



**PDF Days
Europe
2025**

AI and Digital Documents

Opportunities and Regulatory Challenges

Prof. Dr. Philipp Hacker, LL.M. (Yale)

Chair for Law and Ethics of the Digital Society

European New School of Digital Studies

Structure

- I. AI – A Loose Concept
- II. AI in Document Workflows
- III. Efficiency vs. Risk
- IV. Regulatory Focus
- V. Compliance Strategies

Overview based on

- Philipp Hacker, Andreas Engel, Marco Mauer, **Regulating ChatGPT and other Large Generative Models**, (2023) ACM Conference on Fairness, Accountability and Transparency (FAccT '23) 1112-1123, http://arxiv.org/a/hacker_p_1
- Philipp Hacker & Matthias Holweg, **The Regulation of Fine-Tuning**, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5289125

Part I.

AI – A Loose Concept

From Rules to Learning Systems

- Rule-based AI vs. Learning Systems
- Machine Learning: from examples

Definition: A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

Mitchell, Machine Learning, 1997, 2

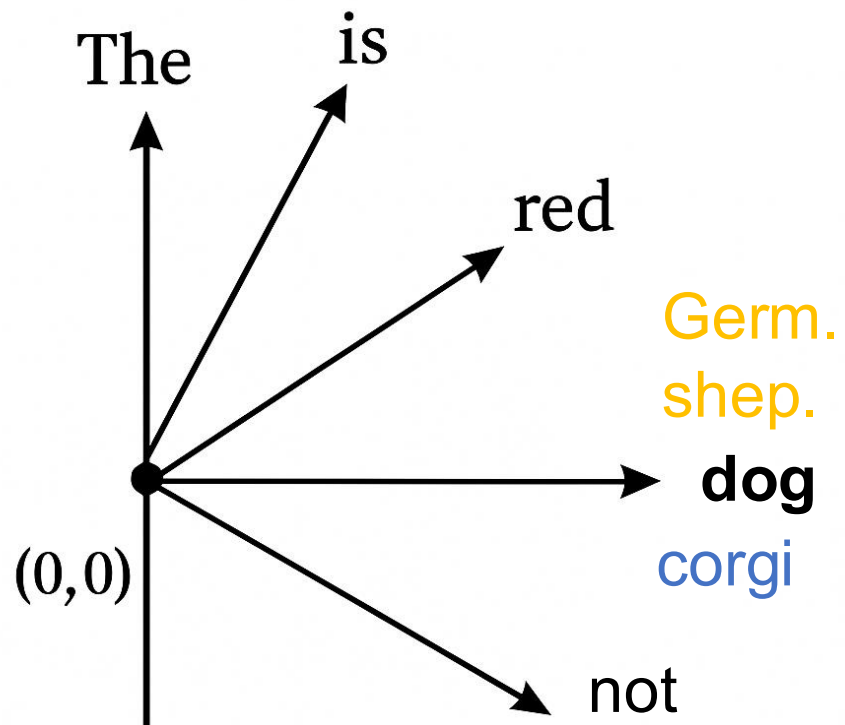
- Generative AI:
 - Output: structured sequences
- Today: often hybrid systems encompassing both rules and AI (including symbolic AI)



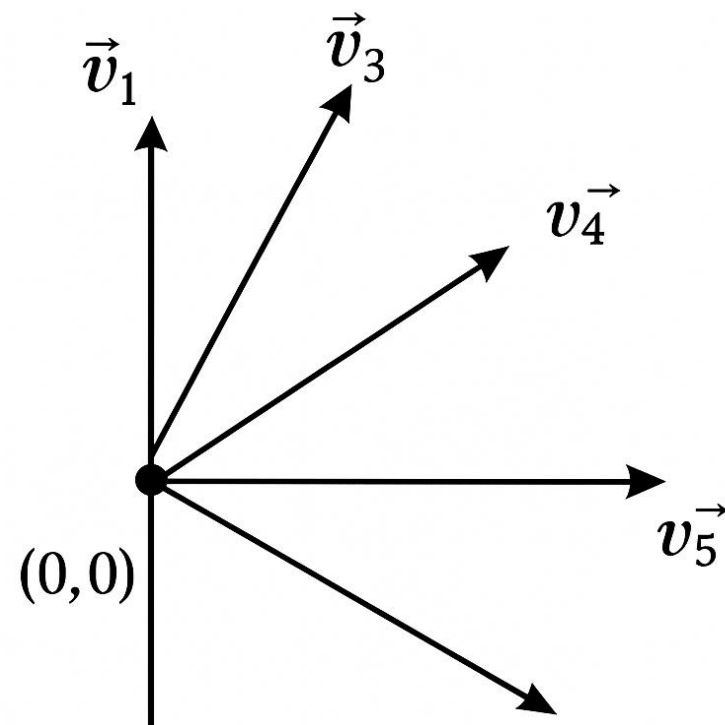
Inside a Large Language Model

- LLMs: deep neural networks trained to predict the next token in a sequence → text generation
- Tokens: mapped to vector space (embeddings for meaning and position)
 - Similar meaning → similar vector

