



Co-funded by
the European Union



The AI Data Robotics
Association

AI Liability in the Era of Agentic Ecosystems: Legal, Ethical, and Operational Frontiers

A Legal & Ethical ADRA TG Webinar

07 July 2026

Speakers



Dr. Marina Cugurra

Ethical and Legal ADRA Topic Group Co-Chair



Argyro Amidi

Ethical and Legal ADRA Topic Group Co-Chair



Amaya Garmendia

ADRA Membership and Policy Manager



Sergiy Barbashyn

President, AI Ethics and Integrity International Association (AIEI)



Alexei Grinbaum

AIOLIA Coordinator, Senior research scientist at CEA-Saclay



Katrin Saaremael

Legal and Policy Officer, DG GROW (European Commission)

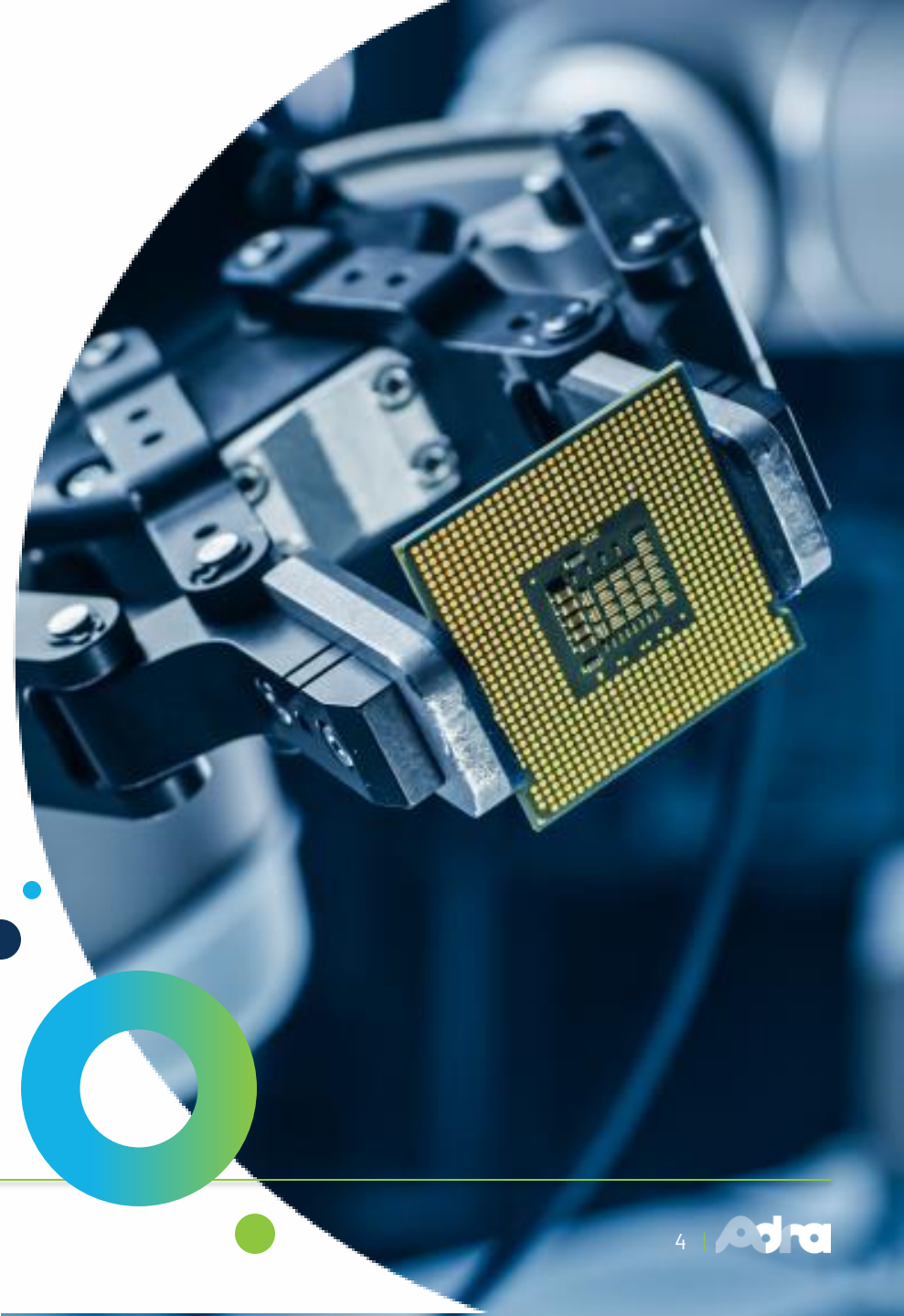
The Context

Artificial Intelligence is evolving from traditional software into increasingly autonomous, adaptive, and agentic systems operating within complex digital ecosystems

At the same time, Europe is entering a **new regulatory era** shaped by the AI Act, the Revised Product Liability Directive and emerging sectoral and governance requirements

3 main Industry Challenges:

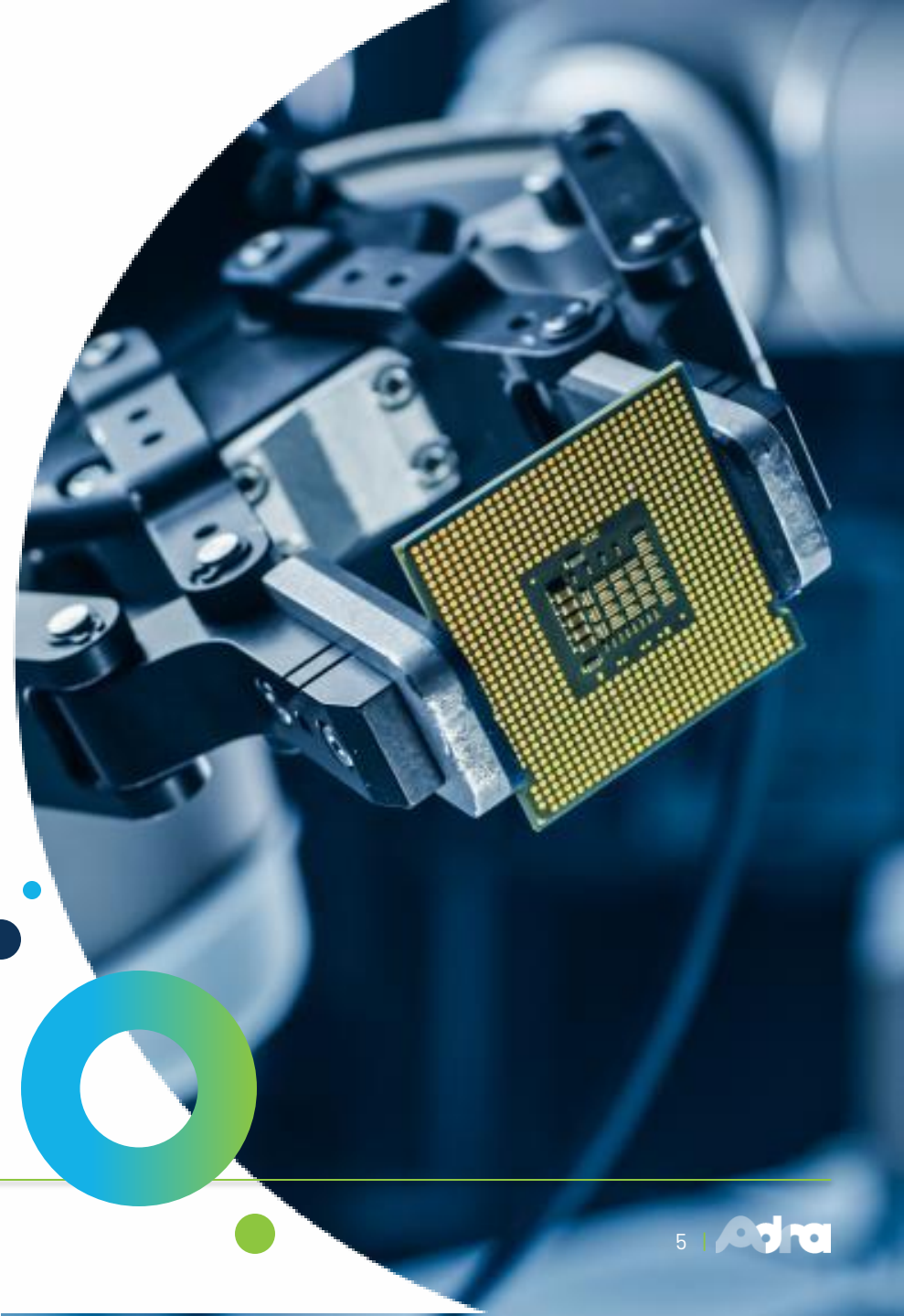
Unclear Liability Allocation Across the AI Value Chain
The Causality and Evidence Problem
From Legal Rules to Operational Practice



Objectives of the Webinar

The webinar is directed to address the **AI liability challenge** across **two interconnected fronts**:

- **Framework Evolution & Deep Dives:** Unpacking the **alleviated burdens of proof, state liability,** and the **distinct challenges** posed by autonomous, agentic AI systems that operate outside traditional human-in-the-loop oversight.
- **Operational Realities & Industry Cases:** Dissecting **practical failures** across **critical sectors**- such as healthcare diagnostics, autonomous mobility, and generative misrepresentation- to bridge the gap between regulatory expectations and engineering constraints.



Agenda

14:30 – 15:00 | Part 1: The Legal and Ethical Framework

- **Welcome & Objectives** (Marina Cugurra)
- **The Tripartite Legal Framework** (Marina Cugurra & Argyro Amidi & Katrin Saaremäel)
- **Procedural Shift:** Navigating the alleviated burden of proof under new PLD. (Katrin Saaremäel)
- **Soft Law & Ethical Grounding** (Marina Cugurra & Argyro Amidi)

15:00 – 15:15 | Part 2: Deep Dives

- **Autonomous Agents & Human Oversight:** Liability in Agentic AI and Multi-Agent Systems (Marina Cugurra)
- **Algorithmic Bias:** Liability for discriminatory outcomes in sensitive sectors (hiring, lending, housing). (Argyro Amidi)

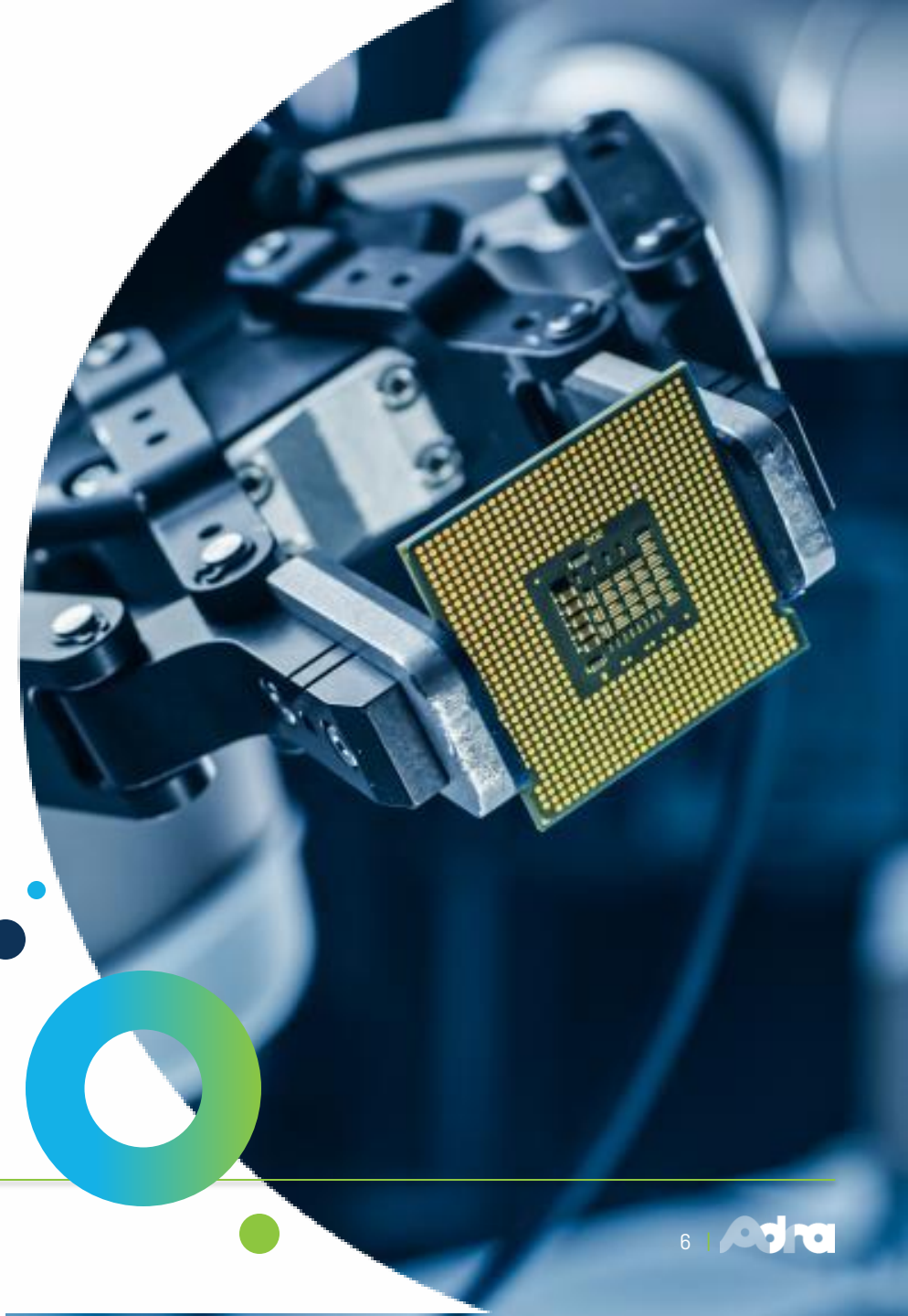
15:15 – 15:30 | Part 3: Examples and Sector-Specific Cases

- **Autonomous Vehicles:** Modern variations of the trolley problem. (Amaya Garmendia)
- **Healthcare Domain:** Diagnostic errors and wrongful recommendations in decision support systems. (Marina Cugurra)
- **Generative AI Misrepresentation:** High-stakes misinformation and toxic outputs (e.g., liability in AI-influenced self-harm cases). (Argyro Amidi)
- **Industrial Automation:** Physical harms and collaborative robots (Cobots). (Marina Cugurra)

15:30 – 15:50 | Expert Roundtable & Panel

- **Sergiy Barbashyn, Alexei Grinbaum, Katrin Saaremäel** - Moderator: Argyro Amidi

15:50 – 16:00 | Q&A and Wrap-up (Argyro Amidi)



Part 1.

The Legal and Ethical Framework

Contractual framework of AI Liability

AI challenges traditional liability because **AI systems are:**

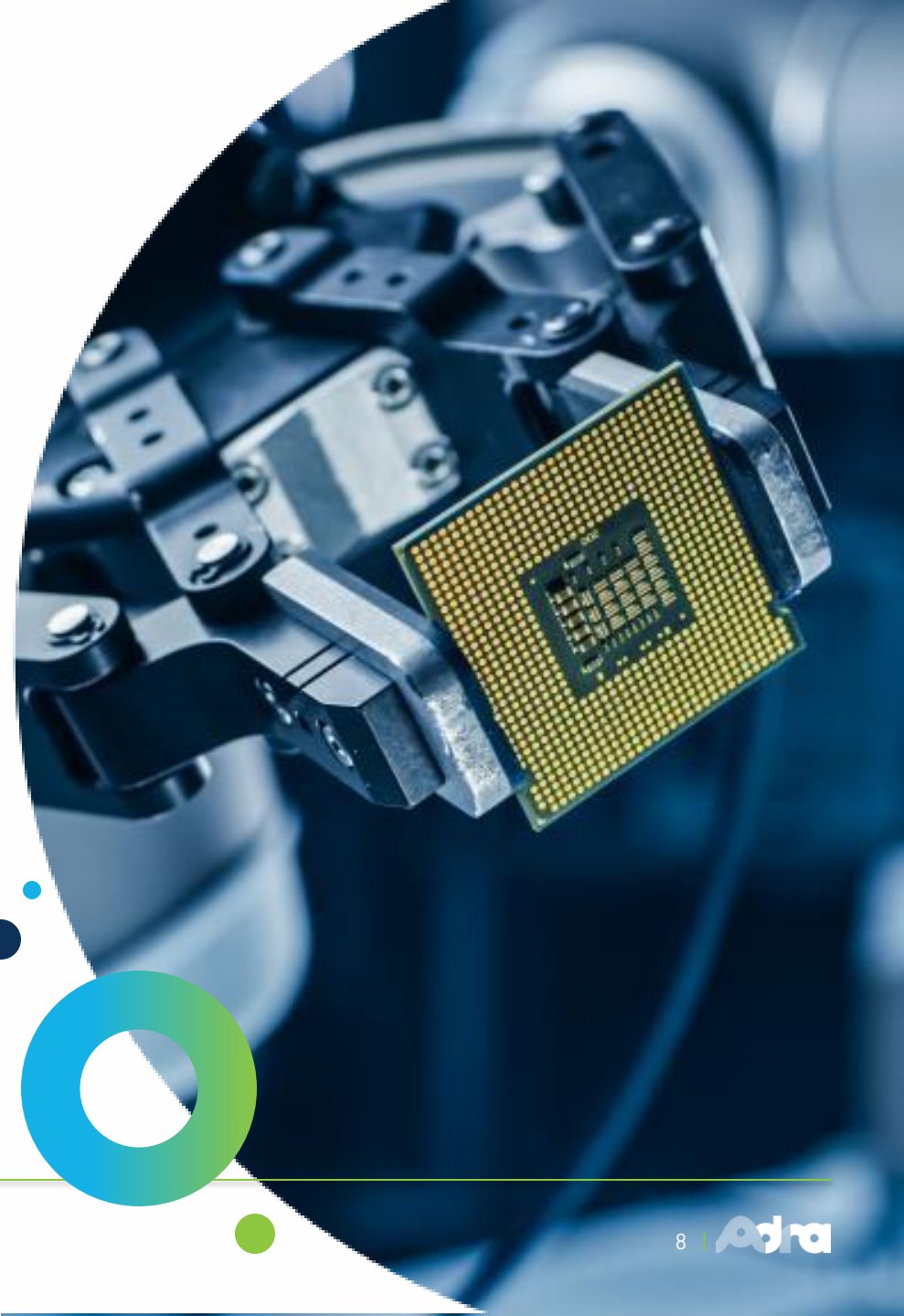
- opaque (“black box”),
- autonomous,
- unpredictable,
- continuously learning.

