the edges like this is legitimate design, not cheating. I suspect “derogation engineering” may become a consulting specialty in itself.

But before you get creative, read the sentence that overrides everything.

> THE ACT SAYS · ARTICLE 6(3)
>
> Notwithstanding the first subparagraph, an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons.

In plain terms: if the system profiles people, meaning it processes their personal data to evaluate aspects of their life such as work performance, economic situation, health or behaviour, the derogation is simply unavailable. This kills two tempting escape plans on the spot. Repackaging the bank loan model as a “merely advisory” second opinion, or as a public website widget that “only informs” you whether you would get the loan? It still evaluates a person’s economic situation, which is profiling, so it stays high-risk no

matter how advisory the framing. An HR tool that "only highlights trends" in employee performance data for a manager? Evaluating work performance is textbook profiling; high-risk, full stop. The derogation is for the genuinely peripheral systems (the file-sorters and schedulers), not for softened versions of the core scoring engines.

And if you do claim the derogation, Article 6(4) attaches three duties, not one. You must **document** your assessment before the system goes to market. You must **register** the system in the EU database; yes, even your self-assessed *non*-high-risk system goes into the register under Article 49(2); the 2026 Omnibus lightened what you must file, but the duty stands. And you must **produce the documentation on request** of national competent authorities. The derogation is not a quiet exit: you leave a paper trail, appear in a database, and a regulator can ask to see your work.

Biometrics

Annex III opens with biometrics, "insofar as their use is permitted under relevant Union or national law", and that opening phrase is a warning.

> WATCH OUT
>
> Several Annex III entries sit right next to an Article 5 ban, and the recitals are explicit that the high-risk entries only cover what is *not* already prohibited (Recital 54). Check the prohibited practices first; this trap appears twice in this first area alone.

Three entries here. First, remote biometric identification: a camera system scanning a train station crowd against a database of missing persons or suspects. Verification is explicitly excluded. The airport e-gate checking you are the person in your own passport is not high-risk, because its sole purpose is confirming you are who you claim to be.

Second, biometric categorisation by sensitive or protected attributes. Here is the first trap. I once watched a conference presentation about interactive marketing banners for bank branches: a camera above the screen reads the waiting customer's face and tailors the advertisement. Inferring gender, age or mood that way is high-risk under this entry.

But the moment the system infers *race or ethnicity* from a face, you are not in Annex III any more. Biometric categorisation deducing race is prohibited under Article 5(1)(g), banned since February 2025. Same camera, different attribute, completely different legal universe.

Third, emotion recognition, and here is the second trap. Remember the fancy-hotel receptionist under the mood camera, with the manager rushing out whenever the smile fades? Such systems genuinely existed. But inferring employees' emotions in the workplace is *banned* under Article 5(1)(f), not high-risk; the same goes for schools, with a narrow exception for medical or safety reasons. The Annex III emotion-recognition entry covers what remains outside the ban: say, a retail system reading *customers'* moods to adjust service. That is high-risk, plus a duty to inform the people exposed to it.

Critical infrastructure

Next: AI used as a safety component in the management and operation of critical digital infrastructure, road traffic, or the supply of water, gas, heating or electricity. Read that as an exhaustive list, because it is one. Banking is not in it. Whatever "critical infrastructure" list your bank landed on during the pandemic has nothing to do with this entry, which takes its definition from the EU's critical-entities directive. And "safety component" here has a specific meaning spelled out in Recital 55: something that directly protects physical integrity or health and safety, without being necessary for the system to function.

> THE ACT SAYS · RECITAL 55
>
> Components intended to be used solely for cybersecurity purposes should not qualify as safety components.

In plain terms: your AI-driven cyber-defence tool guarding a cloud platform is *not* an Annex III safety component, however critical it feels. What is? The recital's own examples: fire-alarm controlling systems in cloud computing centres, water-pressure

monitoring. Add AI-optimised traffic signals at a complex junction, smart-grid balancing (a friend working on UK electricity optimisation tells me demand spikes at half-time of big football matches, a nation switching on its kettles), or automated chemical dosing at a water treatment plant. Failure there endangers people at scale. That is the test.

Education and vocational training

Four entries: AI deciding admission to educational institutions; AI evaluating learning outcomes (grading essays, projects, open-ended exams; a multiple-choice test with predetermined answers involves no inference and is likely not AI at all); AI assessing what level of education someone can access, including access to subsidised reskilling programmes; and AI proctoring, the systems watching your webcam and keystrokes during online certification exams. I will admit this is the area of Annex III I like most. Universities *will* need AI grading tools as generative AI multiplies student output; the Act does not forbid them, it just insists they be built carefully.

Employment and workers' management

AI for recruitment and selection (CV screeners ranking hundreds of applicants, targeted job-advertisement platforms) and AI making decisions about promotion, termination, task allocation or performance monitoring. The failure mode here is quiet: train a hiring model on the decisions of past managers who preferred one gender or nationality, and the model dutifully inherits the bias. And to every hiring manager proudly posting on LinkedIn about evaluating candidates with ChatGPT: you are running a high-risk AI use case on a tool that hallucinates. Please stop. One boundary from the biometrics area applies here too: monitoring *performance* is high-risk; inferring employees' *emotions* is banned.

Access to essential services

A mixed bag of four entries, and honestly not my favourite grouping: eligibility for public benefits, creditworthiness, insurance and emergency dispatch have quite

different stakes. Public authorities using AI to grant or revoke benefits like unemployment support: high-risk. Creditworthiness evaluation of natural persons, our bank loan model, is high-risk too, with one carve-out: fraud detection is excepted. Former banking colleagues explained the logic to me: fraudsters change tactics weekly, and fraud models need free hands to react. Then risk assessment and pricing in *life and health* insurance specifically (not car insurance), and finally AI that classifies emergency calls or prioritises the dispatch of police, firefighters and ambulances, including patient triage.

Law enforcement

Predicting who might become a crime victim, AI polygraphs, evaluating the reliability of evidence, and recidivism prediction. This last one needs care, because it sits directly against a prohibition. Read the Annex III entry closely: it covers assessing the risk of a person offending or re-offending *not solely on the basis of profiling*. Why the odd wording? Because Article 5(1)(d) *bans* criminal-risk prediction based solely on profiling or personality traits: the Minority Report scenario. What survives as high-risk is the narrower case, a tool supporting a human assessment already grounded in objective, verifiable facts directly linked to criminal activity.

Even the permitted version carries a famous trap: the predictive-policing feedback loop. Dispatch patrols where the model predicts crime, and officers will find more crime exactly there, which retrains the model to send even more patrols. The system feeds itself and drifts into bias. Scale is what makes this area dangerous.

Migration, asylum and border control

AI assessing security, migration or health risks of people entering a Member State; AI assisting with asylum, visa and residence applications (say, a consulate tool estimating an applicant's likelihood of overstaying from travel history and financial status); AI polygraphs; and identification systems at borders, except for verifying travel documents. Human border checks are untouched; it is the automation of these judgments that triggers the classification.

Administration of justice and democratic processes

Two entries here. First, AI assisting a *judicial authority* in researching and applying the law: legal research tools suggesting precedents to a court, sentencing-recommendation systems. Note the limit. This covers courts, not the ten-euro-a-month legal chatbot sold to consumers. A curious gap, to my mind, given how confidently generative AI hallucinates case law.

Second, AI intended to be used for influencing the outcome of an election or referendum, with an express carve-out for back-office campaign logistics. The load-bearing word is *intended*. A campaign's micro-targeting tool built to steer voters fits this entry. A social platform's general-purpose feed recommender does not: its intended purpose is engagement, not elections, and that layer is governed by the Digital Services Act and the EU's political-advertising regulation.

> WATCH OUT
>
> After Cambridge Analytica, it is tempting to read this entry as "the AI Act regulates social media feeds". It does not; classification rides on intended purpose.

Can the list change?

Annex III is not carved in stone, but the Commission's hands are less free than you might fear. Under Article 7, the Commission can add or modify *use cases* by delegated act, but only within the eight existing areas, and only where the new use case poses a risk equivalent to or greater than those already listed. It cannot conjure a ninth area out of thin air; that would take the full legislative process. The assessment criteria in Article 7(2) cut both ways, too: alongside the harm-focused factors sit the magnitude of the system's *benefits* and whether existing EU law already provides redress, both of which argue against adding items. Removal is possible under its own two cumulative conditions, and Parliament and Council can object to any delegated act.

One more thing for your radar: Article 6(5) required the Commission to publish practical classification guidelines, with worked examples, by February 2026.

> AS OF MID-2026
>
> The Article 6(5) guidelines arrived late and in draft: published 19 May 2026, consultation open until 23 July 2026, final version promised by year end. No Article 7 delegated act has been adopted, and the 2026 Omnibus left the Annex III list untouched.

The draft's Annex III volume alone runs to nearly 150 pages of worked examples, the closest thing to official answers for "is my system high-risk?" questions. Treat it as strong indicative guidance, not settled law; I would not bet a compliance programme on wording that may still shift by December.

Where does this leave you? With a checklist. Is it AI? Is it in a regulated product (Annex I), as safety component or as the product itself? Is its intended purpose on the Annex III list? If yes, does a derogation condition genuinely apply, and does the system profile people, which slams that door shut? Document, register, be ready to show your work. Classification is answerable; it just rewards reading the list slowly.

Check yourself

1. A company sells standalone AI software that analyses X-ray images to detect fractures. It is regulated as a medical device requiring third-party conformity assessment. Is it high-risk under the AI Act? A) No; it is not a safety component of any product. B) Yes, under the Annex I route, because the AI system is itself the product. C) Only if it also appears in Annex III. D) No; medical software is exempt from the AI Act.

2. A hotel installs a camera that reads receptionists' facial expressions so a manager can intervene when their mood dips. Under the AI Act, this system is: A) High-risk under Annex III point 1(c), emotion recognition. B) Limited-risk, requiring only a transparency notice. C) Prohibited; inferring employees' emotions

in the workplace is banned under Article 5(1)(f). D) Outside the Act, because the camera does not face customers.

3. A bank repositions its loan-approval model as "purely advisory": a human always makes the final decision. Can the bank use the Article 6(3) derogation to escape high-risk classification? A) Yes; the system now merely improves a previously completed human activity. B) Yes; advisory systems are excluded from Annex III entirely. C) Only if it registers the system first. D) No; creditworthiness evaluation profiles natural persons, and profiling systems in Annex III are always high-risk.

4. A provider self-assesses its Annex III system as not high-risk under the derogation. What must it do? A) Nothing; the derogation applies automatically. B) Document the assessment, register the system in the EU database, and provide the documentation to authorities on request. C) Obtain approval from a notified body before launch. D) Publish the assessment on its website within 30 days.

5. Which of the following can the European Commission do to Annex III by delegated act under Article 7? A) Add an entirely new area, such as "retail and e-commerce". B) Add or modify use cases within the eight existing areas, where the risk is at least equivalent to the use cases already listed. C) Rewrite the classification rules in Article 6. D) Nothing; Annex III can only be changed by a new regulation.