The dog is not red



The dog is not red



Inside a Large Language Model

- LLMs: deep neural networks trained to predict the next token in a sequence → text generation.
- Tokens: mapped vector space (embeddings for meaning and position)
 - Similar meaning → similar vector
- **Attention mechanism: analyzing vectors to understand tokens in context**

How does GenAI work?

I went to the yellow **bank**, and ...



... put my feet in the sand
by the river.



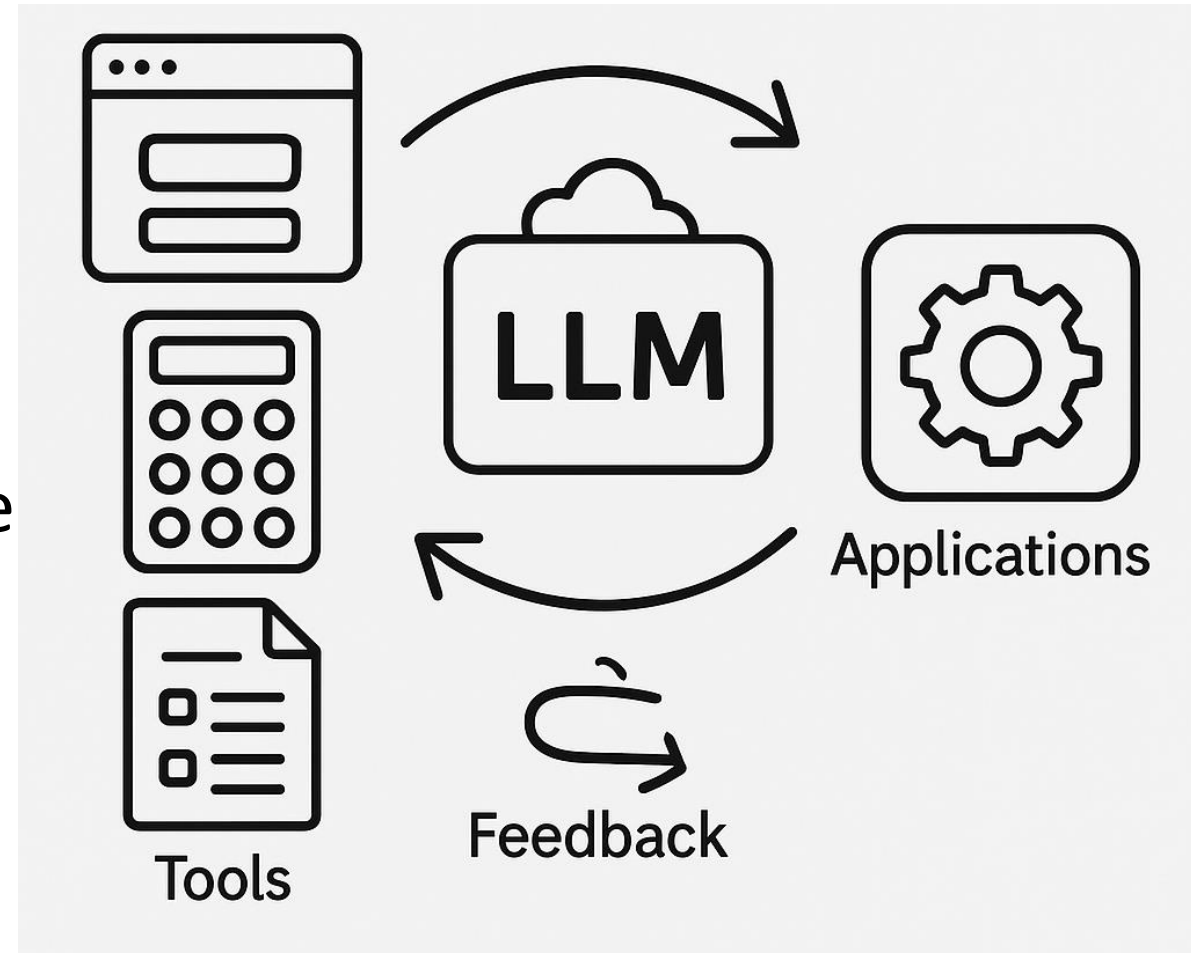
...withdrew some money.