Traditional concepts such as:

- breach,
 - defect,
 - causation,
- become difficult to apply.

Liability is **fragmented because multiple actors** may be involved: manufacturer, software developer, deployer, integrator, user

Contractual governance of liability allocation between chain actors in multi-layered AI pipelines, plus **risk management tools** (clear indemnity, logging, and audit obligations).



Non-contractual (tort) Liability: key concepts

Definition

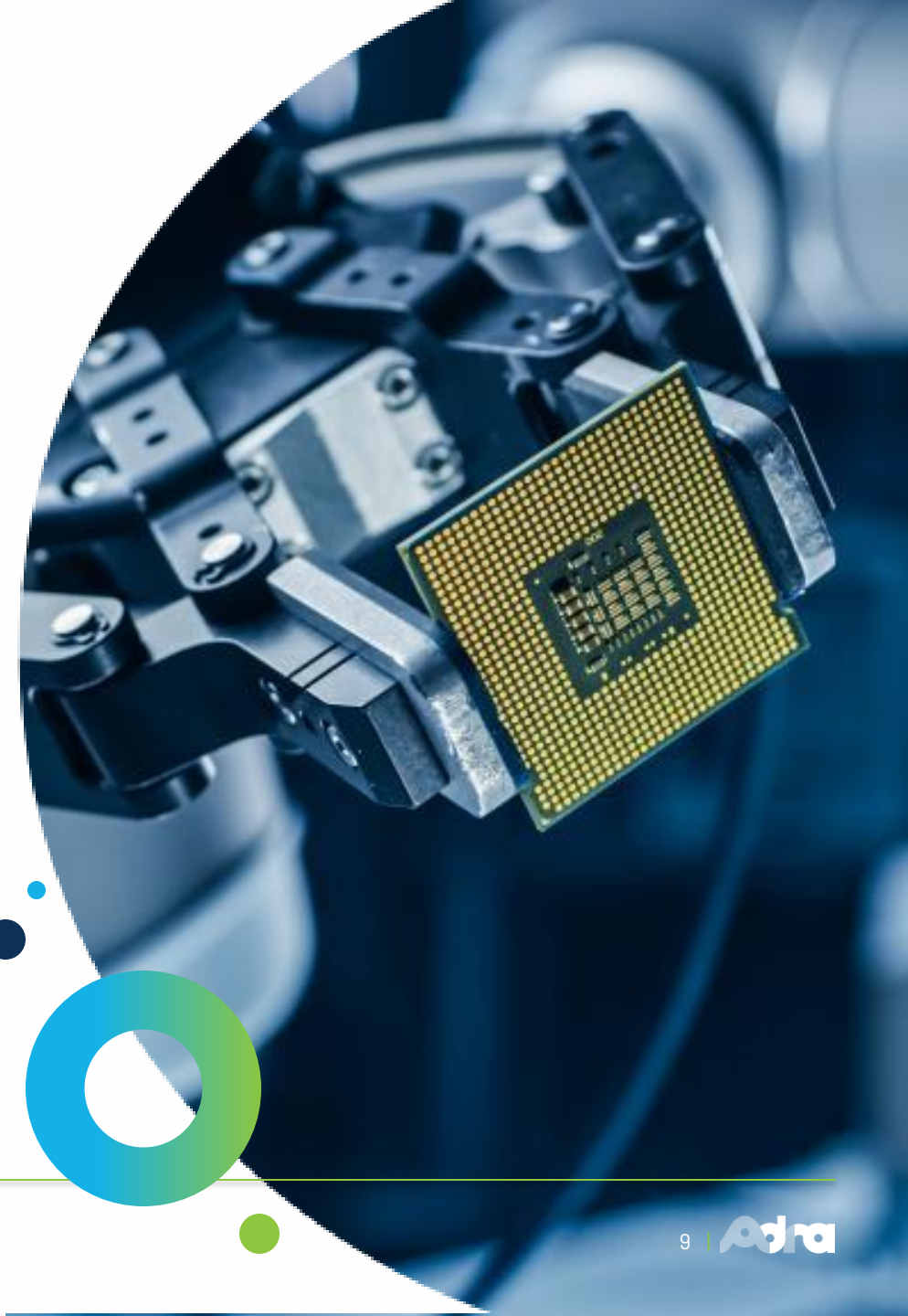
Tort liability arises **outside any prior agreement**. It applies when one party causes harm to a third party with no contractual bond between them. For **AI**, it is the **primary route to redress** for patients, bystanders, and job applicants harmed by systems they never contracted with.

The Three Classic Elements

- **Fault:** negligent or intentional conduct breaching a duty of care
- **Causation:** direct causal link between conduct and harm (conditio sine qua non)
- **Damage:** actual harm suffered – physical, economic, or non-material

Why AI Strains All Three Elements

AI decisions emerge from **model weights and training data**. Opacity makes it hard to identify who acted wrongly (**fault**), what caused the output (**causation**), and which harm is attributable to the AI vs. other factors (**damage**). The result is an **access-to-justice deficit** for victims.



The *Conditio sine qua non* Problem

The *conditio sine qua non* test

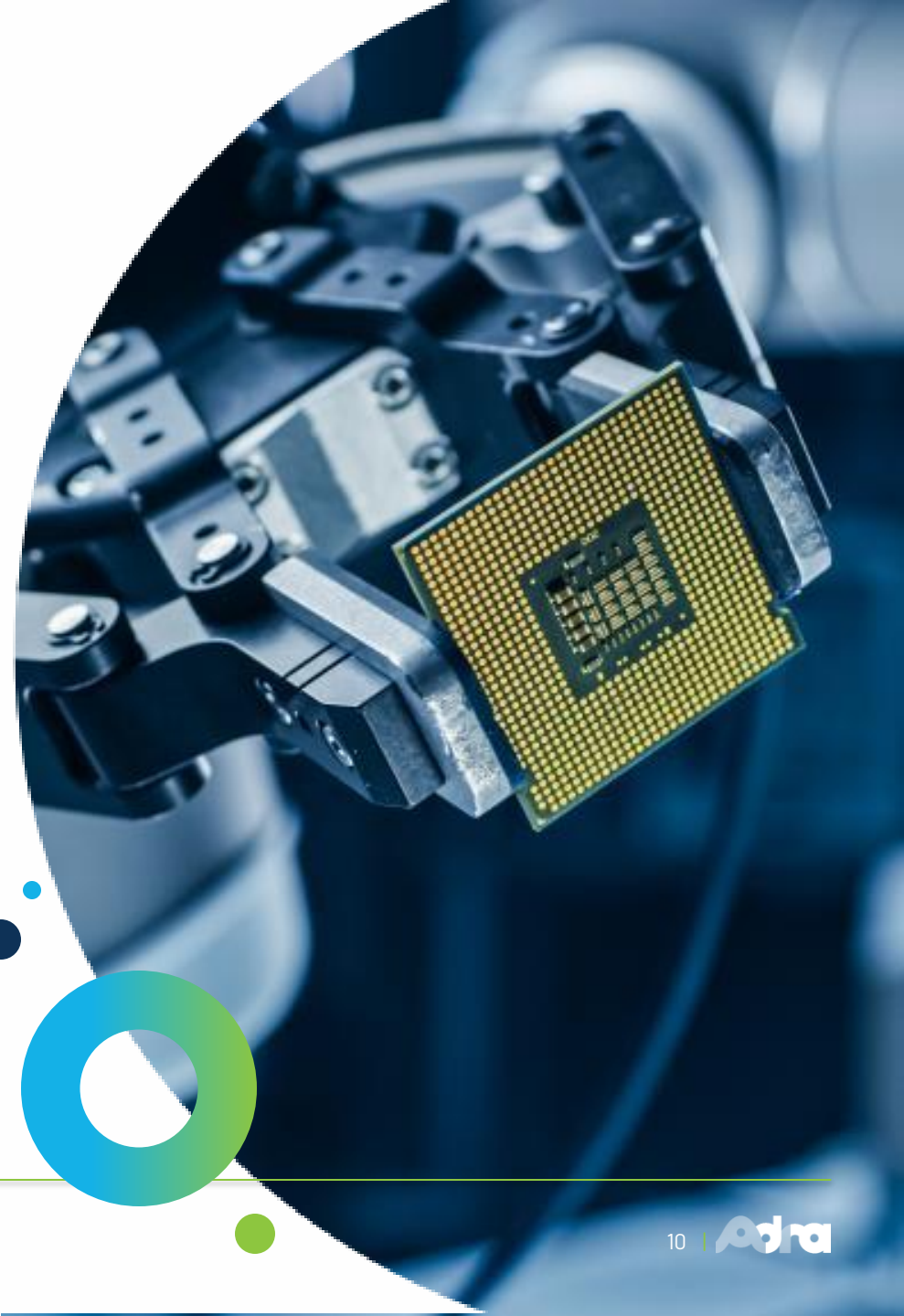
Would the harm have occurred anyway, even without the conduct at stake? If the answer is no, causation is established. **Counterfactual question**: what would have happened where the act did not take place?

Why AI breaks the test: the explainability deficit

Answering “would the harm have occurred without this AI output?” requires **understanding why** the system produced that output. With **opaque models**, this is structurally impossible. Without **explainability**, the *conditio sine qua non* test can’t be applied in any meaningful sense.

The EU legislative response

The **proposed AI Liability Directive (withdrawn)** had introduced a **rebuttable presumption of causality**: where a claimant demonstrates that the defendant violated an obligation under the AI Act and that a plausible link to harm exists, causation is presumed.



A key distinction

Fault-based liability

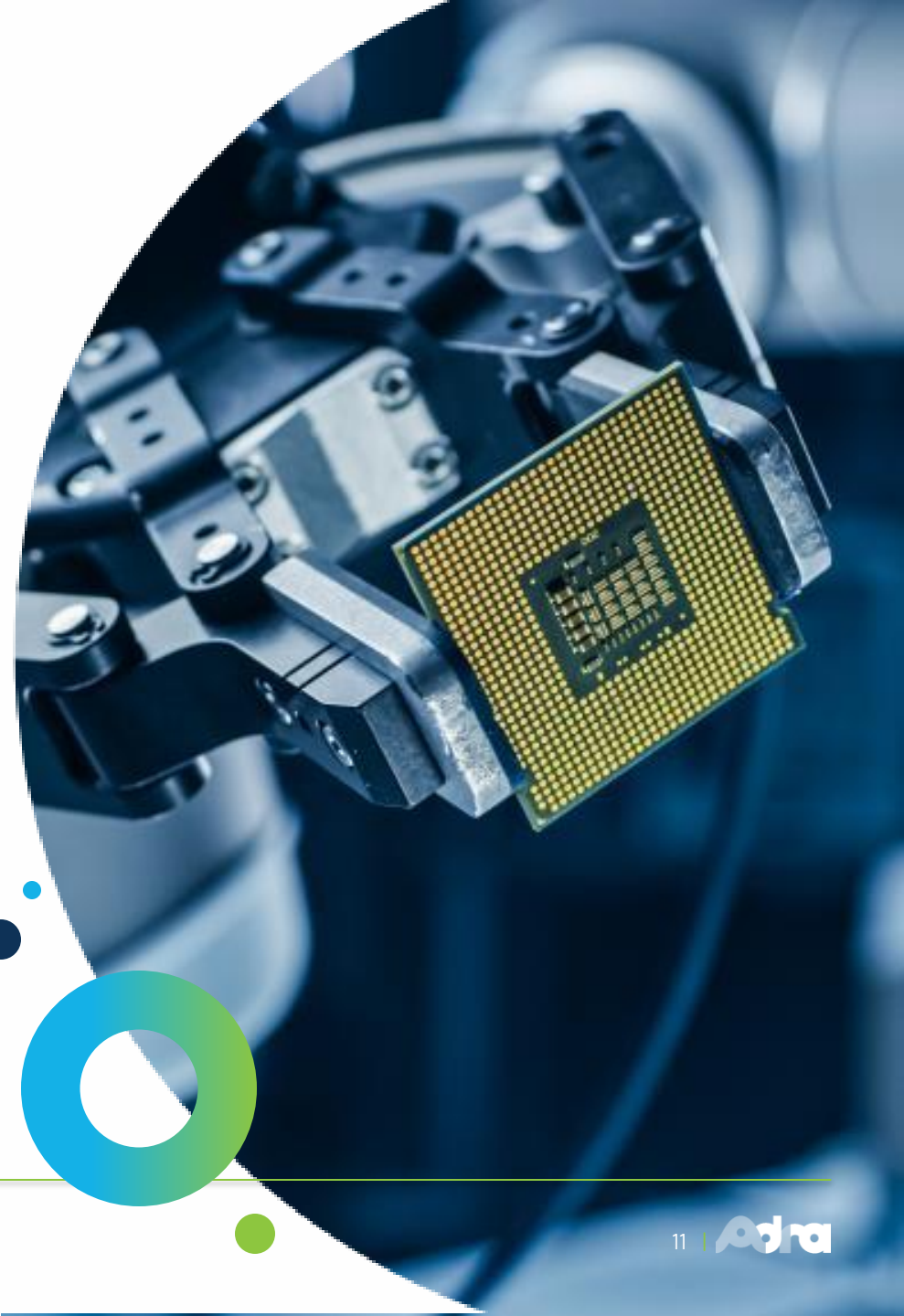
The claimant must prove that the defendant acted **negligently or intentionally**. Liability is conditioned on a **breach of a standard of care**. The defendant is liable because he/she **acted wrongly**.

Strict liability

The claimant need **only prove harm and causation**, not fault. Liability attaches to the **activity or the product** itself, regardless of how carefully the defendant behaved. It is typically reserved for **abnormally dangerous activities or defective products**. Examples: Product Liability under the revised **PLD**.