Answers: 1: B. Article 6(1) covers AI that is a safety component *or is itself a product* under Annex I legislation requiring third-party conformity assessment. 2: C. Workplace emotion inference has been a prohibited practice since February 2025; the Annex III emotion-recognition entry only covers systems the ban does not reach, such as customer-facing ones. 3: D. Article 6(3)'s final subparagraph makes profiling systems always high-risk, so no amount of "advisory" repositioning reaches the derogation. 4: B. Article 6(4) imposes three duties: documentation before market, registration under Article 49(2), and production on request. 5: B. Article 7 permits adding or modifying use cases only inside the existing areas and only under cumulative risk conditions; new areas would need the full legislative process.

So your system is on the list, profiling people, with no way out: the next chapter walks through what a high-risk provider actually has to build, document and prove.

CHAPTER 8

High-Risk AI: The Obligations

So the verdict is in. Your bank's loan model, the one deciding mortgages while the customer sips complimentary coffee, sits on the Annex III list, none of the previous chapter's escape routes apply, and it is officially high-risk. What now?

Now comes the part of the Act where the real work lives: Articles 8 to 15, Chapter III, Section 2. Seven substantive requirements: risk management, data governance, technical documentation, record-keeping, transparency towards deployers, human oversight, and the triple-header of accuracy, robustness and cybersecurity. In real organisations this is a cross-functional team (data scientists, lawyers, compliance officers, domain experts) working for months. This chapter will not turn you into that team; it will give you an honest picture of what each article demands, so that when the team is assembled, you know what they are building and why.

Orientation first: for Annex III systems these obligations apply from 2 December 2027, moved to that date by the 2026 Omnibus amendment, and for Annex I embedded systems from 2 August 2028. Fixed calendar dates: distant, but the work is months-deep, which is exactly why they were pushed.

Article 9: the risk management system

Say our bank builds an AI hiring tool, high-risk under Annex III's employment category. Article 9 says a risk management system shall be established, implemented, documented and maintained: not a checklist you complete once, but a continuous, iterative process across the system's entire lifecycle. A scheduled check-up, not a one-time vaccination.

The process has four recurring steps: identify and analyse the known and reasonably foreseeable risks to health, safety or fundamental rights; estimate and evaluate them under intended use and reasonably foreseeable misuse; keep evaluating new risks from post-market monitoring data once the system is live; and adopt targeted measures for what you found. Article 9(3) helpfully limits the exercise to risks you can actually

address through design, development, or adequate deployer information. After mitigation, the residual risk must be judged acceptable; what cannot be engineered away, you mitigate and disclose.

In practice the bank appoints a cross-functional team (usually people giving 10 to 20 percent of their week) that runs workshops on what could go wrong: bias in candidate selection, misread CVs, privacy leaks. Each risk gets a likelihood, a severity, a documented mitigation; vulnerable groups, including minors, get extra care under Article 9(9).

Then comes testing. Article 9(8) requires testing at any appropriate point during development and, in any event, before market placement, against metrics and probabilistic thresholds you defined in advance. That much is mandatory. Testing in real-world conditions, meaning trying the tool on live job applicants, is optional, and regulated: Article 9(7) allows it only "in accordance with Article 60", which brings a real-world testing plan, registration, market-surveillance oversight and, under Article 61, the informed consent of the people you test on. Quietly running your unfinished hiring model on applicants who never agreed to be test subjects is not diligence; it is itself a violation.

Article 10: data and data governance

Raise Article 10 with clients and someone says: "Data governance? We've done that for years. GDPR, done." That reflex is exactly wrong. GDPR governs personal data wherever it flows. Article 10 governs something narrower and deeper: the training, validation and testing datasets your model learns from, because, as chapter 1 established, whoever shapes the data shapes the behaviour.

Take the bank's creditworthiness model. Article 10(2) demands documented practices covering the design choices, the data's origin and original purpose, its preparation (cleaning, labelling, enrichment), the assumptions it embeds, whether there is enough of it, and, crucially, an examination for biases likely to harm health, safety or fundamental rights, plus measures to detect, prevent and mitigate them. Article 10(3) adds that the datasets must be relevant, sufficiently representative, and to the best

extent possible free of errors and complete for the intended purpose. If the model will score applicants across several regions, the data had better reflect those regions and their socioeconomic mix. A compliance gap can exist before a line of code is written.

One genuinely surprising provision. To check whether the model discriminates by ethnicity or health status, you need the sensitive attribute, the very data you normally must not touch. Article 10(5) allows providers, exceptionally, to process special categories of personal data strictly for bias detection and correction, under stacked safeguards: the job must be impossible with other data (synthetic or anonymised included), access strictly controlled and documented, no onward transfer, deletion once the bias is corrected. The Act would rather you handle sensitive data carefully than stay ignorant of your model's discrimination.

Article 11: technical documentation

Article 11 sounds like the boring one and is anything but optional. The technical documentation must be drawn up *before* the system is placed on the market, not reconstructed afterwards, and kept up to date. Its job: demonstrate, to a national authority or notified body reading it cold, that the system complies with everything in this section.

You do not get to invent the outline: Article 11(1) says the documentation shall contain, at a minimum, the elements of Annex IV, nine of them:

1. A general description: intended purpose, provider, versioning, interactions with other hardware and software, the forms it ships in (API, embedded, download), the hardware it runs on, the deployer's interface and instructions.
2. A detailed description of the system and its development: methods, pre-trained third-party components, design choices and trade-offs, architecture, data provenance, human oversight assessment, validation and testing with signed test reports, cybersecurity measures.
3. Detailed information on monitoring, functioning and control: capabilities and limits, accuracy for specific groups, foreseeable unintended outcomes, input specifications.

4. Why your chosen performance metrics are appropriate for this system.
5. A detailed description of the Article 9 risk management system.
6. Relevant changes made through the lifecycle.
7. A list of harmonised standards applied or, where none, a description of the solutions you adopted instead (as of mid-2026 that is everyone; more on this at the end).
8. A copy of the EU declaration of conformity (Article 47).
9. A description of the post-market monitoring system and plan (Article 72).

Two mercies: the contents apply "as applicable", proportionate to the system, and SMEs, including start-ups, may use a simplified form the Commission is to provide, which notified bodies must accept.

Article 12: record-keeping

Article 12 asks that high-risk systems technically allow the automatic recording of events (logs) over their lifetime. Picture a city's AI traffic-signal controller: every signal change, every anomaly, every alert for human intervention, timestamped, so a 3 a.m. malfunction can be reconstructed: what did the system see, what did it decide?

But the article is graded. For most high-risk systems, Article 12(2) requires logging *appropriate to the intended purpose*: enough to identify situations where the system may present a risk or has been substantially modified, to feed post-market monitoring, and to let deployers monitor operation. What "enough" means for a traffic controller differs from a credit model. You design it, you justify it.

The famous detailed list (start and end of every use, the reference database checked against, the input data that produced a match, the identity of the humans who verified the result) is Article 12(3), and it applies only to remote biometric identification systems under Annex III point 1(a).

> WATCH OUT
>
> “Reference database” here means the database of faces the system matches against, not your data warehouse. And if a vendor’s brochure claims every high-risk system must log those four items, the vendor is wrong.

Article 13: transparency towards deployers

A situation the Act anticipates neatly: you are a provider selling your high-risk loan-scoring system to other banks, deployers in the Act’s vocabulary. Your Article 11 documentation contains everything: architecture, training data, trade secrets, know-how. Hand that to customers and you have handed it to competitors.

Article 13 is the answer: a second, outward-facing document, the instructions for use, which must be concise, complete, correct and clear. At minimum: who you are; the system’s intended purpose, capabilities and limitations; the accuracy metrics it was tested against, its robustness and cybersecurity, and what could degrade them; foreseeable risks and misuse; where applicable, performance on specific groups; the input data it needs; how to interpret the output (“scores above 70 typically indicate approval”); the human oversight measures; computational resources and expected lifetime; and how the deployer can collect and read the logs. So a provider maintains two artefacts: the thick internal file for the regulator, and the honest user manual for the customer, which Article 26 expects the deployer to actually use.

Article 14: human oversight

Article 14 requires high-risk systems to be designed so that natural persons can effectively oversee them while in use, to prevent or minimise risks to health, safety and fundamental rights. Not oversight as a policy statement: oversight designed into the product, through the human-machine interface itself.

Article 14(4) spells out what the overseeing human must be enabled to do: understand the system’s capacities and limits; stay aware of **automation bias**, our well-documented tendency to over-trust a confident machine; correctly interpret the output;

decide not to use the system or to override its result; and intervene or halt it through a stop button or similar. Some measures the provider builds in, others it identifies for the deployer to implement (Article 14(3)): a shared design problem.

Two sketches. MediScan AI suggests diagnoses from X-rays with a confidence score; every suggestion is reviewed by a radiologist, low-confidence cases go to a second radiologist, and a stop function reverts to manual analysis if the system starts producing nonsense. LoanDecider AI scores applications from 0 to 100; mid-range scores go to a loan officer, consequential denials alert a senior one, every override is logged. Different domains, same grammar: thresholds, escalation, training against automation bias, override, audit trail.

Then there is the biometric special case, where the Act stops suggesting and starts ordering.

> **THE ACT SAYS · ARTICLE 14(5)**
>
> For high-risk AI systems referred to in point 1(a) of Annex III, no action or decision may be taken by the deployer on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority.

Two humans, by law, with a carve-out where Union or national law considers that disproportionate in law-enforcement, migration, border-control or asylum contexts. A statutory rule, not a design suggestion.

And be careful with the scenario people most readily imagine here: police scanning live camera feeds in public spaces for wanted individuals. That is mostly not an oversight design problem: real-time remote biometric identification in publicly accessible spaces for law enforcement is a *prohibited* practice under Article 5(1)(h), lawful only for three narrow objectives, only where national law enables it, and only with prior judicial or independent authorisation. Good oversight design cannot launder a prohibited practice into a permitted one. The two-person rule lives in lawful biometric territory, such as after-the-fact identification against footage of a serious crime.

Article 15: accuracy, robustness, cybersecurity

Article 15 is short and readable. It is also, I would argue, the most expensive article in the section: each of its three headline words hides an engineering discipline.

Accuracy. The system must achieve "an appropriate level of accuracy" and perform consistently throughout its lifecycle, with the levels and metrics declared in the instructions for use (Article 15(3)). Who decides what "appropriate" means? Mostly you do. That is not a loophole; it is the design. There are no mandatory EU accuracy benchmarks: Article 15(2) only tasks the Commission with *encouraging* benchmarks and measurement methodologies, and as of July 2026 none exist. For most Annex III systems, Article 43(2) sends you down the internal-control route of Annex VI: you assess your own conformity and must be able to defend it. Declare a metric, justify it, meet it, keep meeting it.