Inside a Large Language Model

- LLMs: deep neural networks trained to predict the next token in a sequence → text generation.
- Tokens: mapped to vectors (embeddings)
 - Similar meaning → similar vector
- Attention mechanism: analyzing vectors to understand tokens **in context**
- **Output token (word unit, pixel, DNA unit):** chosen by computing probability distributions over possible continuations
 - Sampling and instruction tuning for variability and performance

AI Agents and Agentic AI

- "Action-Taking LLMs"
- Integrated into **systems** via, e.g., interface
- Tool use: web, calculators, forms
- Feedback loop
 - Outputs from tools fed back into the model, enabling **multi-step workflows** and adaptive behavior
- Applications: task automation, personal assistants, research agents, autonomous software control, ...



Part II.
**AI in Document
Workflows**

AI in Document Workflows – Overview

- Automation: drafting, data extraction, analysis
 - Particularly for unstructured data
- Processing at scale with NLP/OCR
- Analytical insights for compliance and contracts
- Potential: Efficiency & consistency across workflows
 - But: hallucinations, GIGO → needs oversight

AI Application – Fraud Detection

- Detect anomalies in documents/transactions
- Real-time prevention in banking & insurance
- Automates KYC & compliance checks
- But challenge: balancing false positives and negatives



AI Application – Automated Content Classification

- Categorize & label documents
 - OCR + NLP/LLM + KNN
- Automate compliance tagging
- Speeds retrieval & retention management
- But: Risk of misclassification – need governance



Part III.
Efficiency vs. Risk

Efficiency Gains from AI

- 50%+ reduction in document handling times in some use cases
- Potentially faster verification cycles
- Lower error rates – improved accuracy
 - Particularly for traditional AI (non-generative)
- Potential Productivity and ROI benefits

Efficiency vs. Risk – The Trade-offs

- Semantic misinterpretation risks
- Reliability & hallucinations

Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity

Joel Becker*, Nate Rush*, Beth Barnes, David Rein

Model Evaluation & Threat Research (METR)

Abstract

Despite widespread adoption, the impact of AI tools on software development in the wild remains understudied. We conduct a randomized controlled trial (RCT) to understand how AI tools at the February–June 2025 frontier affect the productivity of experienced open-source developers. 16 developers with moderate AI experience complete 246 tasks in mature projects on which they have an average of 5 years of prior experience. Each task is randomly assigned to allow or disallow usage of early-2025 AI tools. When AI tools are allowed, developers primarily use Cursor Pro, a popular code editor, and Claude 3.5/3.7 Sonnet. Be-

Surprisingly, we find that allowing AI actually increases completion time by 19%--AI tooling slowed developers down.

Efficiency vs. Risk – The Trade-offs

- Semantic misinterpretation risks
- Reliability & hallucinations
 - Human oversight essential
 - Addressing bias & fairness

No cover
image
available

The Oxford Handbook of the Foundations and Regulation of Generative AI

(In Progress)

Philipp Hacker (ed.) et al.

Search in this book



Contents

Introduction to the Foundations and Regulation of Generative AI

Explaining and Interpreting Generative AI

The Challenges of Agentic AI Safety

Building GenAI Benchmarks: A Case Study in Legal Applications

Generative AI and the Problem of

CHAPTER

Generative Discrimination: What Happens When Generative AI Exhibits Bias, and What Can Be Done About It