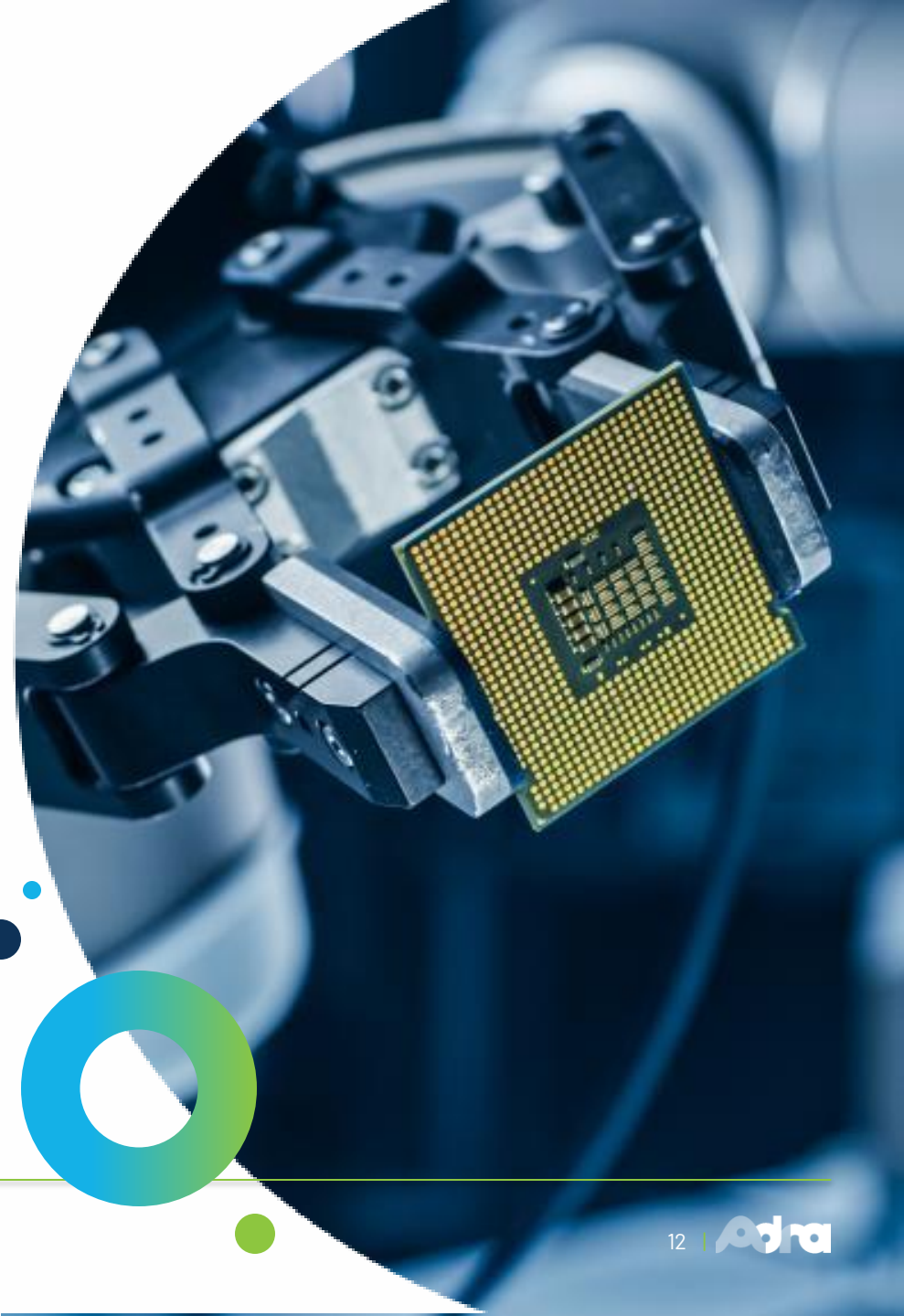
Why the distinction is critical for AI

With fault-based liability, opacity is critical to the claimant: without insight into the model's internal reasoning, proving negligence is structurally impossible. **Strict liability removes that obstacle** entirely, but with **risks** of penalising developers of systems that cause rare, unforeseeable harms. The **choice** reflects a **fundamental policy choice** about who bears the residual **risk of AI harm**.



Tort Liability: Challenges for AI

- **The attribution gap:** tort law was designed for **identifiable human actors**. In a **multi-layer AI pipeline no single party** makes the “decision” that causes harm. Fault becomes diffuse and legally unassignable.
- **The opacity problem:** Without explainability, claimants cannot **demonstrate fault**, establish **causation** or quantify which **part of the damage** is attributable to the AI.
- **Non-determinism:** the **same input** may produce **different outputs** across runs. This makes **counterfactual reasoning** (the basis of the *conditio sine qua non*) practically unreliable.
- **The autonomy paradox:** the more autonomous the system, the weaker the **causal link** between any human actor’s choice and the harmful output. **Agentic AI** can effectively sever the human causal chain entirely.
- **Evidentiary asymmetry:** all **relevant evidence** (model architecture, training data, logs, evaluation reports) is held by the **defendant**. **Victims** can’t access it without litigation



Divergent approaches across Europe

France – Strict Liability Orientation (Art. 1242 Code Civil)

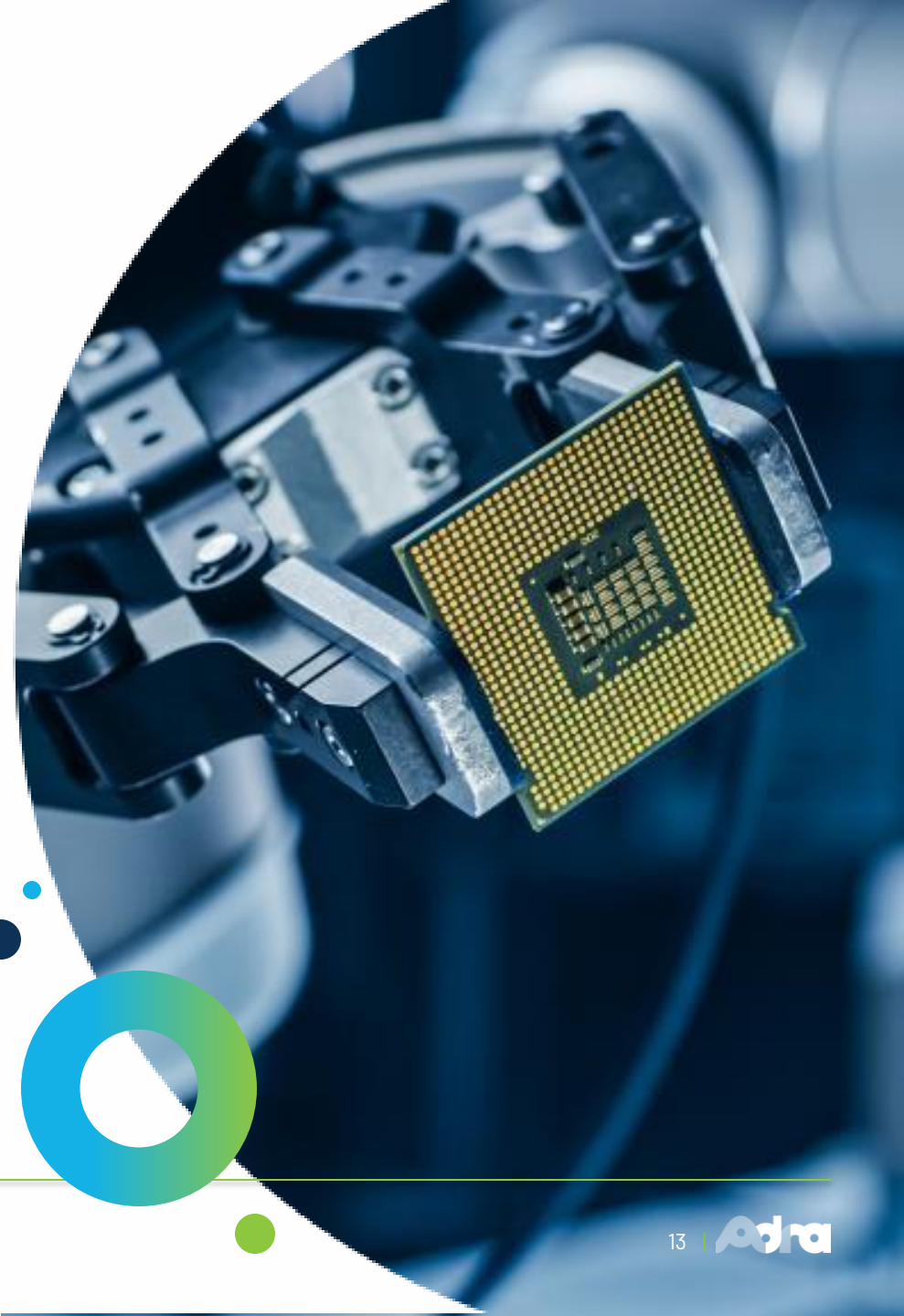
French law imposes liability on the “custodian” (gardien) of a thing that causes harm, irrespective of fault. Courts are applying this to AI: the **deployer in operational control** may be liable simply because harm occurred – **no need to prove negligence**. A major advantage for claimants.

Germany – Fault-Based Default with Sector Exceptions (§823 BGB)

German tort law is **fault-based**: the claimant must prove that the defendant acted **negligently or intentionally** and violated a **protected right** or a statutory norm. For **AI**, this means demonstrating a **defect in design, training, or deployment**. However, the **Product Liability Act** (Produkthaftungsgesetz) provides **strict liability for defective products**, which the revised **EU PLD** now extends explicitly to **software and AI**.

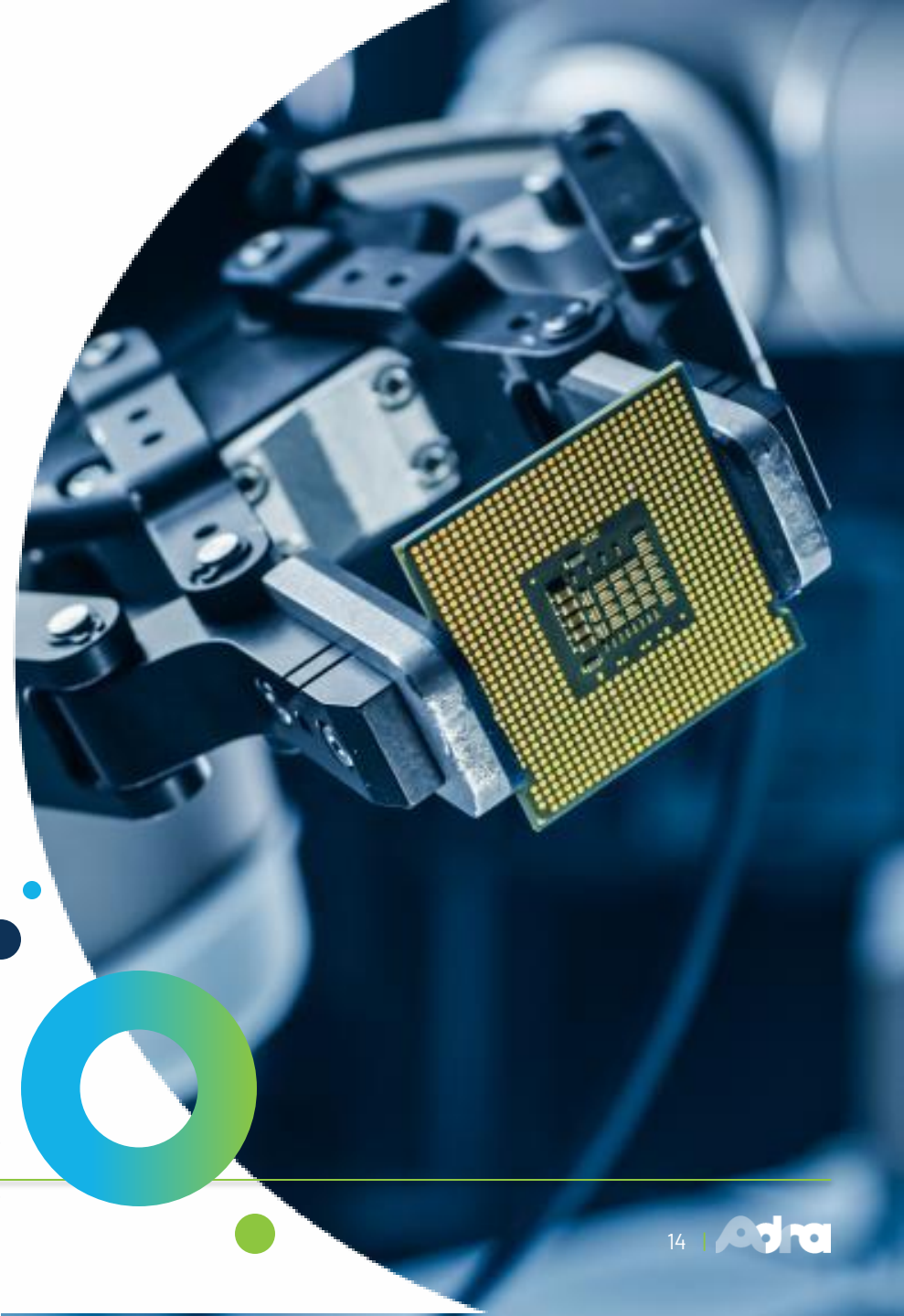
Italy – Dual Track: Art. 2043 and Art. 2050 Codice Civile

Art. 2043 requires proof of dolus or culpa (**fault-based**). **Art. 2050** governs “**dangerous activities**,” reversing the burden: the defendant must prove all possible precautions were taken. Early **case law** suggests **high-risk AI deployment** may qualify as a dangerous activity, a significant advantage for claimants.



The EU Legislative Response

- **Why divergence is a problem:** Forum shopping risks, unequal victim protection and uncertainty for AI providers.
- **AI Liability Directive Proposal – [WITHDRAWN]:** It introduced a **rebuttable presumption of causality** and mandatory **disclosure rights**. It was withdrawn in **2025**. The **EC** reserved the right to assess how to proceed (revised proposal / alternative approach).
- **Revised Product Liability Directive (RPLD):** Directive (EU) 2024/2853 applies from December **2026**, replacing the **1985 PLD** and explicitly bringing **AI systems and software** within the scope of **strict product liability** for the first time.
- AI Act as the **perimeter within which liability is assessed**. Non-compliance does not automatically entail civil liability, but makes successful **defence** in tort or product liability litigation
- **European Parliament: 2025 Study** (EPRS, 2025/776426): neither the RPLD nor the withdrawn AILD are sufficient to address the **liability risks of high-risk AI systems**. It called for a **dedicated strict liability framework** aligned with the AI Act's risk classification



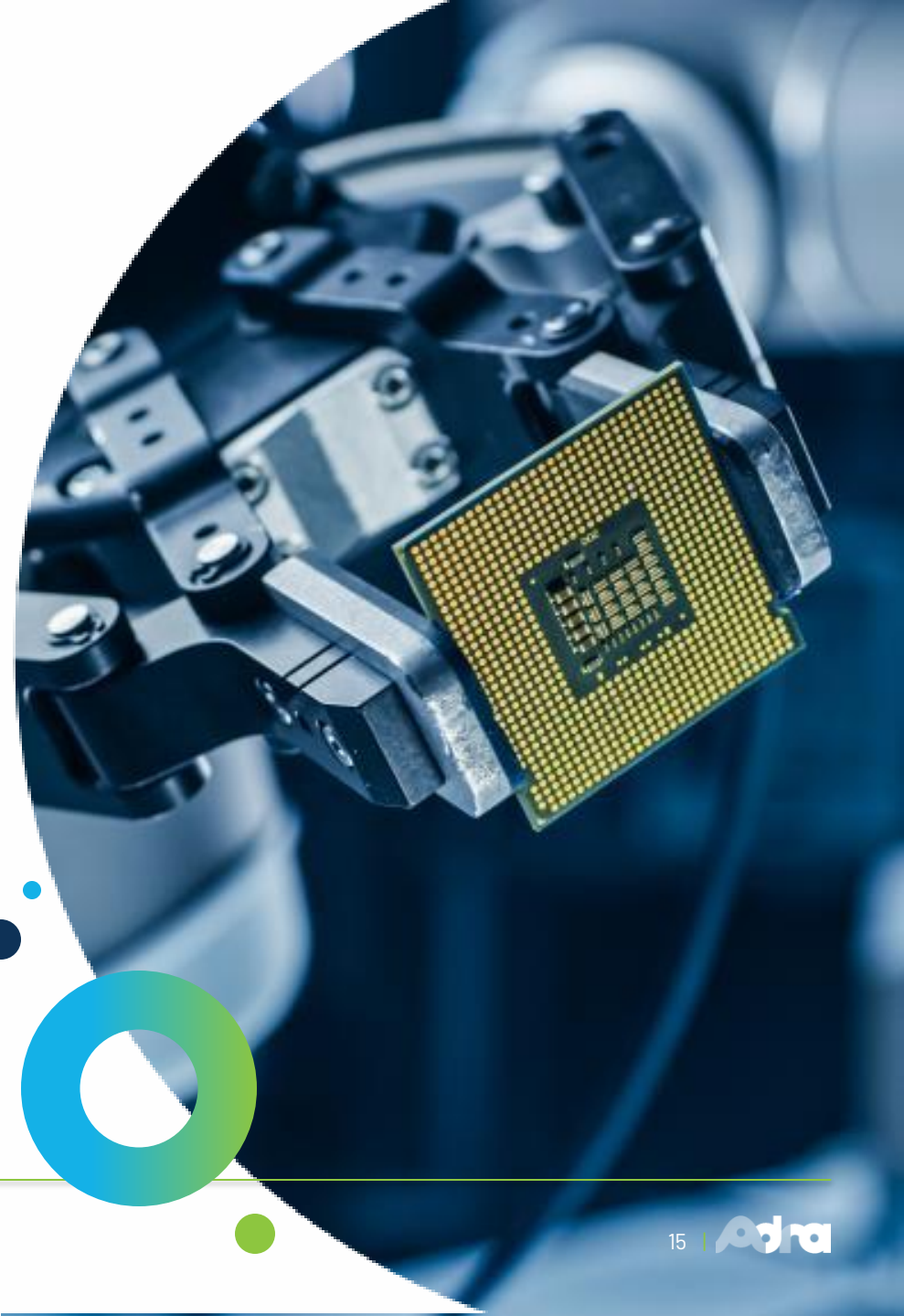
Revised Product Liability Directive

What the revised PLD covers for AI

It applies to **AI systems** placed on the **EU market** as **products** or embedded in products. It covers **harms caused by defective AI** (physical injury, property damage, data loss, and medically certified psychological harm). The **manufacturer, importer, and authorised representative, fulfilment service provider or a provider of an platform through which the product has been purchased** may all be liable. **Deployers** who **substantially modify** a system after placing on the market become liable as manufacturers.