Choosing the metric is where domain judgment bites. Take a model that screens for Parkinson's disease from a person's gait on video. Its headline "accuracy", the share of all predictions that are correct, can be a nearly pointless number. The catastrophic error is the false negative: telling someone who has the disease that they are fine, so they never see a doctor. A false positive merely sends a healthy person for an all-clear. For that system you care about recall: of the people who truly have the disease, what fraction do we catch? A different system demands a different choice, which is why Annex IV point 4 makes you justify your metrics in writing.

Then there is "consistently throughout their lifecycle". Models are frozen at deployment; the world is not. Patterns drift, like inflation slowly invalidating a rule learned on last year's incomes, or snap, as COVID-19 snapped consumer behaviour overnight. Compliance here means monitoring for drift and retraining on fresh data before performance sinks below what you declared.

Robustness. The system must be as resilient as possible against errors, faults and inconsistencies, through technical and organisational measures: redundancy, backups, fail-safe plans.

> MY TAKE
>
> This is the requirement that will hurt culturally. Data scientists were trained for a decade to build systems that *learn* well; making them *fail* well was someone else's job. Expect the robustness bill to be paid in engineering practices (careful feature pipelines, handling of malformed inputs, stress tests) that many AI teams have never been asked to master.

Cybersecurity. Article 15(5) is the most concrete: the system must resist attempts by unauthorised third parties to alter its use, outputs or performance, with measures, where appropriate, against four named AI-specific attacks. Data poisoning is corrupting the training data: the trick from chapter 1, where fake "customers" fed fabricated records into the bank's loan model until its decision boundary shifted. Model poisoning aims the same idea at pre-trained components: the open-source model you bolted into your pipeline arrives with someone else's tampering baked in. Adversarial examples are inputs crafted to make the model err. Confidentiality attacks coax the model into leaking its training data or internals. Classic IT security still applies; this list is what is new now that the software learns.

Practical implications

Step back and the shape is clear: Articles 8 to 15 describe a mature engineering organisation, one that knows its risks, its data, its performance, and can prove all of it on paper, before and after deployment. Article 8 calibrates it all: compliance is judged against the system's intended purpose and the state of the art.

A worry I hear from teams constantly: "If our prototype counts as high-risk AI, doesn't every experiment now need documentation, risk files, some regulator's approval? That would triple the cost of trying ideas." The Act answers it directly.

> THE ACT SAYS · ARTICLE 2(8)
>
> This Regulation does not apply to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service. Such activities shall be conducted in accordance with applicable Union law. Testing in real world conditions shall not be covered by that exclusion.

In plain terms: the obligations attach when you place the system on the market or put it into service, not while you tinker in the lab. Your backlog of raw ideas, your three-month prototypes, the nine of ten experiments that quietly die: none of it is regulated by the AI Act. No duty to submit prototypes anywhere; the regulatory sandboxes (Articles 57 to 59) are voluntary support, not a gate.

Notice the last sentence, though: the moment your "testing" involves real people, trying your hiring model on actual job applicants, you are back inside the Article 60 regime from Article 9.

And the detailed "how" is still forming.

> AS OF MID-2026
>
> Zero harmonised standards under the AI Act have been cited in the Official Journal, so nobody can yet claim the presumption of conformity that following one will eventually give; everyone documents their own solutions under Annex IV point 7. An ISO/IEC 42001 certificate, whatever the consultant selling it says, is useful scaffolding, not AI Act compliance, and grants no legal presumption.

Take any "AI-Act-certified" badge with a large grain of salt, and run your specific case past your legal counsel. The comfort is the calendar: December 2027 is far enough away to do this properly.

Check yourself

1. Your team wants to build a quick internal prototype of a high-risk-candidate AI idea. The Articles 8 to 15 obligations: A) Apply from the first line of code, since the use case is high-risk B) Apply only at placing on the market or putting into service; pre-market R&D is excluded (Article 2(8)) C) Apply only after a national sandbox approves the prototype D) Never apply to systems developed in-house

2. The bank wants to test its unreleased AI hiring tool on real job applicants. Under the Act, this is: A) Mandatory under Article 9 before market placement B) Freely allowed, since the system is not yet on the market C) Optional, and lawful only under the Article 60/61 real-world-testing regime, including informed consent of the subjects D) Prohibited in all circumstances

3. The detailed logging list in Article 12(3) (period of each use, reference database, matching input data, identity of verifying persons) applies to: A) All high-risk AI systems without exception B) Only remote biometric identification systems under Annex III point 1(a); others need purpose-appropriate logging C) Only systems used by public authorities D) Only systems that continue to learn after deployment

4. Which statement about accuracy under Article 15 is correct as of 2026? A) The EU has published mandatory minimum accuracy benchmarks per Annex III use case B) A notified body sets the required accuracy level for every high-risk system C) Accuracy only matters at the moment of market placement, not afterwards D) The provider declares an appropriate accuracy level and metrics in the instructions for use; the Commission only encourages benchmark development, and most Annex III systems self-assess

5. Before a deployer acts on a match from an Annex III point 1(a) biometric identification system, the identification must generally be: A) Confirmed by the provider's customer support B) Separately verified and confirmed by at least two competent natural persons C) Logged, with no human confirmation required D) Re-run three times by the system itself

Answers: 1: B. Article 2(8) excludes pre-market research, testing and development from scope; sandboxes are voluntary. 2: C. Article 9(7) says testing *may* include real-

world testing “in accordance with Article 60”, which brings a plan, registration, oversight and informed consent under Article 61. 3: B. Article 12(3) is expressly limited to Annex III point 1(a) biometric identification; everyone else gets purpose-appropriate logging under Article 12(2). 4: D. Article 15(2) only asks the Commission to encourage benchmark development, Article 15(3) makes the provider declare the metrics, and Article 43(2) routes most Annex III systems to self-assessment. 5: B. Article 14(5) requires separate verification by at least two competent natural persons, with a narrow carve-out where law deems it disproportionate.

Those were the heavyweight duties for the high-risk minority; next come the transparency obligations that reach far more systems, including the chatbot you talked to this morning.

CHAPTER 9

Transparency Obligations

It is two in the morning and you are typing into a customer-support window: “Hi, I’m Anna, how can I help you today?” Anna answers instantly, in full sentences, with a cheerfulness no human possesses at 2 a.m. Is Anna a person? You suspect not. But suspecting is not knowing, and the company has every commercial incentive to keep you guessing, because research keeps suggesting that people value AI-produced work less than the human kind. If I replaced myself in this course with a digital avatar and admitted it, you would probably rate the course lower for that reason alone.

Chapter IV of the AI Act exists to close that gap: the gap between what the AI is and what you are told it is. It is the shortest chapter in the whole Regulation: one article, Article 50. But do not let the size fool you. While the high-risk machinery of Chapter III applies to a relatively narrow list of systems, Article 50 touches almost every company that runs a chatbot, generates images, or publishes AI-written text. For many readers of this book, this chapter is the part of the AI Act that will actually land on your desk first.

And it lands soon. The 2026 Digital Omnibus amendment, agreed by Parliament on 16 June and Council on 29 June 2026 and still awaiting Official Journal publication as I write, pushed the big high-risk deadlines back to December 2027 and August 2028. It did **not** push Article 50. The transparency obligations apply from **2 August 2026**, exactly as originally scheduled.

> WATCH OUT
>
> If you filed the whole AI Act under “that’s a 2027 problem now”, this chapter is the correction. The Omnibus left Article 50 alone: for transparency, compliance is a next-month problem, not a next-year one.

Article 50 contains four distinct duties. The easiest way to keep them straight is to ask two questions about each: *who* owes the duty (the provider who built the system, or the

deployer who uses it), and *what* must be disclosed. We will take the people-facing duties first, then the content duties.

Tell people they are talking to a machine

Article 50(1) is the Anna rule. Providers must ensure that AI systems intended to interact directly with natural persons are designed so that those persons are informed they are dealing with an AI system.

Notice the narrowing. This does not cover every AI system you operate, only those intended to interact *directly* with people. Your customer-facing chatbot: yes. The credit scoring model from our bank example, quietly deciding loan applications in the back office: no. Customers never converse with it, so there is nothing to announce. (The bank model has its own, much heavier obligations; that was the last two chapters.)

You have seen this rule in the wild for years: the little robot icon, the "Hi, I'm a virtual assistant" opener. From August 2026 that courtesy becomes a legal requirement. I have to say I like this one. Given the incentive to pass AI off as human, a blunt "you must say so" is exactly the kind of rule that costs honest companies nothing and catches the others.

There are two escape hatches, and both are sensible. First, no disclosure is needed where it is obvious "from the point of view of a natural person who is reasonably well-informed, observant and circumspect". If a customer walks into your branch and talks to a visibly physical robot, you do not need a sign saying "this robot is a robot". Second, the duty does not apply to AI systems authorised by law for detecting, investigating or prosecuting criminal offences, the same law-enforcement carve-out we saw in the prohibited practices chapter. Imagine police using a synthetic voice in a sting operation; they will not pause to disclose it.

When the system reads you: emotion recognition and biometric categorisation

The second people-facing duty is the one almost nobody talks about, which is precisely why I want you to remember it. Article 50(3): deployers of an emotion recognition system or a biometric categorisation system must inform the natural persons exposed to it that the system is operating, and must process the personal data involved in line with the GDPR and its law-enforcement siblings.

Notice who owes this one: the *deployer*, the organisation using the system, not the company that built it. And notice the trigger: mere *exposure*. The person does not have to interact with anything; if the camera is reading them, they must be told.

You may remember our receptionist under the mood camera from the prohibited practices chapter, the employer inferring emotions of a person at work. That exact scenario does not end up here; it ends up banned outright under Article 5, because emotion recognition in the workplace and in education is prohibited. Article 50(3) governs what is left: the lawful uses. A retailer running emotion analysis on customers browsing the store, a call centre whose software estimates the caller's frustration level, a system sorting people by biometric categories at an event: the system may be allowed to run, but the people exposed to it have a right to know it is running. Quietly reading customers' faces without a word stops being a growth hack in August 2026 and becomes an infringement.

The usual law-enforcement exception applies, where permitted by law and with safeguards. For everyone else: if your system reads people, you tell people.

Synthetic content versus deepfakes

The remaining two duties are about the content AI produces, and before either of them makes sense, we need two terms straight: *synthetic content* and *deepfake*.

Synthetic content is the broad category: anything AI-generated, whether audio, image, video or text. It is tempting to say "anything that passes through an AI is synthetic content", and if you read only Recital 133, which talks broadly about AI systems generating "large quantities of synthetic content", you might land there. But the Act itself draws a boundary, right inside Article 50(2), the marking provision we will meet in a moment: the obligation does not apply where the AI performs "an assistive function

for standard editing" or does "not substantially alter the input data provided by the deployer or the semantics thereof".

So: you write your own report and ask ChatGPT to proofread it. Grammar, typos, a smoother sentence here and there. The meaning is yours; the AI assisted. That output does not need marking. You type "write me an essay on the dangers of AI" and paste the result: that is synthetic content, and the marking duty applies. The line is whether the AI substantially altered your input or its meaning, not whether an AI was anywhere in the room. The Commission's draft guidelines push the boundary further out on the other side too: recommendation lists, generated source code, short sequences of numbers, machine-to-machine outputs. None of these count as synthetic content for the marking duty.