Philipp Hacker, Frederik Zuiderveen Borgesius, Brent Mittelstadt, Sandra Wachter


<https://doi.org/10.1093/oxfordhb/9780198940272.013.0016>

Published: 22 April 2025





PDF

 Split View

 Annotate

 Cite

 Permissions

 Share ▼

Abstract

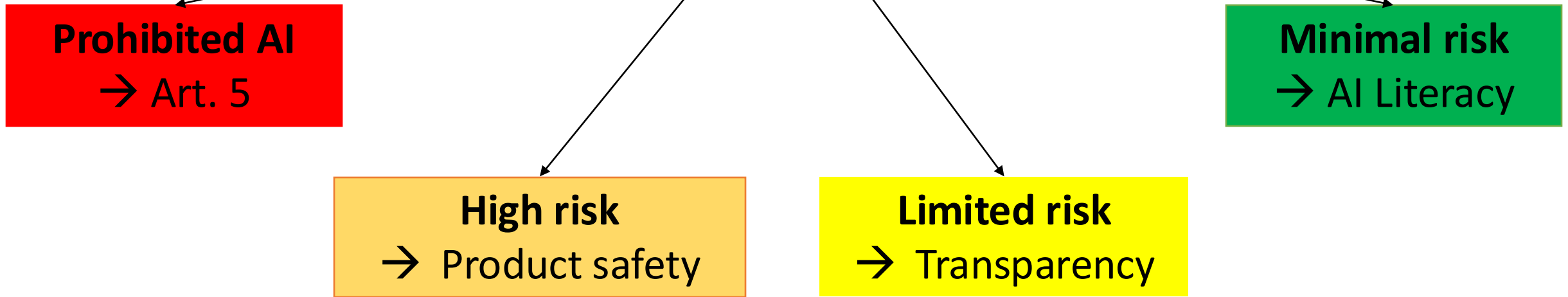
As generative Artificial Intelligence (genAI) is increasingly used across sectors, its potential for societal benefit is paired with risks of discrimination. This chapter explores how genAI challenges non-discrimination law, identifying two primary types of discriminatory outputs: (i) demeaning and abusive content; and (ii) subtler biases from inadequate representation of protected groups. The latter includes genAI output that, while not discriminatory in a single instance, has discriminatory effects over time. For example, a genAI system may predominantly display white men when asked for examples of people in important jobs. The chapter examines the resources of existing EU law in addressing such cases and demonstrates how traditional legal categories, such as direct and indirect discrimination and harassment, are sometimes inadequate for genAI. The final part also offers suggestions on updating EU laws and mitigating biases preemptively in training and input data.

Part IV.

Regulatory Focus

Structure of the AI Act

Four Risk Levels

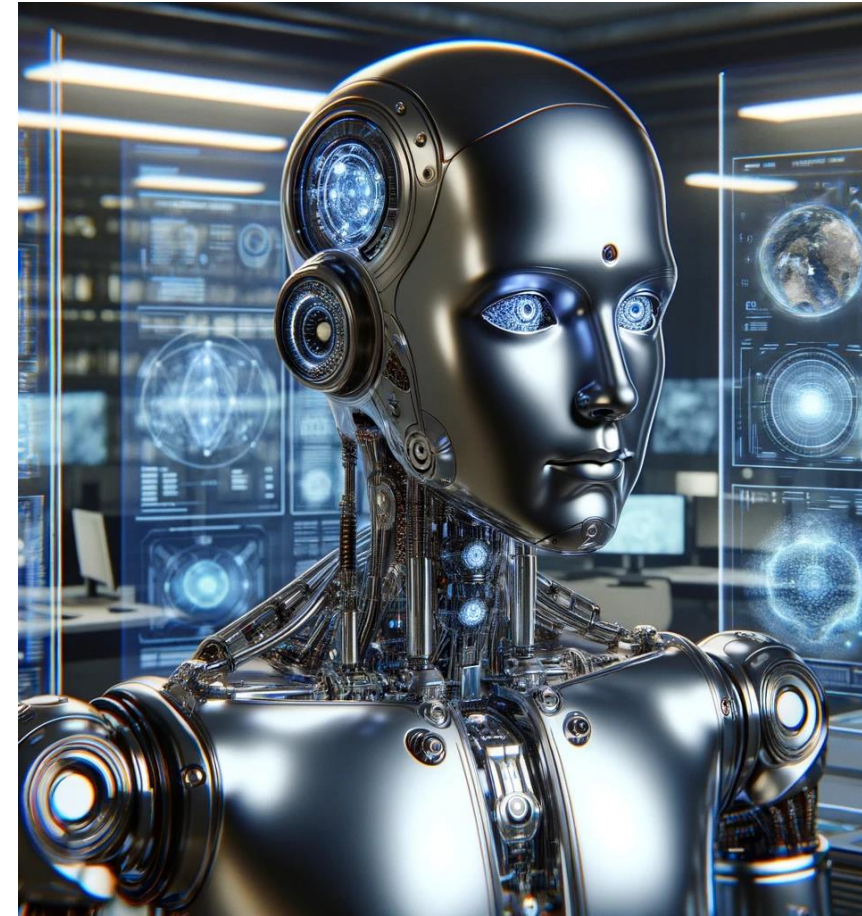


General-purpose AI models

AI Act: Definition of AI

- Art. 3(1): **machine-based** system
 - with varying degrees of **autonomy** and that may show adaptiveness after deployment,
 - that **infers**
 - from the **input** it receives,
 - how to generate **outputs**
 - such as predictions, content, recommendations or decisions,
 - that can **influence** physical or virtual environments.