Key Innovations

- **Software and AI as “products”** (both standalone software and AI systems embedded in physical products)
- **Strict liability** without proof of fault.
- **Data corruption as damage**
- **Post-market software updates** (after the product is placed on the market) which makes a product defective
- **Disclosure obligations:** claimants may request courts to order manufacturers to disclose relevant technical documentation
- **Rebuttable presumptions of defect and causation:** where the manufacturer fails to disclose the evidence, the product does not comply with product safety legislation or the damage was caused by an obvious malfunction, or the claimant faces despite of the disclosure of evidence excessive technical difficulty in proving the defect or the causal link but succeeds to demonstrate that the defectiveness and the causal link are likely, courts may presume both



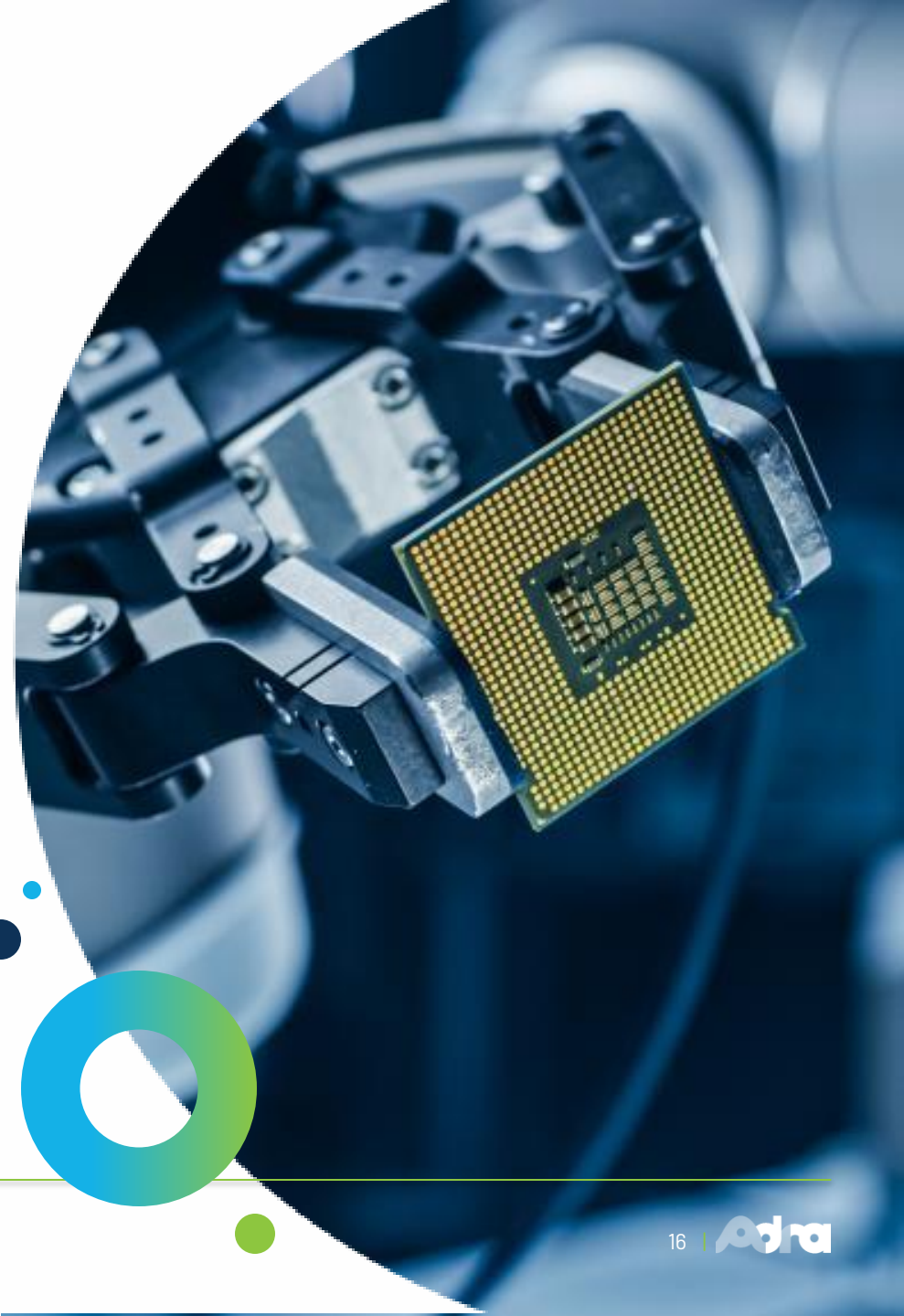
Shifting the burden of proof under new PLD

Burden of proof remains on injured person, but **is eased**:

- Manufacturers obliged to **disclose necessary information** in court, subject to protection of trade secrets / confidentiality
- **Presumptions** of defectiveness/causal link where:
 - a) the manufacturer fails to disclose information
 - b) the product does not comply with safety requirements,
 - c) the product obviously malfunctions,
 - d) the damage is typically consistent with the defect OR
 - e) technical or scientific complexity causes excessive difficulties to prove liability (e.g., 'black box' AI systems or pharmaceuticals)

Extension of 10-year period of liability to 25 years for **latent health injuries**

Removal of financial ceilings to ensure full compensation of victims



The case of state liability

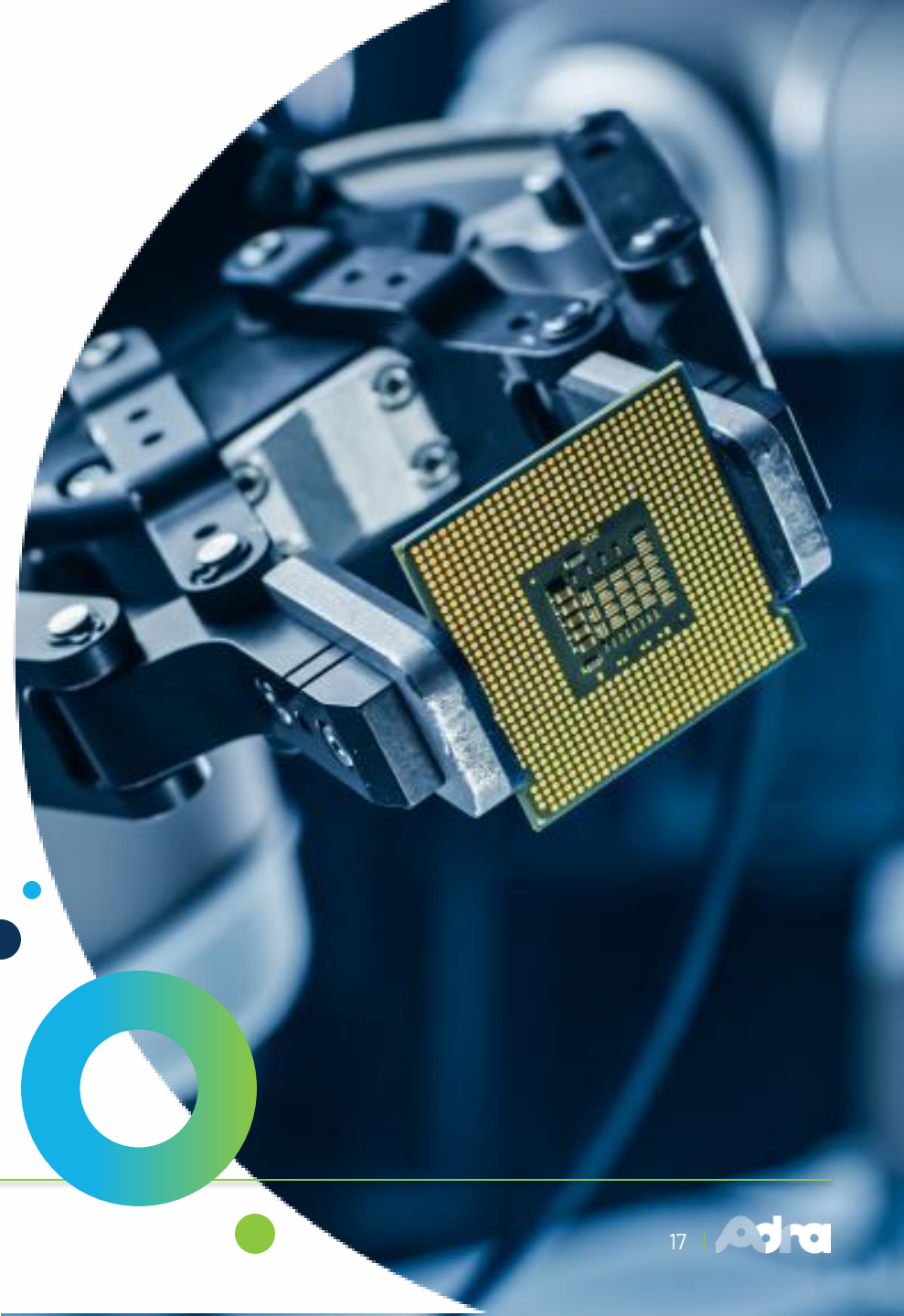
Why discussing particularly about state liability by the use of AI systems?

Public authorities exercise specific public power that:

- is restrictively predefined and described by the law
- is obligatory and produces direct legal effects to citizens

Result:

- a) the use of AI must be also provided by the law
- b) public decisions and acts, even AI-generated, must comply with the rule of law, obligations for explainability and accountability



Challenges – examples

- **Different jurisdictions may have different rules** regarding state liability and stricter procedural provisions for the claimant's compensation
- **Lack of knowledge** and **control to the AI system** on behalf of the public authority – space to deny liability

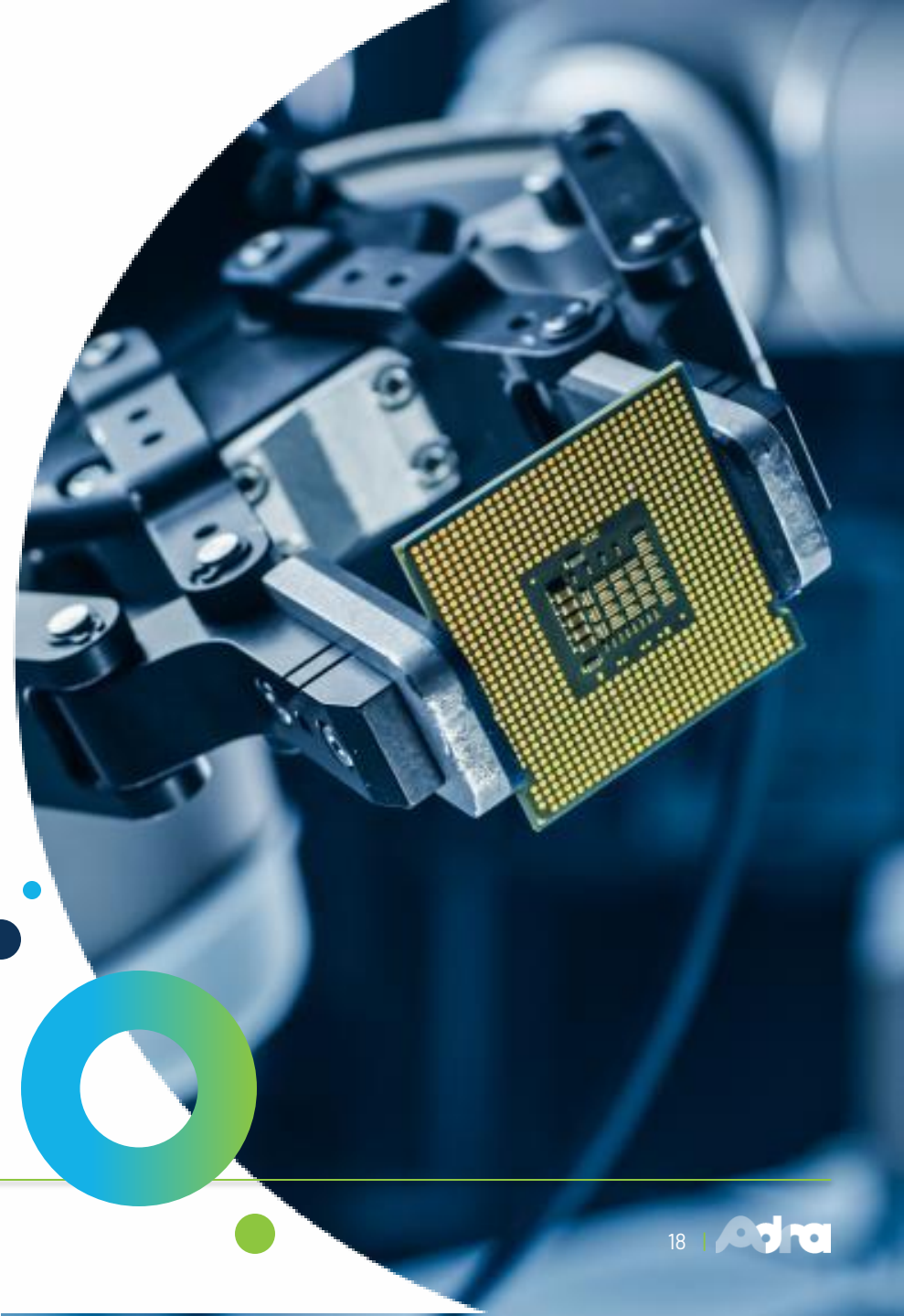
Dutch child care benefits fraud scandal (2012-2019)

(self-learning algorithm used by the Dutch tax authorities to create fraud profiles to spot child care benefits fraud – tens of thousands families were asked to return amounts given as child care benefits – the scandal was revealed in 2019 – the government resigned – fines to the Dutch tax authority due to violation of data protection rules and discrimination)

COMPAS (a support software used by US courts to predict the likelihood of a defendant to commit a crime – Loomis case before the Supreme Court – the algorithm violates the defendant's right to a due process, mainly due to its black box character and the inherent biases)

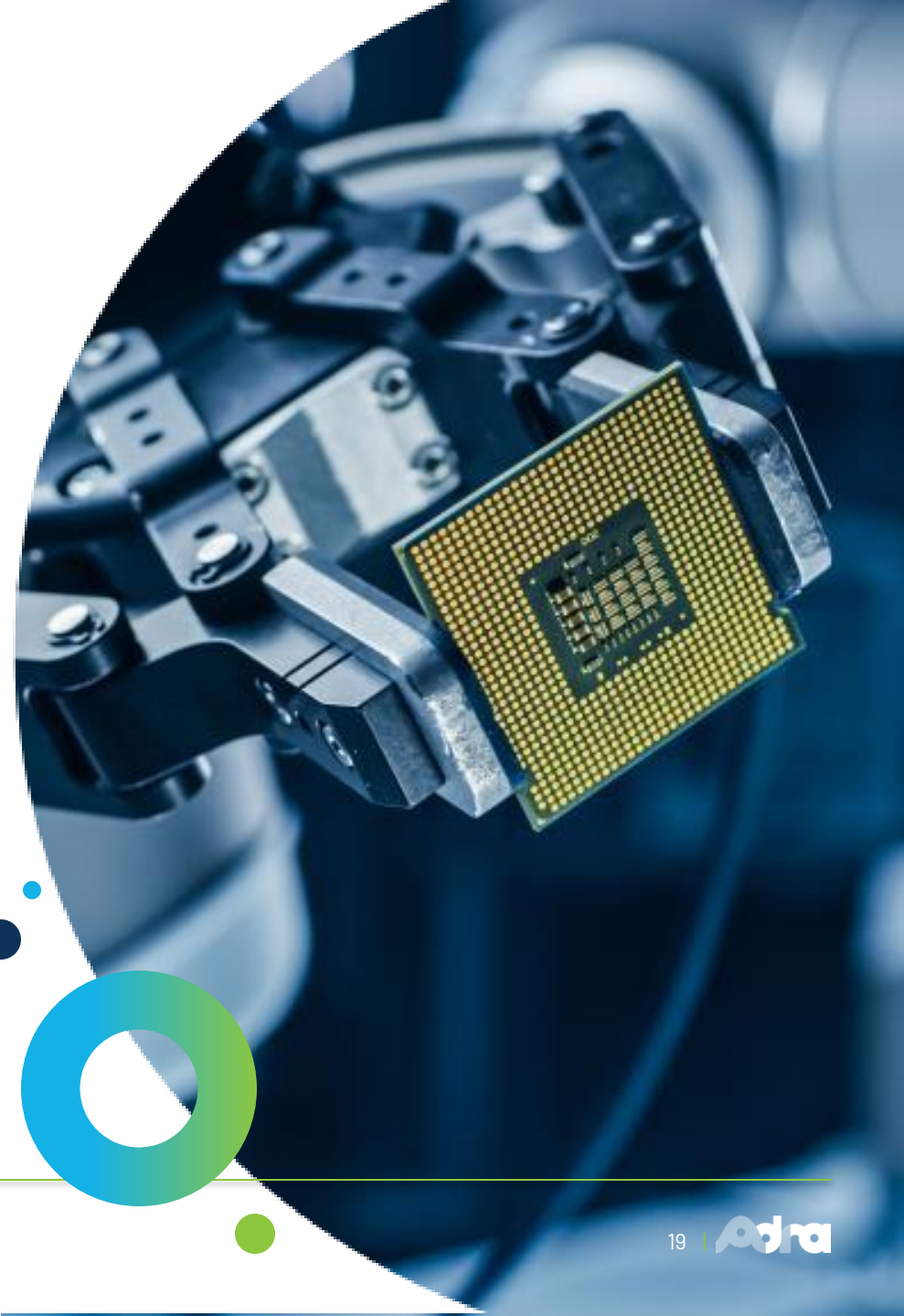
Datajust (France, 2020)

An AI system designed to process case law data relating to personal injury and bodily harm compensation intended to help both lawyers and judges in calculating potential compensations, processing past law cases – due to its limited use and the criticism for unfair results, calculated human suffering and restricting judges' discretionary power, the system was abandoned two years after its introduction



How to address particularly issues regarding the use of AI by public authorities

- The AI Act already has identified as **High-Risk AI systems**, systems used in the exercise of public power (administration of justice and democratic processes, education, access to and enjoyment of essential public services and benefits, law enforcement, migration, asylum and border management)
- **Fundamental rights impact assessment** (Art. 27 AI Act)
- **Ethical principles**: human oversight, transparency, explainability, accountability
- **EU Guidelines concerning High-Risk AI systems** (recently issued for public consultation)
- Provision of the use of AI systems in public sector in **specific law** (requirements, safeguards, structured liability rules depending on the context of the relevant AI-assisted public service)
- **Art. 22 GDPR**: The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects (right to human intervention). Unless the decision:
 - is necessary for entering into, or performance of, a contract between the data subject and a data controller;
 - is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
 - is based on the data subject's explicit consent



Ethical Foundation of AI Liability

Key ethical principles relevant to AI liability span from Accountability, to Explainability, to Human Oversight

Common ethical expectations

- **Explainability** proportional to risk
- **Auditability** throughout the lifecycle
- **Human contestability** of decisions
- **Documentation** and traceability
- Continuous **monitoring** and risk mitigation

Liability as the legal expression of ethical expectations regarding responsibility, control and prevention of harm, which are increasingly being translated into liability provisions.