This also answers the long-running Photoshop debate. AI-powered spot removal and colour correction on your own photo is assistive standard editing. Prompting a full scene into existence from nothing is generation. The two ends are clear; the middle will keep lawyers busy.

The second term the Act does define directly, and the definition has a twist.

> THE ACT SAYS · ARTICLE 3(60)
>
> 'deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful;

In plain terms: a deepfake is the subset of synthetic content that imitates something *real* (an existing person, place or event) convincingly enough that someone could take it for genuine. Both halves must hold.

MY TAKE

That second half genuinely bothers me: it hands a defence to exactly the wrong people. “Yes, we generated it, but look how unconvincing it is. Most viewers spotted it, so legally it is not even a deepfake.” Why give bad actors that argument? I do not have a good answer; the definition is what it is.

Keep the intuitive picture: synthetic content is the big circle, deepfakes the smaller circle inside it. The big circle must be *marked* so machines can spot it; the smaller one must additionally be *disclosed* to the humans looking at it. In that order.

Marking synthetic content

Now the duty where the ground has moved most recently. Article 50(2) targets providers of AI systems, explicitly including general-purpose AI, that generate synthetic audio, image, video or text. Think OpenAI, think Midjourney. Here is the operative sentence, because the exact wording matters.

THE ACT SAYS · ARTICLE 50(2)

Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.

In plain terms: another machine must be able to look at the content and tell that AI made it. Note that these are *two* cumulative requirements, not one: the output must carry a machine-readable mark, **and** there must be a working way to detect it. The Commission’s draft guidelines are explicit that a watermark nobody can actually read does not count. Marking without detectability will not suffice. The boundary from the previous section does the gatekeeping: the proofread report stays out, the from-scratch essay is in.

How do you actually mark content? When the Act was written, this was open speculation. Watermarks? Metadata? New file formats? Text was the obvious headache: an invisible watermark survives inside an image file, but text gets copy-pasted, and a visible “written by ChatGPT” footer is deleted in one keystroke. The “how” finally has official shape, in two documents.

> AS OF MID-2026
>
> In May 2026 the Commission published draft guidelines on the Article 50 obligations, with the final version expected before the August application date. On 10 June 2026 the final Code of Practice on Transparency of AI-Generated Content followed, a voluntary compliance tool for the marking and labelling duties, now in the Commission’s adequacy assessment. Both documents are still moving; check the current versions before relying on either.

It is the same mechanism you will meet with general-purpose AI in the next chapter: sign the code, follow it, and you have a strong practical argument that you comply. If you provide a generative system, these two documents are your reading list; invisible watermarking and embedded metadata are exactly what they discuss.

Two clarifications from the guidelines cut against intuitions you might have.

First, the “as far as technically feasible” language in Article 50(2), the part about costs of implementation and the state of the art, conditions the *quality* of your technical solution, not *whether* you must mark. You cannot reason “my images are stylised fantasy art, not realistic, so the marking rules apply to me less”. There is no realism gradient in Article 50(2); realism matters for the separate deepfake question, which is coming below. The guidelines do allow narrow, cumulative exceptions on proportionality grounds (closed embedded environments like in-vehicle navigation, or strictly technical business-to-business outputs like engineering designs), but public-facing content does not get in through that door.

Second, if you were wondering whether social media platforms would be forced to scan and label every upload: under Article 50, no. A platform that merely disseminates

content, with no authority over the AI system that made it, is not a deployer here. The guidelines only *encourage* platforms to preserve the marking. Platform-side labelling duties, where they exist, live in other legislation, not in this article.

One date wrinkle, courtesy of the Omnibus: the marking duty gets a grace period to **2 December 2026**, but only for generative systems already on the market by 2 August 2026. Put a new generative system on the market after that date and you must comply from day one. This is a four-month accommodation for existing systems, not a deferral of the rule.

Deepfakes: allowed, but labelled

Finally, the duty everyone has heard of. Article 50(4) says deployers of an AI system that generates or manipulates image, audio or video content constituting a deepfake must disclose that the content has been artificially generated or manipulated. Notice what that does not say: it does not say deepfakes are forbidden. When I first read the AI Act, this was my biggest surprise: deepfakes are, as a baseline, *allowed*; you just have to label them. There is a softened rule for evidently artistic, satirical or fictional works, where disclosure only needs to happen in a way that does not spoil the enjoyment of the work: a discreet note in the credits rather than a banner across the film. And yes, the law-enforcement exception appears here too.

I have been open about disliking this design. Deepfakes get cheaper every year: less reference data needed, no technical skill required, a few dollars a month. Meanwhile voice-verification banking and “mum, I need money” phone scams show exactly where this goes. A disclosure label does very little against someone whose entire business model is deception.

The EU has, since, partially conceded the point. The 2026 Digital Omnibus adds a new prohibition to Article 5: AI-generated **non-consensual intimate imagery** (the “nudifier” apps) and **child sexual abuse material**, applicable after a transition running to **2 December 2026**. So the single worst deepfake category is moving from “disclose it” to “banned outright”, with liability for providers where such output is a reasonably foreseeable and reproducible outcome of their system, and for deployers

who misuse systems deliberately. Everything else (the fake politician, the fake product endorsement) remains in the disclosure regime. Whether that holds, we will see.

A second paragraph of Article 50(4) is easy to miss and genuinely matters. Deployers who use AI to generate or manipulate **text** that is *published with the purpose of informing the public on matters of public interest* must disclose that too. AI-written news is covered, not just AI-made video. But there is a carve-out that will do a lot of work in practice: the duty does not apply where the AI-generated text has undergone human review or editorial control and a natural or legal person holds editorial responsibility for it. In plain terms: a newsroom where an editor reviews the AI draft, signs off, and stands behind it publicly owes no AI-disclosure label. A website auto-publishing AI-written "news" with no human accountable for it does. The rule is less about how the words were produced and more about whether a human is answerable for them.

The small print that ties it together

Article 50(5) sets the manner and the moment: all these disclosures must be clear, distinguishable, and delivered at the latest at the *first* interaction or exposure, not buried in paragraph 14 of the terms of service, and they must meet accessibility requirements. Article 50(6) adds that none of this replaces Chapter III: a high-risk system that also chats with customers owes both sets of obligations.

Because dates are where this chapter is most easily gotten wrong, here is the timeline in one place:

DATE	WHAT APPLIES
2 August 2026	All four Article 50 duties (not deferred by the Omnibus)
2 December 2026	End of the marking grace period for generative systems already on the market by 2 Aug 2026
2 December 2026	New Article 5 prohibition on AI-generated NCII and CSAM becomes applicable

DATE	WHAT APPLIES
2 December 2027	For contrast: high-risk Annex III obligations. This is what moved, not Article 50

The usual caveat: take this with a grain of salt and check the final texts with your legal representative. The Omnibus was, as I write, still awaiting Official Journal publication, and the Article 50 guidelines are still in draft. The dates above are the agreed ones, but "agreed" and "in the Journal" are not quite the same thing.

Check yourself

1. Your bank's website chatbot answers customer questions. Under Article 50(1), who must ensure that customers are told they are talking to an AI system? A) The deployer of the chatbot B) The provider, by designing the system so users are informed C) The national market surveillance authority D) Nobody, because chatbot disclosure is voluntary until December 2027

2. A marketing manager writes her own press text and asks a generative AI tool only to fix grammar and typos. Under Article 50(2), does the output need machine-readable marking? A) Yes, anything that passes through an AI system is synthetic content B) Yes, but only if the text looks realistic C) No, because assistive standard editing that does not substantially alter the input or its meaning is outside the marking duty D) No, because text is never covered by Article 50(2)

3. What did the 2026 Digital Omnibus amendment do to the Article 50 transparency obligations? A) Deferred them to December 2027 together with the high-risk rules B) Left the 2 August 2026 date in place, adding only a grace period to December 2026 for marking by generative systems already on the market C) Made them conditional on harmonised standards being published D) Repealed the deepfake disclosure duty

4. A retail chain runs software that estimates the emotions of customers from in-store cameras (outside any prohibited setting). What does Article 50(3) require? A) Nothing, as long as the footage is deleted within 24 hours B) A

conformity assessment before deployment C) The provider must watermark the camera feed D) The deployer must inform the exposed persons that the system is operating, and handle the personal data in line with the GDPR

5. A news outlet publishes an article on election results drafted by an AI system, reviewed by a human editor who holds editorial responsibility for it. Under Article 50(4), must the outlet label the article as AI-generated? A) No, because the human review and editorial responsibility carve-out removes the disclosure duty B) Yes, all AI-generated text must always carry a label C) Yes, political topics can never use the editorial carve-out D) No, because text is only covered when it constitutes a deepfake

Answers: 1. B: Article 50(1) places the duty on providers, who must design and develop the system so that natural persons are informed, unless it is obvious from context. 2. C: Article 50(2) expressly excludes AI performing an assistive function for standard editing or not substantially altering the input data or its semantics. 3. B: Article 50 was not deferred and applies from 2 August 2026; the Omnibus only granted a marking grace period to 2 December 2026 for pre-existing generative systems. 4. D: Article 50(3) obliges deployers of emotion recognition and biometric categorisation systems to inform exposed persons and to process personal data under the GDPR and related rules. 5. A: The second subparagraph of Article 50(4) disapplies the disclosure duty where the text underwent human review or editorial control and a person holds editorial responsibility for the publication.

The marking duty you just met already named the systems that will dominate the next chapter: general-purpose AI models, the ChatGPTs of the world, and the special regime the Act built just for them.

CHAPTER 10

General-Purpose AI

Everything in this book so far has started from the same question: what does this system *do*? The CV screener ranks job applicants, so it lands in Annex III. The mood camera watches the receptionist's face, so it collides with the emotion-recognition rules. The whole risk ladder assumes you can point at a purpose and classify it.

Now try that with ChatGPT. It translates contracts in the morning, drafts a breakup text at lunch, plays the interviewer sitting across from you at a mock job interview in the afternoon, and offers a passable opinion on your skin rash in the evening. What is its purpose? All of them. None in particular. The purpose-based ladder we spent three chapters climbing simply has no rung for a system that can do almost anything.

The EU's answer is Chapter V of the Act: a separate regime, aimed squarely at the very large models behind these tools, the ones the industry calls foundation models and the Act calls **general-purpose AI models**, GPAI for short. This chapter is about that regime: who falls into it, what every provider in it owes, and the second ladder inside it, which climbs from ordinary GPAI to GPAI with systemic risk. That ladder mirrors the high-risk ladder you already know, but it must never be confused with it, and we will be careful about that when we get there.

A model is not a system

Before anything else, one distinction that the rest of this chapter stands on. ChatGPT is not a model. ChatGPT is an *application* (a chat interface, safety filters, a subscription page) built on top of a model called GPT. The model is the raw engine; the system is the car built around it. The Act keeps these two apart with almost religious discipline, and so should you.

Here is the definition, from Article 3(63), and this is one of the few places where the exact wording earns a box of its own.

> THE ACT SAYS · ARTICLE 3(63)
>
> ‘general-purpose AI model’ means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications…

In plain terms: it is an engine that (1) can competently do many different things, not one thing, and (2) is built to be embedded into other people’s products. Both halves matter. A model that only predicts loan defaults, however sophisticated, is not general-purpose; it does one job. And the “integrated into downstream systems” part is the quiet heart of the definition: the real business of these models is not the chat window you type into, it is the thousands of applications other companies build on top of them. A startup wiring GPT into a medical-triage product is exactly the scenario the drafters had in mind.

The definition also name-checks *self-supervision at scale*, which deserves thirty seconds. In supervised learning, the bank’s historical data tells the model which customers repaid their loans; the labels supervise. Large language models use a different trick. Take any sentence from any book, say “the cat sat on the…”, hide the last word, and ask the model to guess it. The original text contains the answer, so the model can grade itself, no human labelling required. Do this across a good fraction of the written internet and you get the fluent generalist we started with. The data supervises itself; hence, self-supervision.

One more definition while we are here: Article 3(66) defines a **general-purpose AI system** as an AI system built on a general-purpose AI model. ChatGPT is the system; GPT is the model. Chapter V regulates the *model* and its provider. The moment somebody builds a system on top (including the model’s own maker), the system rules from the rest of the Act apply to that system as usual.

What every GPAI provider owes: Article 53

Suppose your model qualifies as general-purpose. What do you owe? Article 53 lists four duties, and here is a date worth pinning: they have applied since **2 August 2025**. The 2026 Digital Omnibus amendment, which moved the high-risk deadlines to December 2027 and August 2028, left this chapter untouched: these duties bind today. (What switches on 2 August 2026 is the formal enforcement layer: the AI Office's power to compel information, order evaluations, and impose the GPAI-specific fines of up to 15 million euros or 3 percent of worldwide turnover under Article 101. Until then, supervision has run as polite but pointed "compliance dialogues".)

The four duties:

Technical documentation for the regulator. Draw up and maintain documentation of the model, covering the training and testing process, the evaluation results, and the minimum content listed in Annex XI, ready to hand to the AI Office or national authorities on request (Article 53(1)(a)).

Documentation for the people building on you. Prepare and keep updated information for providers of AI systems who integrate your model, so they understand its capabilities and limitations well enough to meet their own obligations (Article 53(1)(b), minimum content in Annex XII). If you remember the high-risk chapter, this rhymes: inward documentation plus outward instructions. But notice how far down the rhyme reaches. In the systems world, documentation duties kick in only at high-risk. In the models world, *every* general-purpose model carries them, systemic risk or not. The bar sits one level lower here.

A copyright policy. Put in place a policy to comply with EU copyright law, including honouring rights-holders' machine-readable opt-outs from text-and-data mining (Article 53(1)(c)). Chapter 12 digs into the copyright story properly; for now, know the duty exists.

A public training-data summary. Publish a "sufficiently detailed summary" of the content used to train the model, following a template from the AI Office (Article 53(1)(d)). When I first read this, I raised an eyebrow: the large labs guard their training-data recipes jealously, and here is a legal duty to summarise them in public. The template is no longer hypothetical, either: the AI Office published the mandatory version, with an

explanatory notice, in December 2025. It requires, among other things, disclosure of the largest web domains in the training data (the top 10 percent of domains by volume; smaller companies get a lighter 5-percent-or-1,000-domains version). One transition softens the blow: models already on the market before 2 August 2025 have until 2 August 2027 to catch up on this documentation. New models comply from day one.

Beyond the four, Article 53 adds the expected housekeeping: cooperate with the Commission and national authorities, and everything you hand over (trade secrets included) is protected by the Act's confidentiality rules.

The open-source discount

Article 53(2) carves out a genuine concession. If you release your model under a free and open-source licence, and genuinely open (parameters, weights, architecture, and usage information all publicly available), then the two documentation duties, (a) and (b), do not apply to you. The logic is fair enough: the transparency those documents would provide is already sitting on the internet for anyone to inspect.

Note what the discount does *not* cover. The copyright policy and the public training-data summary still apply, because openness about weights is not openness about what you trained on. And the whole exemption evaporates the moment the model carries systemic risk. Open-sourcing a frontier model does not buy its provider out of the heavy tier. Given that some major labs release their most capable models openly, that exception is not a footnote; it is the clause doing the real work.

Classification as systemic risk: a presumption, not a verdict

Chapter V has its own internal escalation, and its shape will feel familiar. Just as an AI system can climb from "ordinary" to "high-risk", a GPAI model can climb from "ordinary GPAI" to **GPAI with systemic risk**: the tier reserved for the true frontier models, with heavier duties attached. The shape is familiar, but the ladders are not the same ladder, and here I want to be firm, because this is where I see people slip.

> WATCH OUT
>
> **High-risk** is a property of *AI systems*: it lives in Chapter III and turns on what the system is used for (recruitment, credit, border control). **Systemic risk** is a property of *models*: it lives in Chapter V and turns on how capable the raw engine is, regardless of use. A frontier model with systemic risk is not "high-risk", and calling it that is not harmless shorthand; it points you at the wrong set of articles. Systems can be high-risk; models can carry systemic risk.

So how does a model climb this second ladder? Article 51 gives two routes.

The first is capability-based: the model has "high-impact capabilities", assessed with technical tools and benchmarks. And because "high-impact" is hard to measure directly, Article 51(2) supplies a proxy: a model is **presumed** to have high-impact capabilities when its training used more than 10^{25} floating-point operations of compute. You do not need to feel the number; just know it is an enormous amount of computation, the kind only a handful of frontier training runs have plausibly reached. For calibration: the Commission's GPAI guidelines treat roughly 10^{23} FLOPs as the indicative scale for merely *being* a general-purpose model. The systemic-risk line sits a hundred times higher. Most commercial models, including many well-known ones, live comfortably below it; this tier was built for the frontier, not the mainstream.

The word doing the legal work in that sentence is *presumed*. Crossing 10^{25} does not stamp your model "systemic risk"; it shifts the burden onto you. Under Article 52(2), a provider can submit substantiated arguments that, despite the compute, the model does not actually present systemic risks, and the Commission accepts or rejects them (Article 52(3)). A rebuttable presumption, in lawyer's terms. This is the single most misreported fact about Chapter V, so plainly: 10^{25} FLOPs is where the conversation starts, not where it ends.

The second route is designation: the Commission can classify a model as systemic-risk on its own initiative, or after an alert from its scientific panel, using the criteria in Annex XIII: number of parameters, dataset size, training compute, input and output modalities, benchmark results, even the number of registered users (Article 51(1)(b),

Article 52(4)). So even a model that never crossed the compute line can be pulled up the ladder if its real-world capabilities and reach warrant it. A designated provider can request reassessment, at the earliest six months after the decision (Article 52(5)).

Which models have actually been classified? Here is the honest answer, and it may surprise you: **nobody outside the process knows for certain.** Article 52(6) obliges the Commission to publish a list of systemic-risk models, but as of July 2026 no such public register exists. Estimates about which frontier models exceed 10^{25} FLOPs circulate constantly, but training compute is rarely disclosed, so they are exactly that: estimates. I will not name names, and I would treat anyone who confidently does with suspicion.

> MY TAKE
>
> What a frontier-scale provider should assume is the conservative reading: if your training run is anywhere near the line, count on qualifying. I would rather count with the worse possibility, at least for now.

Two weeks to raise your hand

There is one procedural duty here with real teeth, easy to miss because it hides in Article 52(1): a provider whose model meets the compute condition must **notify the Commission within two weeks**. Two weeks from the moment the threshold is met, or from the moment it becomes known that it *will* be met. Since training runs are planned and budgeted long in advance, that second clause means the clock can start before training even finishes. The notification can carry the rebuttal arguments we just discussed. If a provider stays quiet, the Commission can simply designate the model anyway when it finds out. This duty, like the rest of the chapter, has been live since August 2025.

Extra duties at the top: Article 55

A model classified with systemic risk keeps all the Article 53 duties and adds four more (Article 55(1)):

- **Model evaluation with adversarial testing.** Evaluate the model against state-of-the-art protocols, including structured attempts to make it misbehave (red-teaming, in industry language). If chapter 1's poisoned-training-data story taught you that a model can carry hidden failure modes its maker never intended, this is the duty that says: go hunting for them before release, and document the hunt.
- **Assess and mitigate systemic risks at Union level**: risks stemming from the model's development, release, or use, tracked to their sources.
- **Serious-incident reporting.** Track, document, and report serious incidents and the corrective measures taken, without undue delay, to the AI Office and national authorities. This too has grown practical scaffolding: the Commission published the incident-reporting template in November 2025.
- **Cybersecurity**, for both the model and the physical infrastructure it runs on. Frontier model weights are among the most valuable files on earth; the Act insists they be defended accordingly.

Notice the register these duties are written in: evaluate, mitigate, report, secure. Not "do not release": Chapter V never prohibits a model. It assumes frontier models will exist and demands their makers manage them like the critical infrastructure they are becoming.

The Code of Practice, and who signed it

How does a provider actually demonstrate all this? Article 53(4) and Article 55(2) sketch a hierarchy of proof, and getting it right matters more than it looks.

At the top sit **harmonised standards**: comply with one, and you earn a formal *presumption of conformity* for whatever the standard covers. One level down sit **codes of practice** under Article 56: adhering to one lets you *demonstrate compliance* until harmonised standards arrive. That is a genuinely useful legal shield, but a weaker one. A code shows you are compliant; a standard makes the regulator presume it. And a provider that follows neither must prove "alternative adequate means of compliance" to the Commission directly, which is the uncomfortable seat.

> AS OF MID-2026
>
> There are **zero** AI Act harmonised standards cited in the Official Journal, so the presumption-of-conformity route exists only on paper and the code of practice is the only game in town. (And no, an ISO/IEC 42001 certificate does not stand in for it: helpful for your management system, silent on AI Act conformity.) Signatory lists move too; the Commission keeps the current one on its digital-strategy site. Re-verify both before relying on this snapshot.

That code exists, and its short history is a better lesson in how EU tech regulation actually works than any article text. The **GPAI Code of Practice** was published on 10 July 2025, in three chapters (Transparency, Copyright, Safety and Security) plus a Model Documentation Form, and endorsed as adequate by the Commission around the time the obligations went live in August 2025. Then came the interesting part: who would sign a voluntary instrument? Roughly two dozen providers did, including most of the biggest names: Amazon, Anthropic, Google, IBM, Microsoft, Mistral, OpenAI. **Meta publicly refused**, arguing the code overreached the Act itself. **xAI signed exactly one chapter**, Safety and Security, and left the other two on the table, an à-la-carte option the code permits for its safety chapter. The major Chinese labs have largely stayed away.

The pattern is worth remembering: even a “voluntary” code becomes a strategic decision when the alternative is proving compliance to the regulator the hard way. Meta has not escaped Article 53 by refusing to sign; it has only chosen the uncomfortable seat.