HLEG-AI **Ethics Guidelines for Trustworthy AI** and the **AI Act** (risk management, documentation, traceability, human oversight, monitoring)

Ethical principles increasingly function as **operational benchmarks** (fault, due diligence)



Part 2. Deep Dives

Autonomous Agents / Key concepts

Autonomous agent

A software entity that **perceives its environment, pursues goals, and takes actions autonomously**, without requiring a human to approve each individual step. It can plan sequences of actions, adapt to new information mid-task, and interact with external systems (APIs, databases, other services).

Agentic AI

A **LLM** or **foundation model** operating in an **agentic mode**: given a high-level goal, it decomposes the task, selects tools, executes sub-tasks iteratively, and refines its outputs across multiple steps, all without human approval at each step. The model observes the result of each action and **adjusts** its plan accordingly.

Multi-agent system (MAS)

A **network of autonomous agents** that interact, coordinate, and delegate tasks among themselves, often without any centralised human control over individual interactions. An **orchestrator agent might** assign sub-tasks to specialised worker agents; each worker may in turn invoke further tools or **sub-agents**.



Oversight Approaches

Different models with different implications for liability:

- **Human-in-Command (HIC):** Humans maintain overarching direction and final authority. They define the limits, quality criteria, and risk parameters, intervening at a strategic level rather than every decision cycle;
- **Human-in-the-Loop (HITL):** A human reviews and approves every individual output or decision before it takes effect.
- **Human-on-the-Loop (HOTL):** The AI operates autonomously, but humans monitor the process and have continuous veto or override power. Humans must have a genuine, timely opportunity to intervene, also considering the speed and volume of decisions.
- **Human-out-of-the-Loop (Operational Autonomy):** The system operates fully autonomously, without any human review or override capability at the operational level. Oversight is exercised only at the design and governance stage, before deployment.

The level of human oversight is a key of liability allocation: the closer the human is to the decision, the more exposed they are to a fault-based claim for having exercised (or failed to exercise) their oversight function adequately.

The **AI Act (Art. 14)** requires that **high-risk AI systems** be designed and deployed to allow **effective human oversight**.



Liability Challenges in Agentic and Multi-Agent Systems

- **Fragmentation of liability across the chain:** in a **MAS**, the harmful output is the product of **interactions between multiple agents** developed and operated by different parties. **No single actor** controlled the outcome: **cause and effect are distributed** across autonomous sub-systems;
- **Emergent harm in multi-agent interaction:** in a **MAS**, **harmful behaviours** (harmful outcomes that arise from the interaction pattern rather than from any single agent's design) might **emerge from the system's architecture**: no party foresaw or chose the harm;
- **The approval paradox:** the approver may have been presented with hundreds of outputs per hour, in a format optimised to drive acceptance, with insufficient time or information to meaningfully evaluate each one. The question is whether **formal approval** constitutes **genuine oversight**, or merely provides a **liability transfer** mechanism that insulates the developer while exposing the operator.

Future directions

- Who is liable when a **human “approved”** but could not meaningfully review?
- Towards **strict liability for high-capability agentic systems? EPRS study** (July 2025): fault-based liability is structurally inadequate for fully autonomous agents. Strict liability (or at least a mandatory presumption of causality) as a valid option
- **Legal personality for AI agents?** A limited legal status for advanced AI agents (comparable to corporate personality) combined with mandatory insurance or compensation fund obligations, despite **so far rejected**, could be an option for the **liability vacuum** created by fully autonomous agents



Liability for discriminatory outcomes

AI systems tend to **replicate social norms embedded in their training data**, that leads to reproduction of stereotypes (no reflection of the evolving social contexts) and discriminatory outcomes

Examples:

- Rejection of loan or mortgage or credit card applications, excluding diverse family structures or diverse types of individuals (Apple's credit card algorithm)
- Discriminative hiring procedures and CV classification (Amazon AI recruitment tool, LinkedIn job recommendation tool)
- Healthcare cost algorithms
- AI Bias in skin cancer diagnosis for dark skin tones
- Facial recognition errors in law enforcement AI systems (under-representativeness of dark-skinned women faces may lead to false arrests, COMPAS AI system)
- Fraud detection and insurance in financial services



Case law 1/2

A growing wave of litigation concerning AI bias is rising, mostly in the U.S. where AI is more broadly used than in Europe.

Legal basis of the lawsuits: mainly anti-discriminatory laws, privacy regulations

Examples:

- **Mobley v. Workday** (*AI hiring system, used both as a platform from job seekers, also as a tool inside companies processes for job applications screening*)
The court has preliminarily highlighted that discriminatory practices can equally come from a human and an AI agent supportive tool, thus both can violate the law and be held liable. (District Court of California 2025, the decision is expected)
- **Louis v. SafeRent solutions** (*automated tenant-screening services*) Credit scoring data that reflected historic racial disparities resulted in disproportionately more housing denials for Black and Spanish applicants (District of Mass. 2023)



Case law 2/2

- **Deliveroo case** (*AI system to calculate the company's riders' performance*) Deliveroo riders accused Deliveroo of discrimination through the platform's algorithm. According to the riders, it ranked them according to their performance and restricted access to working slots because of riders' absences due to strikes or sickness. As a result, Deliveroo was sentenced to a €50 000 fine (December 2020, Court of Bologna)
- **Syrl case** (*System Risk Indication, an algorithm used by the Dutch government to profile individuals and detect potential welfare, tax and social security fraud*) The Court ruled that SyRI violated among others, the right to privacy (art 8 ECHR), posing further important discriminatory issues against marginalized groups of people (2020, District Court of the Hague)
- **Loomis case** (*COMPAS an AI system for calculating risk of re-offending*) COMPAS was considered violating due process and fair trial rights, due to its black-box character and the inherent bias of taking gender and race into account to lead to its outcome (U.S. Supreme Court)



Part 3.

Examples and Sector-specific Cases

Autonomous vehicles and the Trolley Problem

The Trolley Problem - Classical ethical dilemma

A vehicle must choose between: harming one person OR harming several others.

Example

An autonomous car may have to decide whether to:

- hit a pedestrian,
- swerve into another vehicle,
- sacrifice the passenger,
- minimize total harm.

Why It Matters Legally

The key legal question is whether developers can pre-program ethical choices into AI systems.

This raises issues concerning:

- negligence,
- product defectiveness,
- human dignity,
- discrimination,
- proportionality.



Human Oversight

Product Liability and Autonomous Vehicles – Autonomous vehicles may trigger strict liability if defective.

A vehicle may be considered **defective even if no human directly caused the accident.**

A central issue is determining when the human remains responsible.

Levels of automation matter: assisted driving, partial autonomy, conditional autonomy, full autonomy.

Legal difficulty- Humans may: overtrust AI, react too slowly, misunderstand alerts, lose situational awareness.

This complicates:

- contributory negligence,
- operator fault,
- allocation of liability.



Case laws

Courts reject abstract philosophical dilemmas (the "Trolley Problem") in favor of concrete product liability, design defect, and operator negligence frameworks.

- **Banner v. Tesla (2025/2026):**

A \$243M verdict upheld the reality of design defect and over-marketing liability when systems fail to guard against known visibility edge-cases.

- **Story v. Tesla (2026):**

Settled pedestrian FSD fatality emphasizing the opacity paradox of camera-only sensor arrays in complex traffic situations.

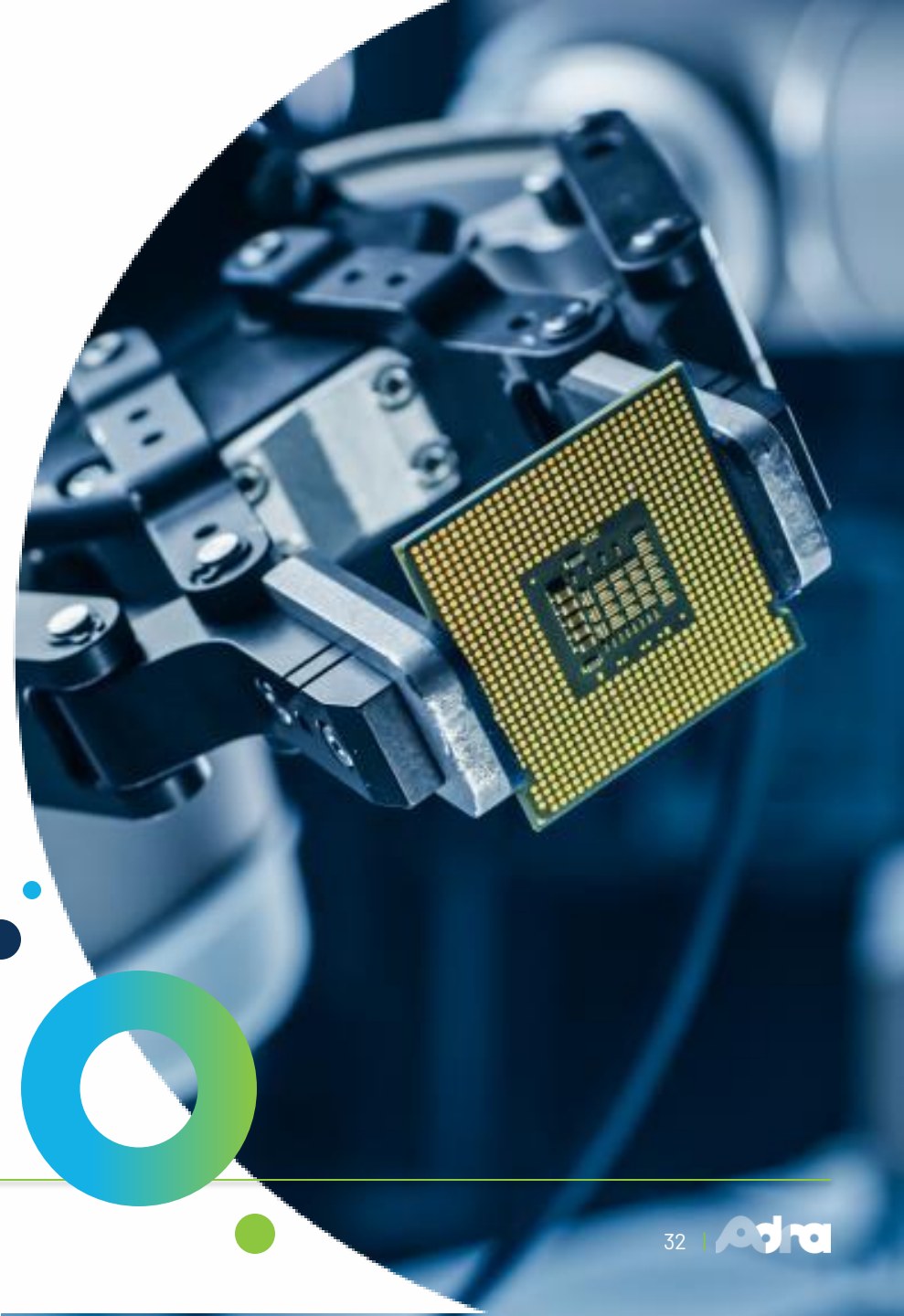
- **State of Arizona v. Vasquez (2023):**

Underscored shared human-in-the-loop duties, holding safety drivers criminally responsible for failure to monitor and intervene.

States response:

German Autonomous Driving Act: Banned target profiling (age, gender, race) and rejected utilitarian trade-offs, making it illegal for an AV to sacrifice a bystander to save a passenger.

California AB 1777 (2026): Shifted traffic ticket liability entirely away from human occupants, authorizing police to cite the AV manufacturer or operator directly for vehicle violations.



Liability in AI-Powered Clinical Decision Support Systems

- A **uniquely high-stakes context**: healthcare AI operates where errors are irreversible (death or permanent harm).
- **Who is liable in case of harm**: the physician, the hospital, or the manufacturer?
- **Courts in the US and EU** have not yet settled a definitive allocation rule, and the three channels can be pursued cumulatively
- **Legal Framework**
 - **EU AI Act**: AI systems intended to be used for diagnosis, prognosis, or treatment decisions are **high-risk** with related obligations
 - **Medical Devices Regulation (MDR, EU 2017/745)**: clinical decision support software that influences clinical decisions is a **medical device** requiring **clinical evaluation and CE marking**.
 - **GDPR**: health data is **special category data (Art. 9)**; processing for AI training or inference requires explicit legal basis, and **automated decisions** with significant effects on patients trigger **Art. 22** rights.



Real Cases

1) IBM Watson for Oncology (2015–2022) – wrongful treatment recommendations, no litigation

Deployed at over **230 hospitals** worldwide. The system generated “**unsafe and incorrect**” cancer treatment recommendations, since it was trained on **synthetic cases** devised by a small group of physicians and reflecting **individual doctors’ preferences**. **MD Anderson Cancer Center** terminated its \$62 million contract in 2016. IBM sold **Watson Health** in 2022. **No litigation**, but example for **training-data liability** and **manufacturer accountability**.

2) Epic Sepsis Model (2021–ongoing) – widespread clinical deployment, no litigation resolved

A 2021 study found that **Epic’s proprietary sepsis prediction model**, deployed at hundreds of US hospitals, **missed two-thirds of sepsis cases** in a sample of 38,500 hospitalisations, while generating **false alerts on 18%** of all hospitalised patients. The **model’s AUC** was 0.63, well below Epic’s published 0.76–0.83 range. Epic subsequently released an **updated model**. **No malpractice case** has been decided yet on these facts.



Coverage Denial AI

Coverage Denial AI: AI systems used in US by **health insurers** to approve or deny coverage duration for **post-acute care**, based on predicted medical necessity or treatment outcomes. It often bypasses physicians' discharge assessments.

Estate of Lokken v. UnitedHealth Group (Minnesota, 2023–ongoing) – first active AI healthcare liability litigation, opened by the **families** of two deceased Medicare Advantage beneficiaries. UHC's **denial rate** for post-acute care more than doubled after nH Predict deployment in 2019. **Court:** breach of contract (2025) and disclosure of AI documentation and other material (2026). The case is **proceeding**.

No direct equivalent exists in the EU, as access to care is primarily mediated through **national health systems**. Closest conceptual framing: **automated decision-making under GDPR Art. 22**, potentially combined with the AI Act's **high-risk classification** if the system affects eligibility for essential benefits (Annex III).