Check yourself

1. Your company builds a medical-advice chatbot on top of a large model licensed from a major AI lab. Under Chapter V, which label fits what? A) Your chatbot is a GPAI model; the lab’s model is a GPAI system B) The lab’s model is a GPAI model; your chatbot is an AI system built on it C) Both are GPAI models with systemic risk D) Neither is covered by the AI Act until 2027

2. Which Article 53 duty still applies to a genuinely free and open-source GPAI model (weights, architecture and usage information all public), assuming no systemic risk? A) Technical documentation for the AI Office under Article 53(1)(a) B) Documentation for downstream system providers under Article 53(1)(b) C) The public training-data summary under Article 53(1)(d) D) None, because open-source models are fully exempt from Chapter V

3. A provider's new training run crosses 10^{25} FLOPs. What is the legal consequence? A) The model is automatically and finally classified as systemic-risk, and appears on a public register B) The model is presumed to have high-impact capabilities; the provider must notify the Commission within two weeks and may submit arguments rebutting the classification C) Nothing, unless the Commission also designates the model under Annex XIII D) The model becomes high-risk under Chapter III

4. Which statement about the GPAI Code of Practice is correct as of mid-2026? A) Adhering to it grants a formal presumption of conformity B) It was never finished, so the Commission imposed common rules by implementing act C) It lets signatories demonstrate compliance until harmonised standards are published, and roughly two dozen providers have signed, though not all of them D) Signing it is mandatory for every GPAI provider operating in the EU

5. Which duty belongs to providers of GPAI models with systemic risk, on top of the ordinary Article 53 duties? A) Registering the model in the EU database for high-risk systems B) Conducting a conformity assessment with a notified body C) Adversarial testing, systemic-risk mitigation, serious-incident reporting and cybersecurity under Article 55 D) Appointing a human overseer for every downstream application

Answers. 1: B. GPT-style engines are GPAI *models*; whatever gets built on top (by the lab or by you) is a *system*, governed by the rest of the Act. 2: C. The open-source exemption in Article 53(2) lifts only the two documentation duties (a) and (b); the copyright policy and the public training-data summary remain, and the whole exemption disappears for systemic-risk models. 3: B. Article 51(2) creates a rebuttable

presumption, and Articles 52(1) and 52(2) set the two-week notification with the option to argue against classification; no public register of designated models exists as of July 2026. 4: C. Codes of practice demonstrate compliance (Articles 53(4) and 55(2)); only harmonised standards grant the presumption of conformity, and none have been cited in the Official Journal yet; Meta refused to sign and xAI signed only the safety chapter. 5: C. Article 55 adds evaluation with adversarial testing, Union-level risk mitigation, incident reporting and cybersecurity; database registration and notified-body conformity assessment belong to the high-risk *systems* regime, a different ladder.

Next: who actually enforces all of this, from the AI Office and the national authorities to the fines and the honest state of the timeline.

CHAPTER 11

Governance, Penalties, Timeline

For ten chapters now, someone has been telling you what you *must* do. Document your bank's loan model. Disclose your chatbot. Take the mood camera off the receptionist. A fair question has probably been building the whole time: says who, exactly? If your company gets it wrong, who knocks on the door: a Brussels inspector, your national telecoms office, some new AI agency? And has anyone, anywhere, actually been fined yet?

This chapter answers those questions honestly, which means the answers will be less dramatic than the headlines. We start with the Act's support arm, the regulatory sandboxes: widely misunderstood, so we take them slowly, and I will tell you frankly what I think of them. Then we meet the institutions that actually run the Act, look at the fine tiers, and compare the law on paper with the enforcement reality of mid-2026. We close with the one table worth printing out, the full application calendar.

Sandboxes: support, not a gate

Imagine a fintech startup, twelve people, building a creditworthiness model, squarely high-risk under Annex III. They have read the requirements chapter and are nervous: risk management, data governance, technical documentation, all new to them. What they would love is someone from the regulator's side to look over their shoulder *before* the stakes are real and say "yes, that's roughly what we meant".

That is what an AI regulatory sandbox is for. The concept is borrowed from fintech, where regulators have run supervised testing environments for years. Article 57(5) describes a sandbox as a controlled environment that fosters innovation and lets you develop, train, test and validate an innovative AI system for a limited time, under a specific sandbox plan agreed between you and the competent authority (we will meet the full cast of authorities in a moment). You get guidance on regulatory expectations, supervision, help identifying risks. SMEs and startups get access free of charge (Article 58), and authorities must decide on your application within three months.

Now the part people get wrong, so let me be blunt about it.

> **WATCH OUT**
>
> A sandbox is not an approval gate. You do not need its permission to put a high-risk system on the market: access runs through the Article 43 conformity assessment, exactly as described earlier in this book, whether or not you ever went near a sandbox. Participation is voluntary innovation support, aimed especially at smaller players without a regulatory-affairs department.

What do you get out of it, then? Article 57(7) spells it out:

> **THE ACT SAYS · ARTICLE 57(7)**
>
> "...the exit reports and the written proof provided by the national competent authority shall be taken positively into account by market surveillance authorities and notified bodies, with a view to accelerating conformity assessment procedures to a reasonable extent."

In plain terms: the sandbox gives you a written record that a competent authority watched you work and saw you take the rules seriously, a record that speeds up your real conformity assessment but does not replace it. One more genuinely valuable perk hides in Article 57(12): follow the agreed sandbox plan and the authority's guidance in good faith, and no administrative fines are imposed for AI Act infringements during the experimentation. A safe space in the literal sense, though you remain liable under ordinary law for damage you cause to third parties.

Two clarifications, because these points get muddled even in professional circles. First, the **EU database of high-risk systems is not the sandboxes' job**. The Commission maintains that database (Article 71), and providers register their systems in it directly, before placing them on the market (Article 49). Sandboxes have no role in it. Second, **real-world testing outside a sandbox**, meaning testing your Annex III system with real people before conformity assessment, is a separate track under Article

60, approved by the market surveillance authority: you submit a testing plan, and silence for 30 days counts as tacit approval. Real-world testing can also happen *inside* a sandbox, supervised there, but then the terms sit in your sandbox plan and no 30-day clock applies.

Where do sandboxes stand today? Behind schedule. The original deadline for every Member State to have at least one operational sandbox was 2 August 2026; that was moved to 2 August 2027 by the 2026 Omnibus amendment, which also added an EU-level sandbox run at the AI Office and extended real-world testing to Annex I systems. If your country's sandbox does not exist yet, that is not a compliance problem, because participation was never mandatory. But the support arm of the Act is arriving later than the obligations arm. Plan accordingly.

My thoughts on sandboxes

> MY TAKE
>
> I will admit I saw the delay coming. The talent pool of people who understand both machine learning and EU product law is thin, and you cannot staff a supervisory sandbox with juniors. Asking 27 countries to hire such people on a two-year timeline was always optimistic.

Two other things worry me once the sandboxes do open.

The first is workload. You have seen how broad the definition of an AI system is and how wide the Annex III categories sweep. The banks and insurers I train are already mapping their use cases; the day a well-run sandbox opens in their country, a first wave of applications will hit it, and a five-person office facing that wave means queues. The three-month decision deadline helps only if the office behind it has capacity. And there is a learning curve on both sides: companies must learn what the authority expects, and the authority must learn to apply a genuinely new law to genuinely varied systems. Neither side has decades of settled practice to lean on.

The second is culture. A sandbox, and real-world testing generally, assumes you work like a scientist: form a hypothesis, design an experiment, test on a limited scale, measure. Medicine works this way, because a new drug has no historical data and a controlled trial is the only honest evidence. Most European companies, in my experience, do not. They build from historical data and observation; structured experimentation is a habit they never formed. Many will discover that the hard part of the sandbox is not the paperwork but the discipline it presumes.

Take all of this as one consultant's read, not a prophecy. The honest summary: sandboxes are a good idea arriving slowly, and the companies that benefit most will treat them as a learning channel rather than a queue to stand in.

Who actually runs this thing

A sandbox supports you; it does not police you. So who does? The AI Act is an EU regulation, but the EU did not build a single AI police force. Enforcement is split between two levels, and knowing which handles what saves you from emailing the wrong institution.

At the national level: your competent authorities. Article 70 requires every Member State to designate at least one *notifying authority* (which accredits the notified bodies from the conformity assessment chapter) and at least one *market surveillance authority* (which checks systems on the market and investigates when something looks wrong). The market surveillance authority is the one that matters for your day-to-day risk. If a competitor reports your emotion-recognition dashboard, or a journalist writes about your scoring system, this is the body that opens the file, including for the Article 5 prohibitions. Each country names one of these authorities as its single point of contact.

At the EU level: the AI Office. Housed inside the European Commission, the AI Office (Article 64) is the closest thing the Act has to a central brain. It supervises general-purpose AI models directly; the GPAI obligations from the previous chapter are its territory, not your national regulator's. And it coordinates everything else: guidance, codes of practice, support for the national authorities.

Around them, three supporting bodies. The European Artificial Intelligence Board (Articles 65 and 66), with one representative per Member State, exists to keep 27 national enforcement practices from drifting apart. A scientific panel of independent experts (Article 68) advises on GPAI matters, including alerting the AI Office to systemic risks; it was constituted in 2025 with sixty experts. And an advisory forum (Article 67) gives industry, SMEs, academia and civil society a formal voice. You will rarely deal with any of these directly, but when the Commission publishes guidelines, this machinery is where they were argued over.

And one thing you should actually go and try. Since October 2025 the Commission runs the **AI Act Service Desk** with a single information platform: a compliance checker, an interactive Act explorer, and a channel for asking questions. It is at ai-act-service-desk.ec.europa.eu, it is free, and it answers the "who do I even ask?" question the Act's early years lacked. If you take one action after this chapter, spend fifteen minutes there with one of your own use cases. The answers are non-binding, but they are the closest thing to an official opinion short of hiring a law firm.

The price list

Now the numbers everyone quotes at conferences. Article 99 sets three tiers of administrative fines, and in each case the ceiling is the fixed amount *or* the percentage of total worldwide annual turnover, whichever is **higher**:

- **35 million euros or 7 percent** for violating the Article 5 prohibitions. Deploy that mood camera on the receptionist, and this is the tier you are in.
- **15 million euros or 3 percent** for breaching the main operational duties: provider obligations under Article 16, deployer obligations under Article 26, importer and distributor duties, notified-body rules, and the Article 50 transparency duties. This is where most realistic corporate failures would land; think of the bank that runs its high-risk loan model without the required documentation and oversight.
- **7.5 million euros or 1 percent** for supplying incorrect, incomplete or misleading information to authorities or notified bodies.

Three softeners are worth knowing. For SMEs and startups, each ceiling flips to whichever is *lower* of the amount and the percentage (Article 99(6)), a genuine mercy for a twelve-person fintech. Fines must be effective, proportionate and dissuasive, and Article 99(7) lists the circumstances that shape the actual amount: gravity and duration, cooperation, self-reporting, intent versus negligence. And these are *ceilings*, not price tags; nothing in the Act says a first offence goes anywhere near the maximum.

Two special regimes sit alongside. When EU institutions themselves break the rules, the European Data Protection Supervisor fines them, at much lower ceilings (up to 1.5 million euros for prohibited practices, Article 100). And for providers of general-purpose AI models, fines come from the Commission itself, not national authorities: up to 15 million euros or 3 percent under Article 101. One timing subtlety: the penalties chapter has applied since 2 August 2025, but the Commission's formal power to fine GPAI providers under Article 101 only bites from 2 August 2026.

Who actually imposes the ordinary fines? The Member States. Article 99(1) requires each country to lay down its own penalty rules within the Act's ceilings, and that detail leads directly to the least discussed fact about the AI Act in 2026.

The scoreboard, honestly

Here is where I would rather give you the uncomfortable truth than the impressive slide.

> AS OF MID-2026
>
> With prohibitions enforceable since February 2025 and the full penalties chapter live since August 2025, there is no confirmed AI Act fine anywhere in the European Union. Not one. No confirmed formal proceeding either. Re-verify before you repeat this in a boardroom; it can change with a single decision.

Why? Look at the plumbing. Member States were supposed to designate their competent authorities by 2 August 2025; by spring 2026, only around 8 of the 27 had actually done so. A fine needs an authority to impose it and a national penalty law to

base it on, and most countries have neither fully in place. Italy got there first and, so far, most completely: Law 132/2025, in force since October 2025, names its authorities and is the only complete national penalty schedule in force. Slovakia has so far passed only Act No. 318/2025 Z.z., a narrow amendment to its conformity-assessment law effective January 2026; the fuller Slovak bill, which would make MIRRI the market surveillance authority, is still at proposal stage. Czechia is at the draft stage too: its adaptation law would hand market surveillance to the ČTÚ, but it has not passed. If you work in Bratislava or Prague, your AI regulator is, quite literally, still being formed.

The closest thing to a test case is Hungary, where civil-society organisations argue that expanded facial-recognition surveillance violates the Article 5 ban on real-time remote biometric identification and have pressed the Commission to act, but no infringement procedure has been confirmed. And the fines you may have seen in the news attached to AI companies (the Italian data-protection authority's 5 million euro fine against the maker of the Replika chatbot, the accumulated fines against Clearview AI) are **GDPR** fines. Different law, different authorities. Keep the regimes straight, because your colleagues will not.

So should you relax? I would not. Every enforcement regime in EU history, GDPR included, started with exactly this quiet phase, and GDPR's quiet phase ended with fines in the hundreds of millions. The authorities being designated now will inherit a backlog of complaints; the Hungarian file shows civil society is already generating them. My honest read: the enforcement risk is ahead of you, not behind you. That is not a reason to ignore the Act; it is a rare, closing window in which you can fix things calmly instead of under investigation.

The calendar

One more moving part before the table: the timeline was reshaped in 2026 by the Digital Omnibus amendment, agreed by the European Parliament on 16 June and the Council on 29 June 2026. As I write in July 2026 it still awaits publication in the Official Journal, so verify the final text before betting the company on it. Its headline effect was to defer the high-risk deadlines by roughly a year or more. One thing it did **not** do, despite persistent rumours: it did not defer the Article 50 transparency duties,

which apply from 2 August 2026 as originally planned. And the new dates are fixed calendar dates. An earlier proposal to make them conditional on harmonised standards being ready was rejected, so do not let anyone tell you the deadlines “only start when the standards arrive”. (There are still zero harmonised standards cited in the Official Journal, and no, an ISO/IEC 42001 certificate is not AI Act compliance; but the clock runs regardless.)

Here is the full picture as it stands in July 2026:

DATE	WHAT APPLIES
1 August 2024	The AI Act enters into force (Regulation (EU) 2024/1689)
2 February 2025	Prohibited practices (Article 5); AI literacy duty (Article 4)
2 August 2025	GPAI model obligations; governance bodies; penalties (Chapter XII); national authority designation deadline
2 August 2026	Article 50 transparency duties (not deferred); Commission’s formal GPAI fining powers (Article 101)
2 December 2026	New Article 5 prohibition on AI-generated non-consensual intimate imagery and CSAM (added by the Omnibus); end of the marking grace period for generative systems already on the market
2 August 2027	National sandbox obligation (moved from 2026 by the Omnibus); EU-level sandbox at the AI Office
2 December 2027	High-risk obligations for Annex III systems (moved from August 2026 by the Omnibus)
2 August 2028	High-risk obligations for Annex I systems, meaning safety components of regulated products (moved from 2027 by the Omnibus)

Read that table from your own seat. If you deploy a chatbot, August 2026 is your date. If you are the bank with the loan model, December 2027 is when the full high-risk machinery becomes enforceable against you, which sounds far away and is not, given

what the requirements chapters asked of you. And if your AI lives inside a machine or a medical device, you have until August 2028, with the sector rules you already follow carrying you in the meantime.

Check yourself

1. Someone files a complaint that your company deployed an emotion-recognition system on employees. Which body investigates, and under which fine tier does the violation fall? A) The AI Office; up to 15 million euros or 3 percent of turnover. B) The national market surveillance authority; up to 35 million euros or 7 percent of turnover. C) The national sandbox; up to 7.5 million euros or 1 percent of turnover. D) The European AI Board; the tier depends on the Member State.

2. A startup building a high-risk Annex III system asks whether it must pass through an AI regulatory sandbox before entering the EU market. The correct answer is: A) Yes. Sandbox approval is a precondition for market access for all high-risk AI. B) Yes, but only until the EU-level sandbox opens. C) No. Sandboxes are voluntary support; market access runs through the Article 43 conformity assessment, and a sandbox exit report merely gets taken positively into account. D) No. Sandboxes are only open to deployers.

3. Who maintains the EU database of high-risk AI systems, and who enters a provider's system into it? A) The national sandboxes maintain it and enter systems after approval. B) The AI Office maintains it; national authorities enter systems. C) The Commission maintains it (Article 71); the provider registers its system directly (Article 49). D) Each Member State maintains its own database; notified bodies enter systems.

4. As of July 2026, which statement about AI Act enforcement is accurate? A) Several GPAI providers have been fined by the Commission under Article 101. B) The Italian fine against the Replika chatbot was the first AI Act fine. C) No confirmed AI Act fine exists anywhere, and only around 8 of 27 Member States had designated their competent authorities on time. D) Enforcement cannot begin until harmonised standards are published.

5. After the 2026 Omnibus amendment, when do the high-risk obligations for Annex III systems apply, and what is the nature of that date? A) 2 August 2026, unchanged from the original text. B) 2 December 2027, a fixed calendar date, not conditional on standards being available. C) 2 December 2027, but only once harmonised standards are cited in the Official Journal. D) 2 August 2028, the same date as Annex I systems.

Answers: 1: B. Workplace emotion inference violates Article 5(1)(f), a prohibited practice enforced by the national market surveillance authority at the top tier of Article 99(3). 2: C. Article 57(5) makes sandboxes voluntary innovation support, and Article 57(7) says exit reports are "taken positively into account" to accelerate, not replace, conformity assessment. 3: C. Article 71 puts the database in the Commission's hands and Article 49 makes registration the provider's own duty; sandboxes play no role. 4: C. The honest scoreboard shows zero confirmed AI Act fines and a national-designation lag, and the Replika and Clearview fines were GDPR cases. 5: B. The Omnibus moved Annex III to 2 December 2027 as a fixed date after legislators rejected making the deadlines conditional on standards.

You now know the rules, the referees and the calendar. The final chapter puts it all together and asks what a sensible organisation actually does on Monday morning.

CHAPTER 12

The AI Act in Practice

Remember the Monday morning from chapter 3? Your CEO forwarded you an article about the AI Act with a one-line note: “Are we affected?” Nine chapters later, you know the honest answer is “almost certainly, somewhere.” So the note changes: “We should probably do something about this. Can you take it?” You now own AI Act compliance for your company. Where do you actually start?

This chapter is my answer to that question: the advice I give when a company asks me “we understood we need to comply, what do we do first?” Then, as a farewell bonus, we will settle a question I get asked constantly: if you write a book or design a campaign with AI, does the Act make you label it, and do you even own it? The answer to both parts is more interesting than you might expect.

Everything here builds on the previous chapters. Nothing here is new law; it is the same Act, seen from the side of the person who has to do something about it by Thursday.

Step one: check yourself against the prohibitions

Not the high-risk rules. Not the documentation templates. The prohibitions.

Here is why. The Article 5 bans have applied since 2 February 2025 (they are not coming, they arrived), and they sit in the top fine tier, up to 35 million euros or 7 percent of worldwide turnover, with the penalty regime itself live since August 2025. It remains true, as of July 2026, that nobody anywhere has been fined under the Act yet. But if there is one place I would not want to be first, it is here.

There is also a genuinely near deadline to know about: the new prohibition on AI-generated non-consensual intimate imagery and CSAM, the “nudifier” ban added to Article 5 by the 2026 Omnibus amendment, becomes applicable on 2 December 2026. And the big high-risk wave for Annex III use cases lands on 2 December 2027. That is your real urgency calendar: December 2026 for the new ban, December 2027 for high-risk. Not a panic, but not infinite runway either.

So the exercise is simple to describe and uncomfortable to do: walk through your productive AI use cases and ask whether any of them brushes against the eight (soon nine) prohibitions. Remember the receptionist under the mood camera: emotion inference in the workplace is banned, and it is exactly the kind of "employee wellbeing analytics" feature that arrives quietly inside an HR platform. If you find a suspect use case, take it seriously early. Shutting down a productive, money-making system is organisationally painful and takes far longer than anyone expects. Better to discover that in a calm quarter than in a regulator's letter.

While you are at it, run an information session for your data and AI people. The Act's AI-literacy duty (Article 4) has applied since the same February 2025 date. The Omnibus softened its wording, but taking measures to build literacy is still expected, and frankly it is the cheapest compliance measure you will ever buy.

Step two: map your roles and mind the trap doors

Once the prohibitions are cleared, go around the organisation and inventory everything that might fall under the definition of an AI system. For each one, answer the question that drives most of your obligations: are we the provider here, or the deployer?

The intuitive test you know from earlier chapters still does most of the work. Did we develop this system? Then we are likely the provider, and the Act regulates us heavily. Did we buy or rent it from somebody else? Then we are likely the deployer, with a lighter set of duties.

But I want you to carry two extra routes into that inventory, because they are exactly where companies misclassify themselves.

First, you do not need to train anything to be a provider. Article 3(3) says you are a provider also when you *have a system developed* by someone else and place it on the market under your own name or trademark. If an agency builds your customer-facing chatbot and it ships as "YourCompany Assistant," the fact that you never wrote a line of code does not make you a deployer. The badge on the product decides.

Second, a deployer can *become* the provider after the fact. For high-risk systems, Article 25(1) lists three trap doors:

> THE ACT SAYS · ARTICLE 25(1)
>
> Any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system… if they put their name or trademark on a high-risk AI system already placed on the market; make a substantial modification to it; or modify the intended purpose of an AI system… so that it becomes high-risk.