Open Points

- **Alert fatigue:** a system with a very high false-positive rate induces clinicians to ignore alerts systematically. When a true positive is missed in this context, the question is who is responsible (fault).
- **Bias in training data:** AI systems trained on non-representative datasets produce systematically inferior performance for under-represented groups.
- **The AI-as-standard-of-care paradox:** as AI tools become widely adopted, **failing to use AI** may itself become negligent. A clinician who disagrees with AI and is wrong now faces **greater liability** than one who had no AI tool at all, creating incentives toward **uncritical deference** to AI outputs.
- **Mandatory disclosure of real-world performance: AI Act** (post-market surveillance obligations) and **RPLD** (disclosure mechanisms)



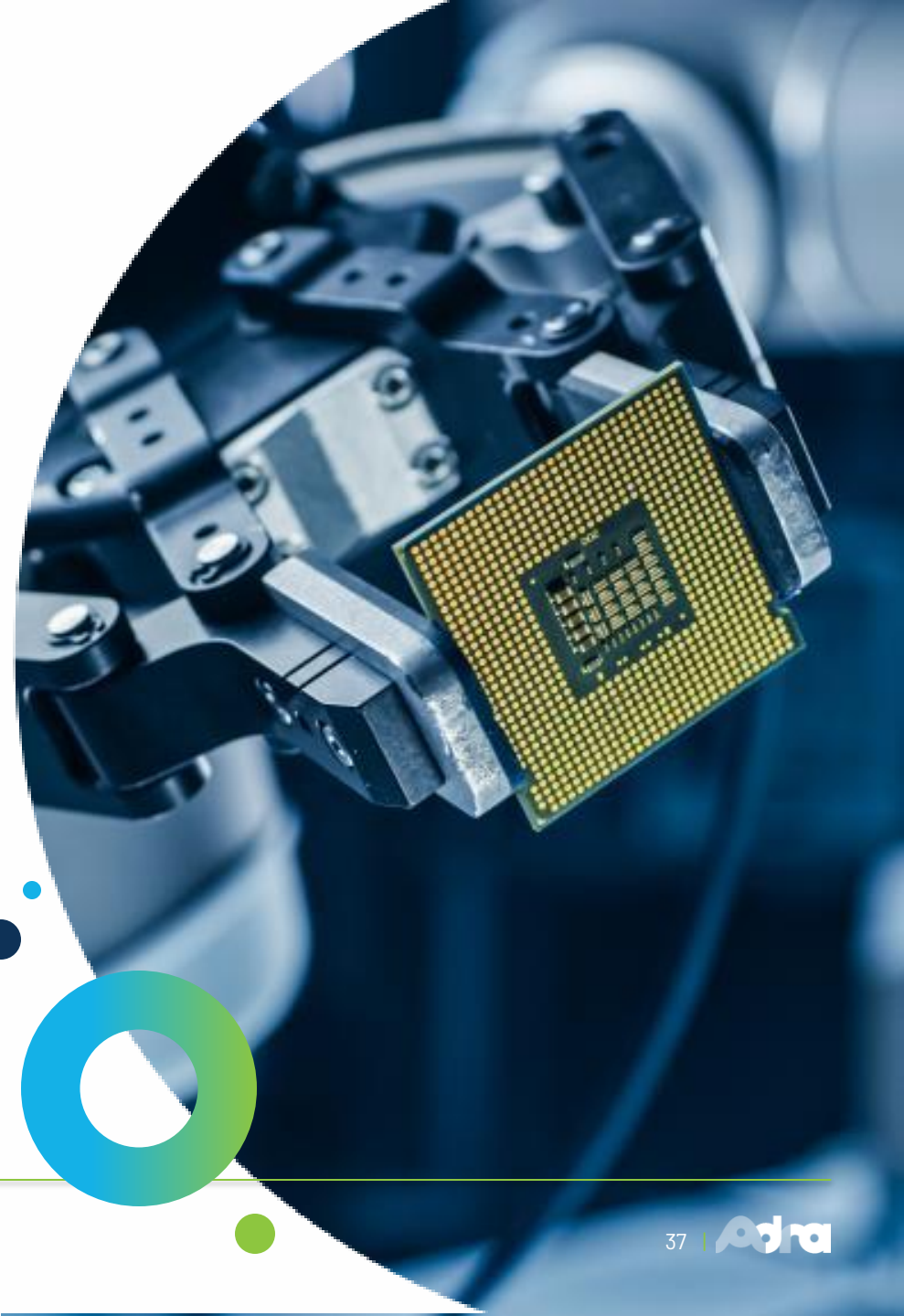
Liability of LLMs and chatbots

Liability of AI systems can be distinguished, in general, between two main categories:

- AI systems developed for a specific context and deployed for a targeted scope by an organization (public or private)
- LLMs and GPAI models used in multiple contexts (direct interaction between the model and the end-user)

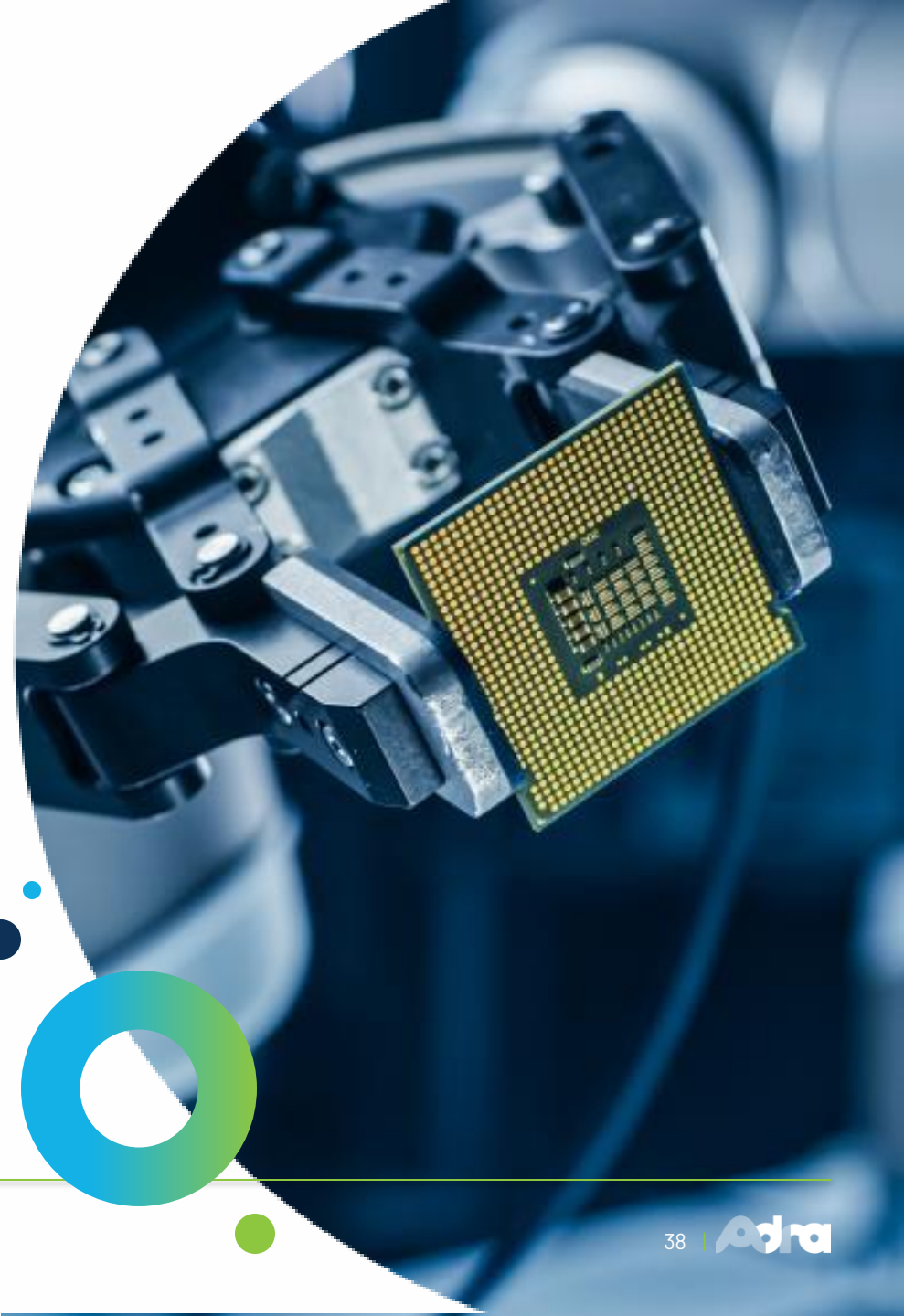
In the A) case: - the deployer/operator of the system is known
- the scope and the context is defined
- easier to define the defect and the causation

In the B) case: - chatbots are meant to be used in multiple contexts
- participation of the victim in the interaction
- questioning legal liability, extremely difficult to prove defect, causation



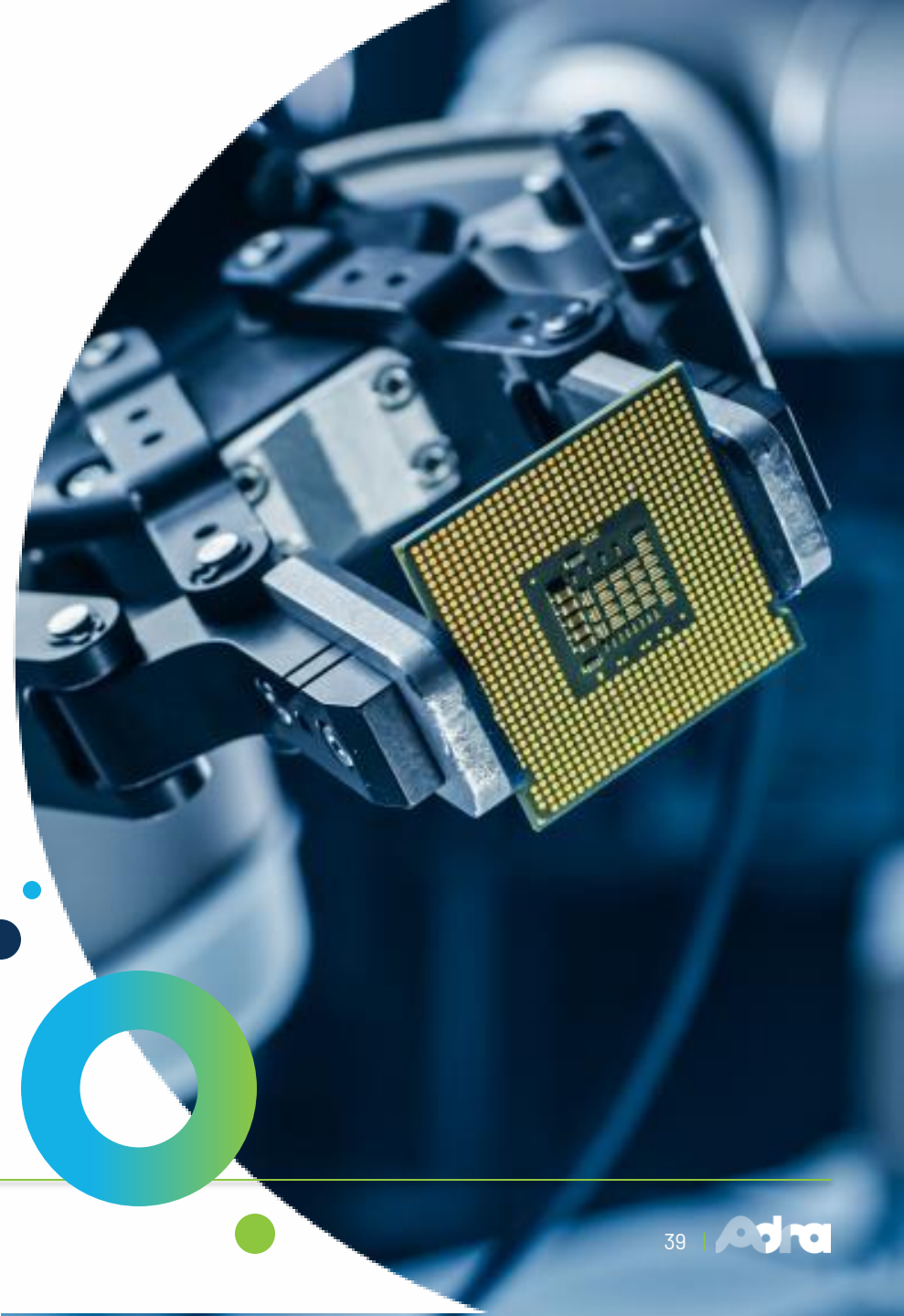
Main contexts of LLMs and GPAI liability

- Damage caused from emotional dependance and addiction to chatbots encouraging self-harm or illegal actions
- Deepfakes/misinformation
- IP infringements (the GPAI outcome in most of the cases violates directly IP rights if there is no consent from the IP owner, so we do not deal with a defective product, that was supposed to operate in a right way, but with an illegal one by default)



A) Emotional dependance 1/2

- **Raine v. OpenAI** (August 2025, San Francisco County Superior Court) Adam's parents sued OpenAI for the death of their 16-years old son, who had committed suicide after his deep engagement with ChatGPT. According to the claimants, the company failed to implement security measures to protect vulnerable users, such as people and mostly children with mental issues.
- Humanlike language – false empathy, emotional attachment that leads to emotional addiction to AI companions – simulated emotional realism
- OpenAI was accused of having removed safeguards in order to be more user friendly
- OpenAI's position: the chatbot had many times consulted Adam to seek for professional care and help



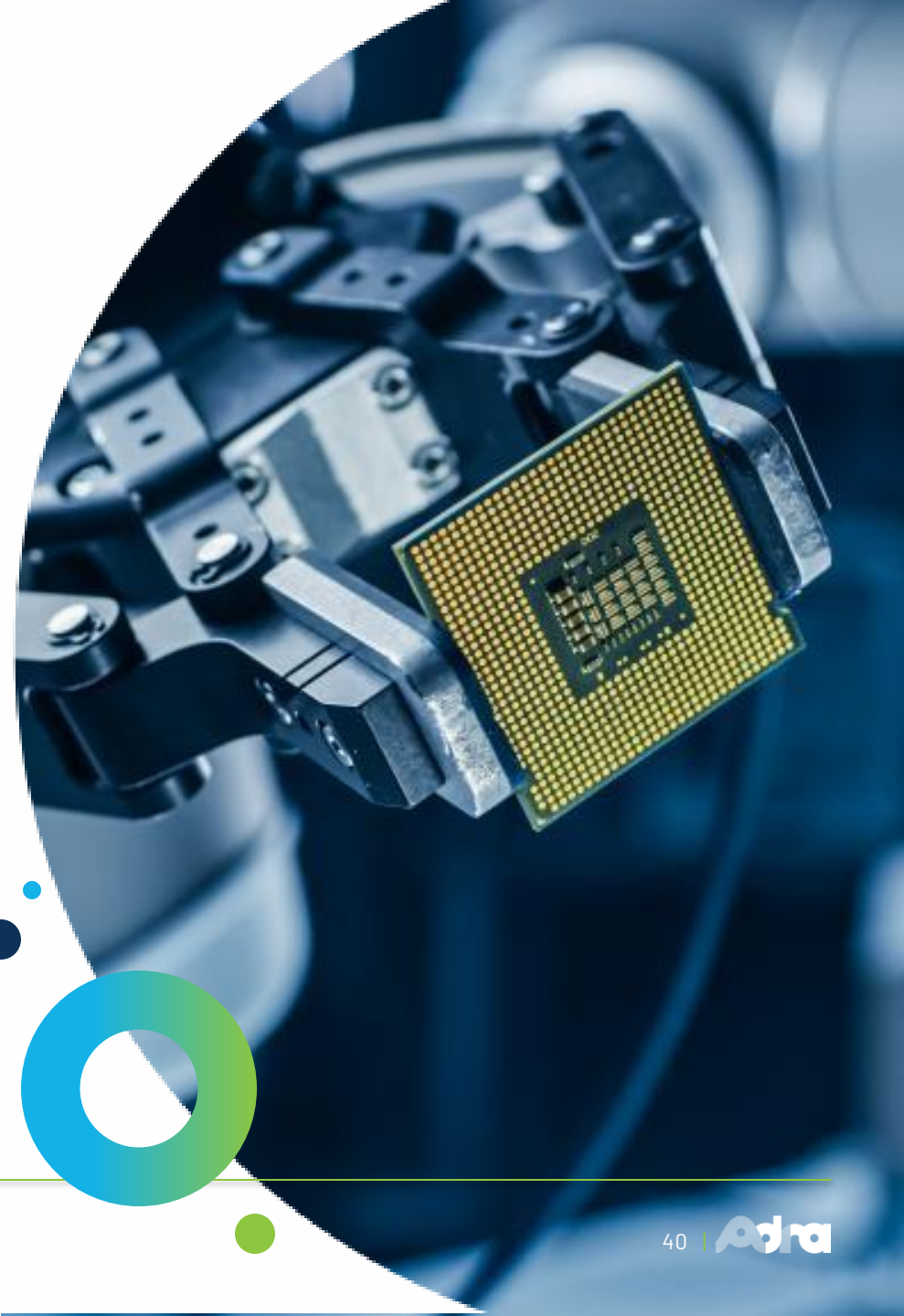
A) Emotional dependance 2/2

Garcia v. Character Technologies (October 2024) chatbot accused for the claimant's son's suicide (Interesting ruling: both Character.AI and Google held liable, the first as the chatbot developer, the latter as the owner of the Cloud' technical infrastructures - detective product: lack of age confirmation and reporting mechanisms, no need to use the expressions used by the chatbot as defective elements)

- **Florida State University mass shooting** (April 2025) Chatbot logs showed ChatGPT giving advice on making the gun operational, shortly before shooting. Investigation was opened against OpenAI.
- **Belgium suicide** (2023) after a six-week correspondence with a chatbot named Eliza on the application Chai.

Big wave of lawsuits against OpenAI, Character.AI for chatbots that encourage self-harm, promote self addiction etc, especially to minors and vulnerable people.

Big companies' reaction: Implement security measures - respond to compensation requests to reduce reputation damage and financial and commercial loses



B) Deep fakes/misinformation

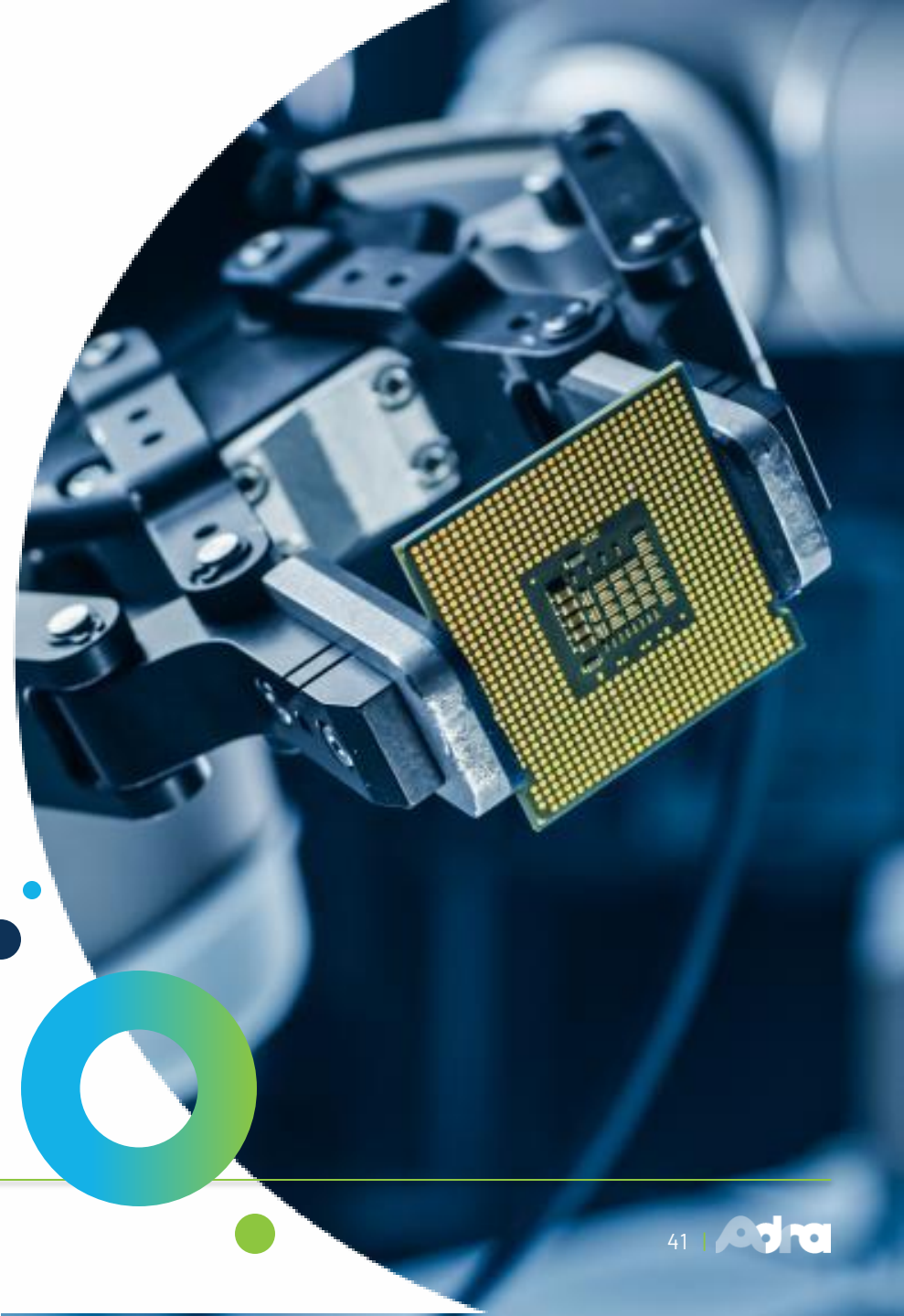
- **Art. 3 AI Act:** 'deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful
 - The AI Act does not prohibit deepfakes, directions on transparency for AI-generated content
 - Other legal basis for their regulation: GDPR, Digital Services Act, IP legislation, sectorial rules covering for instance press, national legislation (penal, civil)

Examples:

- AI-generated video, voice cloning, image AI-generated
- AI-generated evidence in front of courts posing significant challenges to judges and attorneys to detect and authenticate evidence (dual challenge: AI-generated evidence is presented as real and real evidence is often challenged to be AI-generated)
- Chatbots: False and inaccurate information, deliberately or accidentally, while LLMs outputs have the form and the confidence of knowledge, without sharing links or the processes that led to this knowledge – LLMs create the illusion of trustworthy information

- **Google AI Overviews** (June 2026 – Regional Court of Munich) According to the court, the AI had invented connections that appeared in none of the linked sources, mixing the claimants with genuinely shady firms. The claimants (publishers's company) sent a cease-and-desist letter, and Google did not respond adequately. What is the interesting in that case?

Search engines have long had limited liability because they merely point to third-party pages. AI Overviews do something different: they generate "independent, new, and substantive statements" in Google's own words, so Google "alone has influence" over them and owns what they produce, thus Google is liable for the false statements in its AI Overviews outcomes.

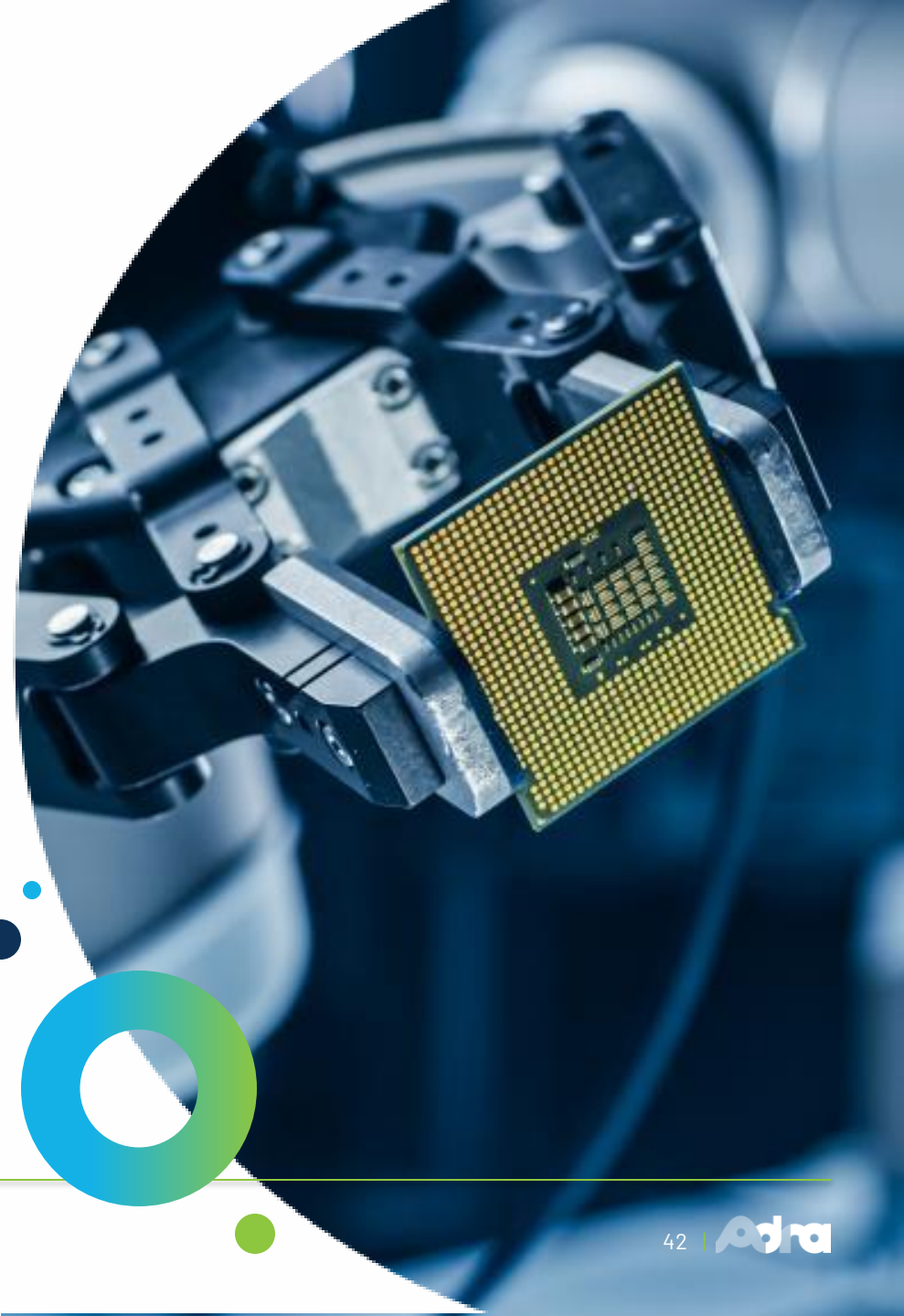


Ways to mitigate risks 1/2

EC **Code of Practice on Transparency of AI-generated Content** (issued in June 2026 – invitation to producers and deployers to sign it and thus prove compliance with the relevant provisions of the AI Act)

Producers: It sets out how to ensure that AI-generated or AI-manipulated audio, images, video or text are marked in a **machine-readable** way and can be detected as artificially generated or manipulated.

Deployers: It explains how they must clearly label deepfakes and AI-generated or AI-manipulated text published to inform the public on matters of public interest when there has been **no human review or editorial control**.

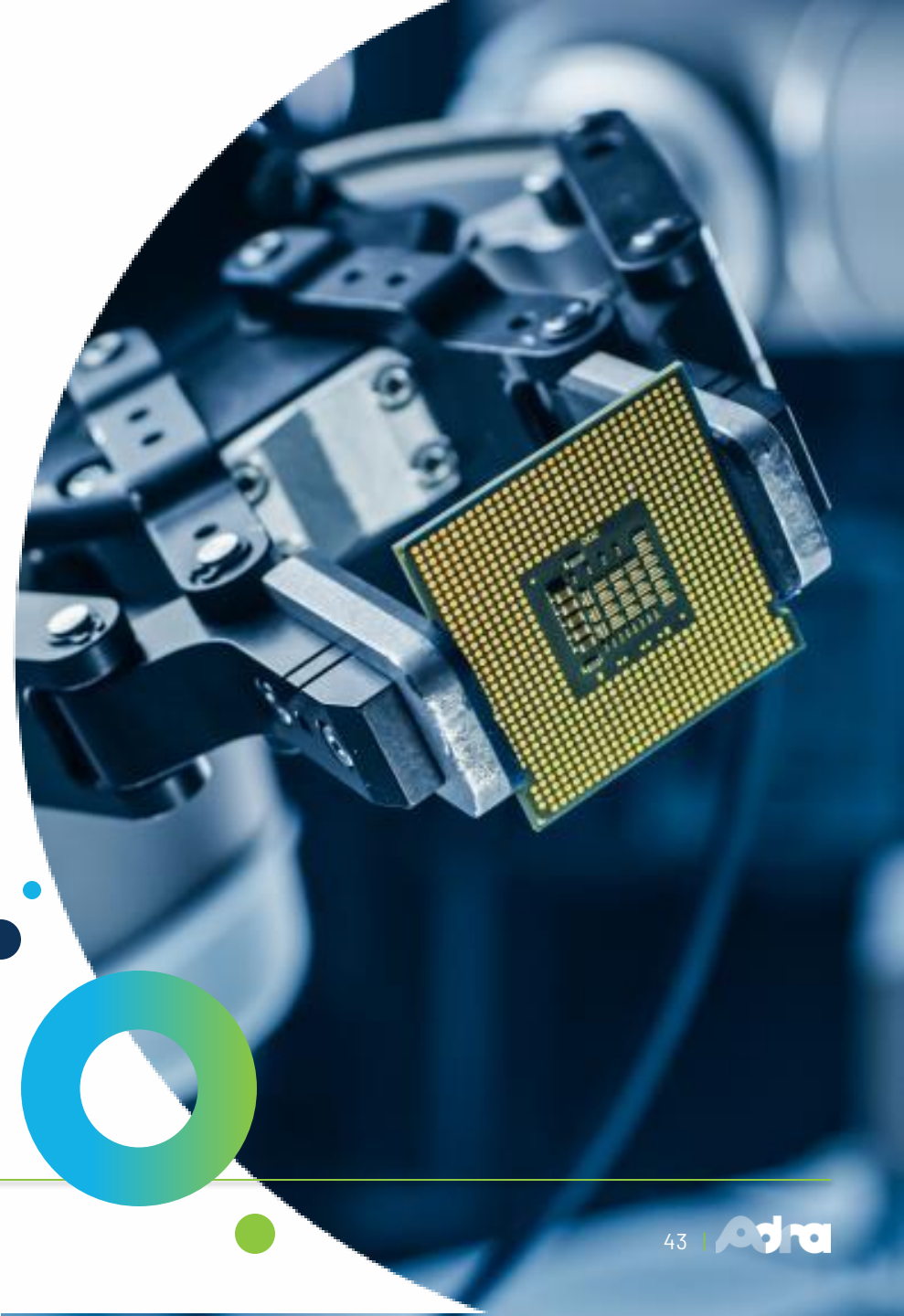


Ways to mitigate risks 2/2

Using the Digital Services Act as a legal basis

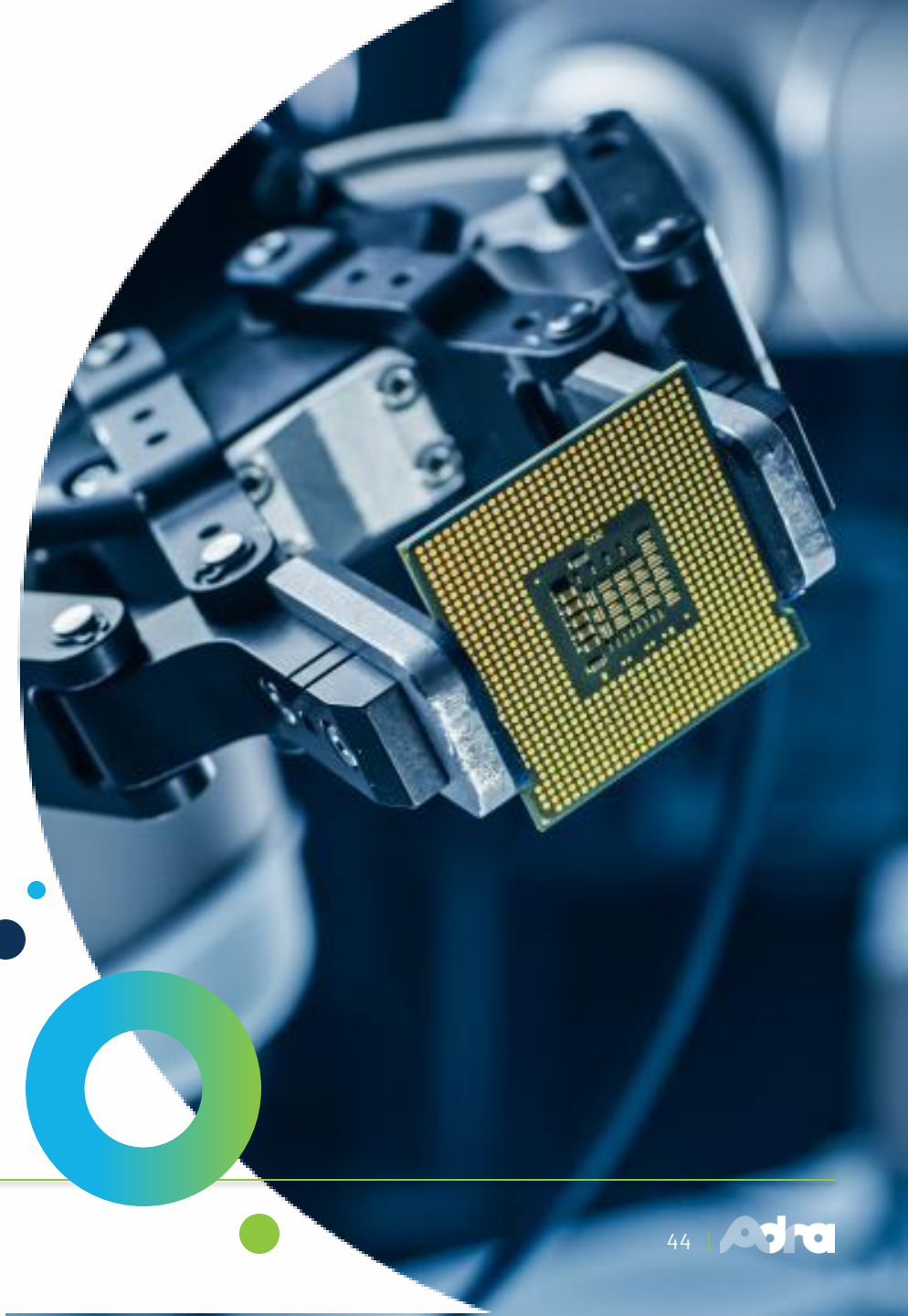
The "Grok" Investigation: Following a global outcry over the X (formerly Twitter) AI chatbot generating nonconsensual intimate images, the European Commission has launched recently an ongoing formal investigation under the Digital Services Act. European regulators are scrutinizing X for failing to prevent the dissemination of illegal deepfakes and are demanding strict compliance.