Cross any of these and you inherit the full provider obligations of Article 16; under Article 25(2), the original provider largely steps out of the picture for that system. So the company that buys an HR screening tool, rebrands it, and heavily customises the model is not "just a deployer" no matter what the invoice says.

The practical takeaway: for every system in your inventory, record not only "provider or deployer" but also "under whose name does it run, and have we modified it or repurposed it?" Trust me: your future self will thank your past self for this mapping.

Step three: document as you go

The Act mentions documentation at almost every turn: technical documentation, assessments, evaluations, logs. But my advice goes beyond the formal requirements: keep at least an informal record of your compliance journey itself. Which systems you found, what you decided about each, why you classified something as out of scope, when you ran that awareness session.

If you ever need to show a regulator (or your own board) that you acted diligently, "we did it" without a paper trail is worth very little. A dated folder of imperfect notes beats a perfect memory every time.

> WATCH OUT
>
> An ISO/IEC 42001 certificate is not AI Act compliance. Useful as a management backbone, but as of July 2026 no harmonised standard has been cited in the Official Journal, so no certificate buys a presumption of conformity.

Step four: follow your regulator, and use the Service Desk

Here I have to be honest about the state of the world: for many of you, "your national regulator" does not fully exist yet.

> AS OF MID-2026
>
> Only around eight of the twenty-seven Member States have designated their national authorities. In Slovakia and Czechia, where many of my students sit, the implementing laws are still drafts.

That is not a reason to relax; it is a reason to watch, because when your authority does stand up, you want to be among the first who know its name.

Two concrete things you can do today. First, bookmark the Commission's AI Act Service Desk (ai-act-service-desk.ec.europa.eu), live since October 2025. It includes a Compliance Checker and an AI Act Explorer, and it is the closest thing to an official help line the Act currently has. Second, keep the regulatory sandboxes on your radar. The obligation for every Member State to run at least one was moved to August 2027 by the 2026 Omnibus amendment, and an EU-level sandbox at the AI Office was added alongside it.

And here is a hint you are free to ignore: if your organisation is heavily exposed, with many use cases likely to be high-risk, consider preparing one or two *toy* use cases now. Not systems you necessarily plan to deploy; drafted, plausible proposals, perhaps one low-risk and one high-risk. When a sandbox opens near you, you then have a ready-made, legitimate reason to apply and learn the process from the inside. Both sides win: you get context and a relationship, the sandbox gets an early guinea pig. I would rather

be the company the regulator already knows than the one it meets for the first time during an inspection.

Step five: several people, not one

A pattern I have watched more than once: a company appoints exactly one person to "own" AI compliance. That person spends months becoming genuinely knowledgeable. And then a recruiter finds them, because trained AI-compliance people are scarce and getting scarcer, and your entire institutional knowledge walks out the door with a 30 percent raise.

So: dedicate multiple people, drawn from both the builder side and the auditor side of your organisation. Redundancy here is not bureaucratic padding; it is how you make the knowledge survive contact with the job market.

> MY TAKE
>
> Do not get defensive. "We will avoid the AI Act by not using AI" is a strategy I have actually heard, and it is a way of paying the compliance cost anyway: in lost competitive advantage instead of lawyer hours. The Act is a set of guardrails, not a wall.

Bonus: AI, copyright, and who owns your book

Now for the question behind half the emails I receive. You are an author using a text model to help write your book, or a marketer generating a poster for a campaign. Two worries, usually tangled together: do I have to label this as AI-made, and do I even own it?

Let's untangle them, because the Act answers the first and deliberately stays out of the second.

Do you have to label AI-assisted work?

Here is the part that surprises people: there is no general duty in the Act to label work you made with AI assistance. The Article 50 transparency duties, which apply from 2 August 2026 and were not deferred by the Omnibus, are specific, and they mostly do not land on you as an author or marketer.

The duty to mark synthetic content in a machine-readable way sits on the *provider* of the generative system, not on you as its user (Article 50(2)). And even that duty has a carve-out worth quoting, because it draws exactly the line people worry about.

> THE ACT SAYS · ARTICLE 50(2)
>
> This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof…

In plain terms: the proofreader end of the spectrum is explicitly out. If the AI polishes your grammar and does not substantially change what you wrote or what it means, no marking duty arises at all. That also answers the "does every Photoshop touch-up make my image AI-generated?" worry.

As a deployer, you owe a disclosure in exactly two situations (Article 50(4)). One: deep fakes, content depicting real people, places or events in a way that would falsely appear authentic, must be disclosed, with a softened, non-intrusive form of disclosure for evidently artistic, creative or satirical works. Two: AI-generated text published to inform the public on matters of public interest (think news) must be disclosed, *unless* the text has gone through human review or editorial control and a person holds editorial responsibility for it. That last carve-out is essentially the newsroom exception: an editor who stands behind the piece replaces the label.

Run your two scenarios through that. An AI-assisted novel is neither a deep fake nor public-interest information, so no deployer duty. An ordinary marketing visual, no real people faked, no deployer duty either. The labels you increasingly see on platforms come from the providers' marking obligation and platform policies, not from a duty on you.

What the Act actually says about copyright

People sometimes expect the AI Act to be a copyright law. It is not. But it does regulate copyright at one specific point in the pipeline: the *input* side, where models are trained on other people's works.

Every provider of a general-purpose AI model must put in place a policy to comply with EU copyright law, and in particular to identify and honour the machine-readable "opt-outs" that rightsholders can declare under Article 4(3) of the 2019 Copyright Directive, the text-and-data-mining rights reservations (Article 53(1)(c)). On top of that, every GPAI provider must publish a sufficiently detailed summary of the content used for training, following the AI Office's template, which has been a mandatory template since December 2025 (Article 53(1)(d)). The GPAI Code of Practice even has a dedicated Copyright chapter walking providers through it. All of this has been in force since August 2025, and the Omnibus did not touch it.

So if your novel was scraped into a training set against your declared opt-out, the AI Act is very much your law. If you are wondering who owns the poster the model generated for you, it is not.

So who owns the output?

For ownership we must leave the AI Act entirely and reach for copyright law, which in the EU means, as a rule of thumb, the national copyright act of the country where you create. I went through this personally: I was among the first people in Slovakia to properly publish a book written with generative models, both text and image. So I sat down with my lawyer, and we drew three scenarios.

Scenario one: the AI is a *brush*. A brush is a static tool. Perform the identical hand movement twice and you paint the identical stroke twice. No uniqueness comes from the tool itself.

Scenario two: the AI is a *dog holding a brush*. Tell a dog to paint a circle twice and you get two different circles; the imperfection of its muscle control adds something. There is

uniqueness now, but the dog is still your instrument; you trained it and told it what to paint.

Scenario three: the AI is a *creative entity* in its own right, in which case it, not you, would hold rights in what you make together.

Which is it? The Slovak Copyright Act (and copyright laws are strongly harmonised across the EU, so yours will read similarly) defines the subject of copyright as a work of literature, art or science "which is the unique result of the creative mental activity of the author perceptible to the senses." The load-bearing words are *unique result* and *creative mental activity*.

Uniqueness first. Ask a text model the same question in two fresh conversations and you get two different answers, because developers deliberately build randomness into generation: the model samples among the most likely next words rather than always picking the top one. So the brush scenario falls away: this tool does not repeat itself.

That leaves the dog and the creative entity. And here a principle older than any AI regulation decides the matter: in copyright law as it stands, internationally, only humans are considered capable of creative activity. Machines do not author. Which means today's generative AI lands squarely in scenario two: a very talented dog. It produces unique outputs, but the creative mental activity, and therefore the copyright, belongs to the human who directed it. Legally, your text model sits on the same shelf as Microsoft Word: a tool, an unusually lively one.

The usual disclaimer applies with extra force here: this is the current settlement, national laws differ in their details, and courts are still testing the edges. How much human direction is enough is exactly the kind of question that will produce interesting judgments in the coming years. If real money rides on your answer, this is a conversation to have with your own lawyer, in your own country. But as a working model, the dog with the brush has served me (and my book) well.

Check yourself

1. Your CEO hands you AI Act compliance today. According to this chapter, your first move is: A) Start drafting technical documentation for every AI system B) Check your productive AI use cases against the Article 5 prohibitions, which have applied since February 2025 C) Wait for your national regulator to be designated D) Apply for ISO/IEC 42001 certification, which counts as AI Act compliance

2. A company buys a high-risk recruitment-screening system, rebrands it under its own trademark and substantially modifies the model. Under Article 25(1), the company: A) Remains a deployer, since it did not develop the system B) Becomes the provider and takes on the Article 16 provider obligations C) Shares provider obligations 50/50 with the original vendor D) Only needs to notify the original provider

3. Where does the AI Act regulate copyright? A) It assigns copyright in AI outputs to the model provider B) It contains a complete EU copyright regime for AI-generated works C) On the training-input side: GPAI providers need a copyright policy honouring TDM opt-outs and must publish a training-content summary D) Nowhere, because the Act never mentions copyright

4. A novelist uses an AI model as a proofreader that makes minimal edits, then publishes the book. Under Article 50: A) The book must carry an "AI-generated" label B) The novelist must register the book in the EU database C) No duty arises, since assistive editing that doesn't substantially alter content is carved out, and a novel is neither a deep fake nor public-interest information D) The publisher must watermark every page machine-readably

5. Under current copyright law, who holds copyright in an image you generate with an AI tool? A) The AI model, since it produced the unique result B) Nobody, because AI-assisted works cannot be copyrighted at all C) The model's developer, through the terms of service D) You, the human who directed the tool, since only humans are considered capable of creative activity

Answers. 1: B. The prohibitions are the oldest live obligations, carry the top fine tier, and unwinding a banned use case takes time, so they come first. 2: B. Rebranding, substantial modification or repurposing into high-risk each make the downstream actor

the provider under Article 25(1), with the original provider stepping back under 25(2). 3: C. Articles 53(1)(c) and (d) regulate the training-input side, while ownership of outputs is left to national copyright law. 4: C. Article 50(2)'s marking duty sits on providers and excludes assistive standard editing, and the deployer disclosure duties in 50(4) cover only deep fakes and public-interest text. 5: D. Today's law treats generative AI as a tool (the dog with the brush): its outputs are unique, but creative activity, and thus copyright, is reserved to humans.

And that is the whole journey. We started with a definition that seemed to swallow every spreadsheet in Europe, and we end with you owning the book your very talented dog helped you write. Along the way you learned to spot a prohibition, classify a high-risk system, tell a provider from a deployer, and read a compliance deadline without panicking. That is genuinely more than most people advising on this regulation can do.

The Act will keep moving: dates shifted once already, guidelines keep arriving, and someday there will even be a first fine for the newspapers to write about. When that happens, don't take anyone's word for what it means, including mine. Open the text. You now know your way around it.

Thank you for reading. Now go map your systems, and maybe let the dog paint you something on the way.