Takedown Rights: Under the DSA, users now have explicit rights to flag and demand the removal of deepfakes on major platforms, with independent out-of-court dispute mechanisms available to challenge wrongful content moderation decisions.



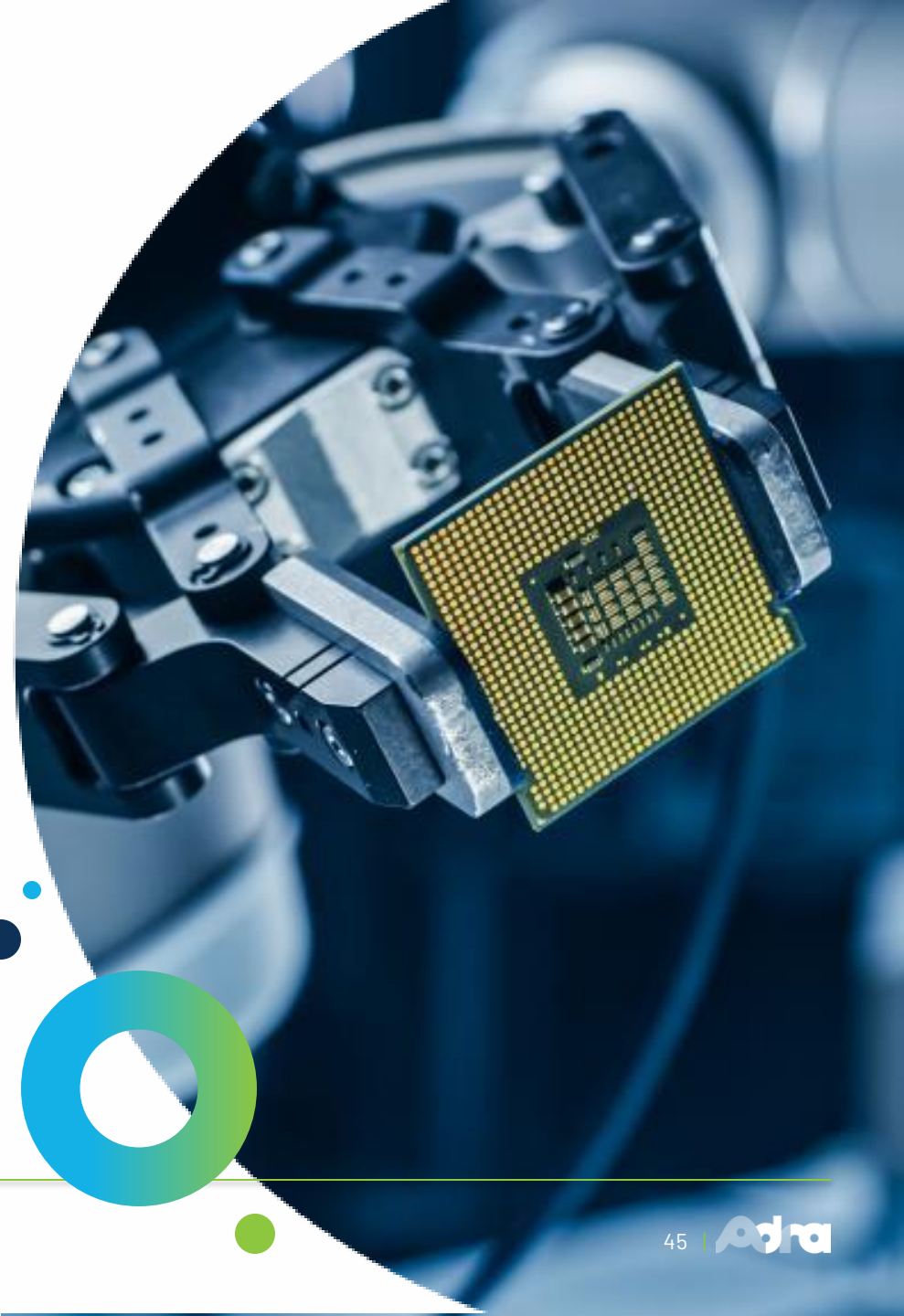
Industrial Automation: Cobots and harms

- Collaborative robots **share physical workspace** with human workers, operating without fixed barriers;
- Global **cobot market** estimated at \$1.2 billion in 2023, projected to exceed \$10 billion by 2030. Over 2.4 million industrial robots in use worldwide. 61 **robot-related deaths** in the US (1992- 2015), expected to increase with **cobots, exoskeletons, and autonomous vehicles** into work environments.
- **Robot-related accident data** remains **under-reported**: only severe injuries are captured (hospitalisation or amputation);
- **Cobot harm involves a multi-party supply chain** (hardware manufacturer, software vendor, system integrator, employer);
- **Physical and non-physical harm**: contact injuries from unexpected movement, musculoskeletal harm, over-reliance, psychological harm from surveillance, emotional dependence and anthropomorphisation, skill degradation.



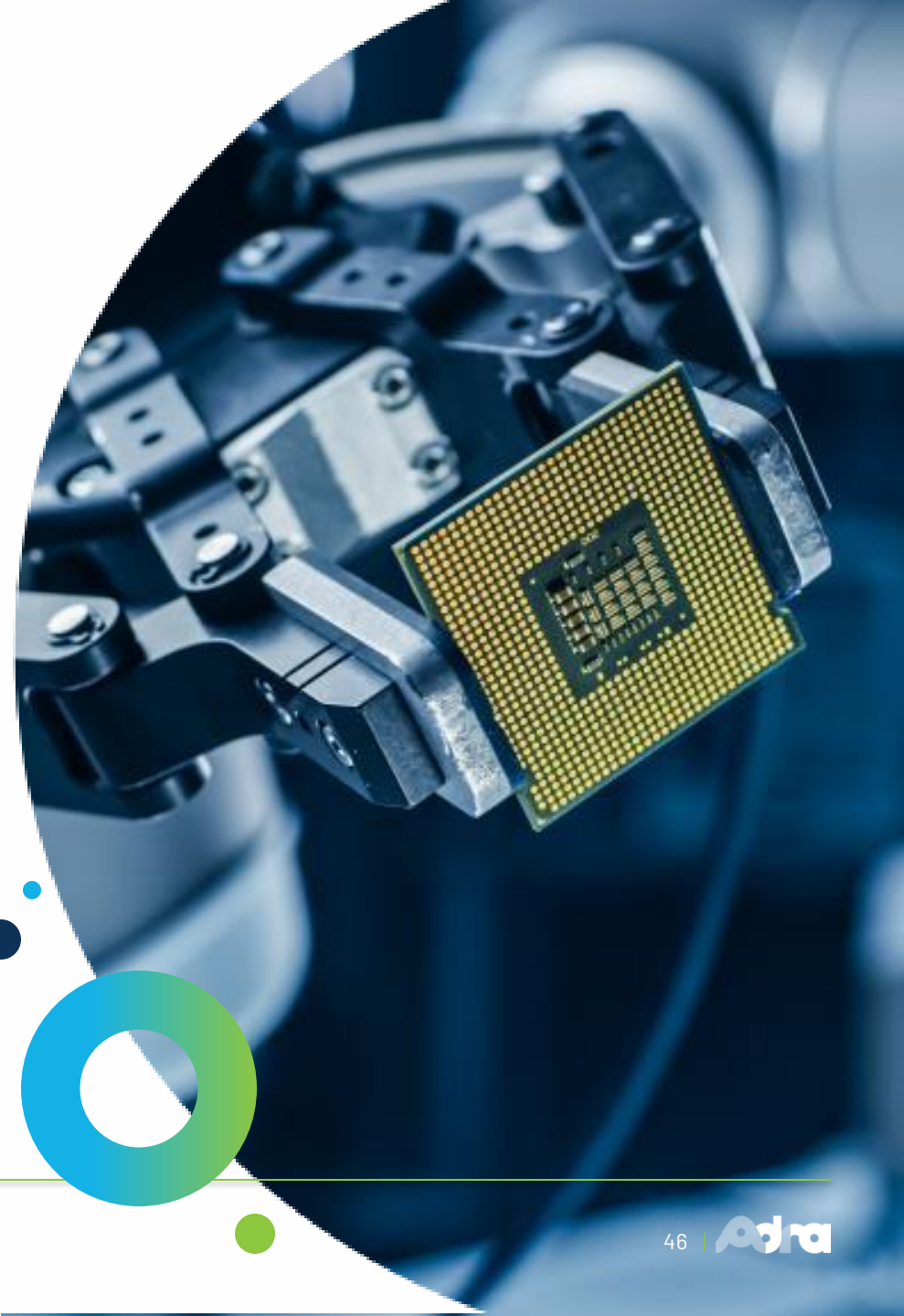
The Regulatory Framework / 1

- **EU Machinery Regulation (EU) 2023/1230** (applicable from January 2027), replacing the Machinery Directive 2006/42/EC; explicitly extends **mandatory safety requirements** to AI-driven cobots, autonomous mobile machinery, and self-evolving software safety components; requires **third-party conformity assessment** by a Notified Body for AI safety components; mandates **cybersecurity** as a machine safety obligation; requires tamper-evidence logs and documentation of **software modifications**; imposes continuous **lifecycle safety assurance** for adaptive systems.
- **EU AI Act**: AI systems performing safety functions in machinery are **high-risk** under Annex III; they face a dedicated **conformity assessment route** that interacts with the Machinery Regulation; Art. 14 requires effective **human oversight**; Art. 72 requires **post-market surveillance**



The Regulatory Framework / 2

- **Revised Product Liability Directive** (EU) 2024/2853 (applicable from December 2026): **cobots** are products; **embedded AI software** is a product component; **post-market OTA updates** that introduce defects trigger liability; **data corruption** is compensable damage; **psychological harm** “of medical significance” is compensable above a minimum threshold; **rebuttable presumptions** of defect and causality assist injured workers; **mandatory disclosure** of technical documentation can be ordered by courts.
- **Occupational health and safety law: D.Lgs. 81/2008** (Italy) and the **EU OSH Framework Directive** impose duties on employers to assess and manage **cobot risks**, including both physical and psychosocial risks; to provide **safe equipment** and a **safe working environment**; the employer may seek contribution from the **manufacturer** where a product defect was the primary cause of harm



Pending Judicial Cases on Cobot-Related Injuries

No significant **published judicial decisions** specifically addressing liability for cobot injuries (recent technology, incidents typically resolved via workers' compensation or settlement)

Closest analogous cases involve traditional industrial robots:

- **Pyeongtaek, South Korea (2022)** – worker crushed by a robot near a conveyor belt at a milk factory; no published court ruling.
- **Volkswagen Baunatal, Germany (2015)** – worker killed by a robot arm during installation; prosecutors closed the case without indicting any individual.

The absence of cobot-specific precedent leaves open how **courts** will allocate fault between manufacturer, integrator, and operator under shared human-robot workspaces



Expert Roundtable & Panel



Sergiy Barbashyn

President, AI Ethics and Integrity International Association (AIEI)



Alexei Grinbaum

AIOLIA Coordinator, Senior research scientist at CEA-Saclay



Katrin Saaremael

Legal and Policy Officer, DG GROW
(European Commission)

Moderated by:



Argyro Amidi

Ethical and Legal ADRA Topic Group Co-Chair

Q&A and Wrap-up

Q&A?

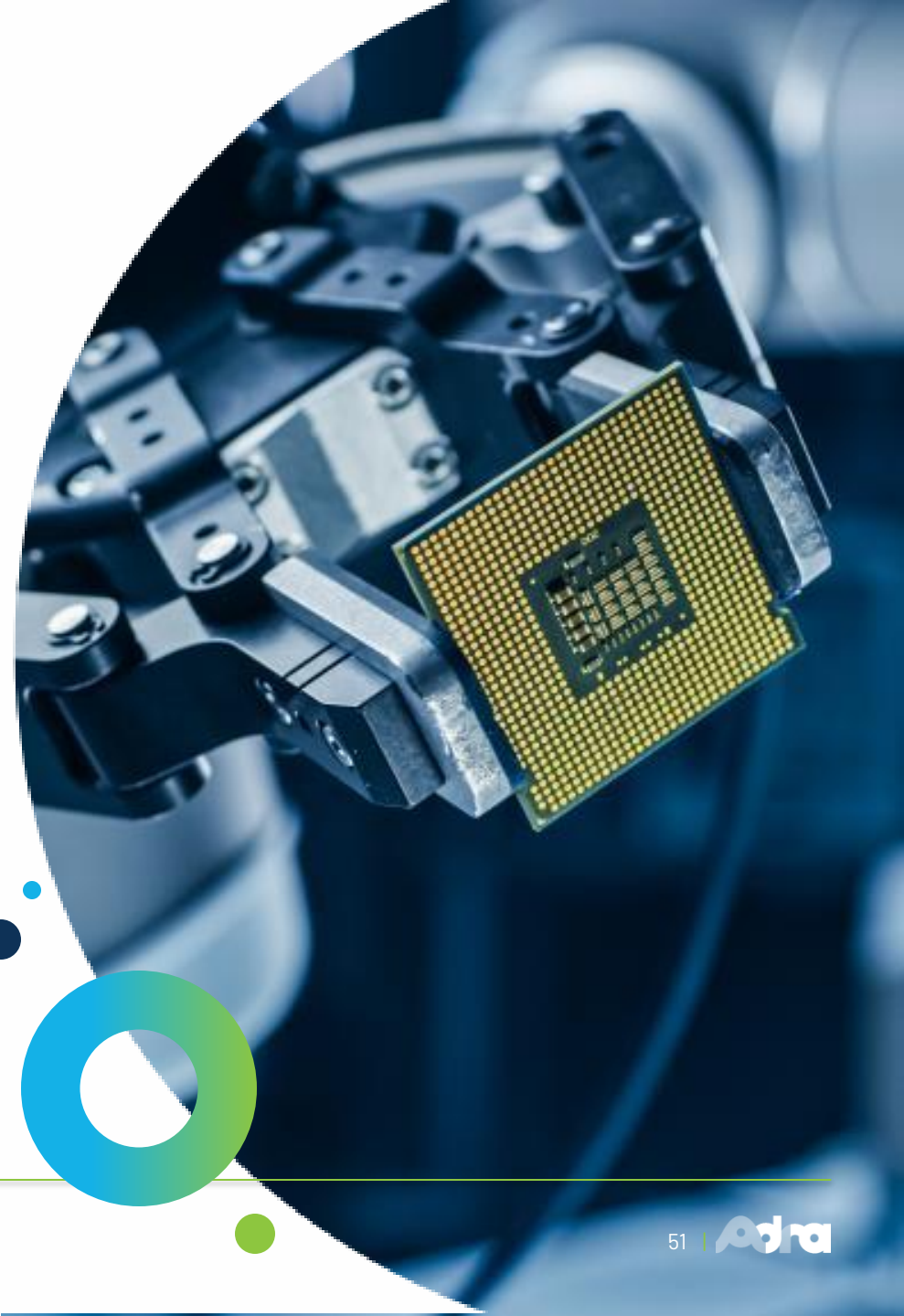
What the **next webinar**
should tackle?

Give us your feedback!



Wrap up

- Litigation for damage caused by AI systems will **increase** (starting from U.S. already)
- EU Member States competent to address this challenging issue due to the lack of EU common civil law and harmonized procedural rules
- Risk for fragmentation among EU national jurisdictions – legal uncertainty that does not facilitate innovation
- Big challenge: how to prove defect for the defective product and causation due to the black-box inherent character of AI systems, their complexity and the multi-stakeholders involved for their development and deployment
- **Revised Product Liability Directive** (including software, ensuring an EU-based liable party, information obligations of the defendant that, should not covered, alleviate the burden of proof concerning the defect of the product, covering psychological damage and data loss as compensable damage)
- Preferable a strict liability regime for high-risk AI systems, or strict sectorial liability in alignment with sectorial legislation, such as the Medical Devices rules
- **Still a big challenge:** How to regulate in a holistic way Liability of LLMs and GPAI



Thank you.



The AI Data Robotics
Association



Adra Association

Rue Belliard 40, Brussels 1040
Belgium

+32 2 786 30 36

www.adr-association.eu