→ AI = machine learning, **broadly understood**

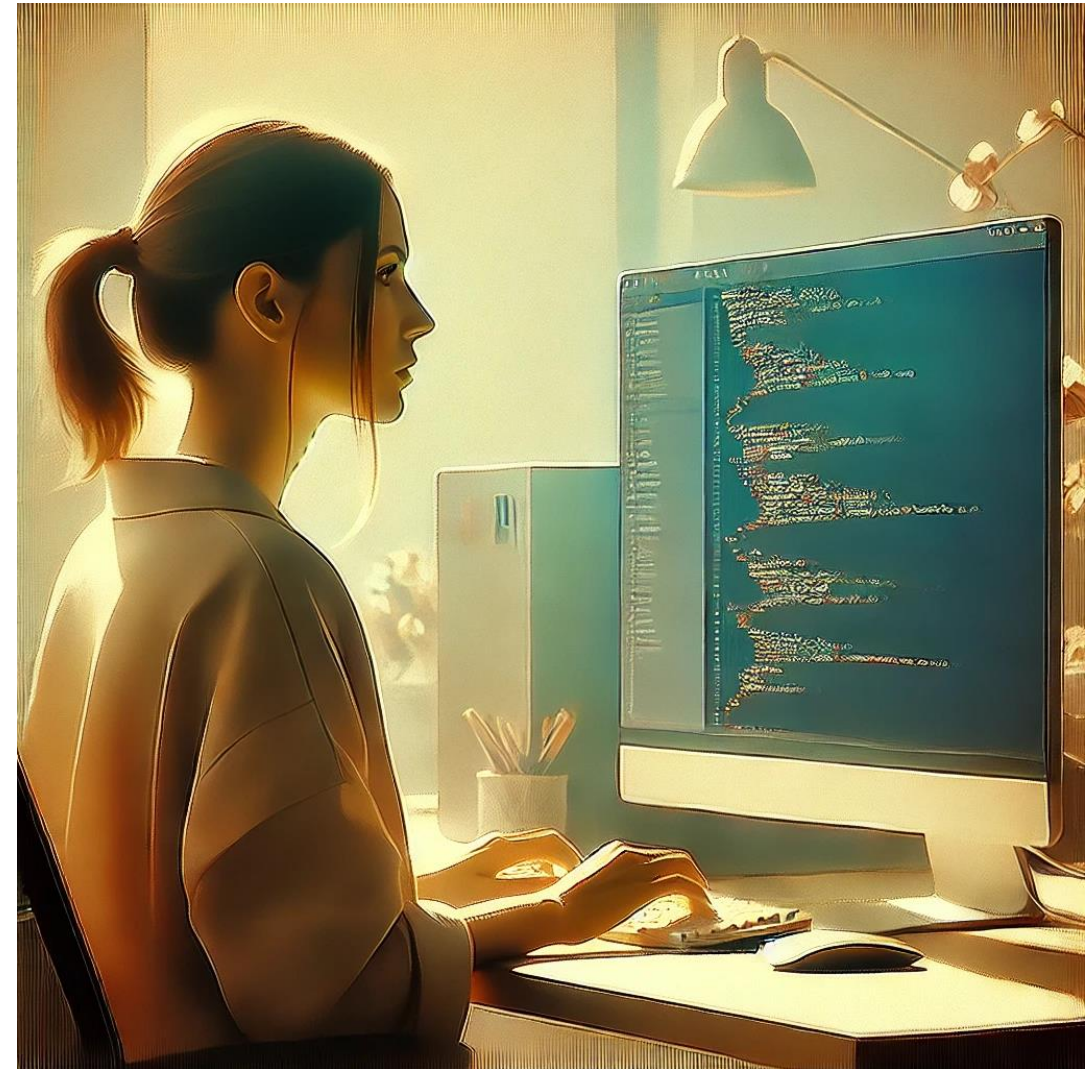


DALL-E 3, "picture of AI, photorealistic", 01/31/2024

Rules for providers

Art. 3(3):

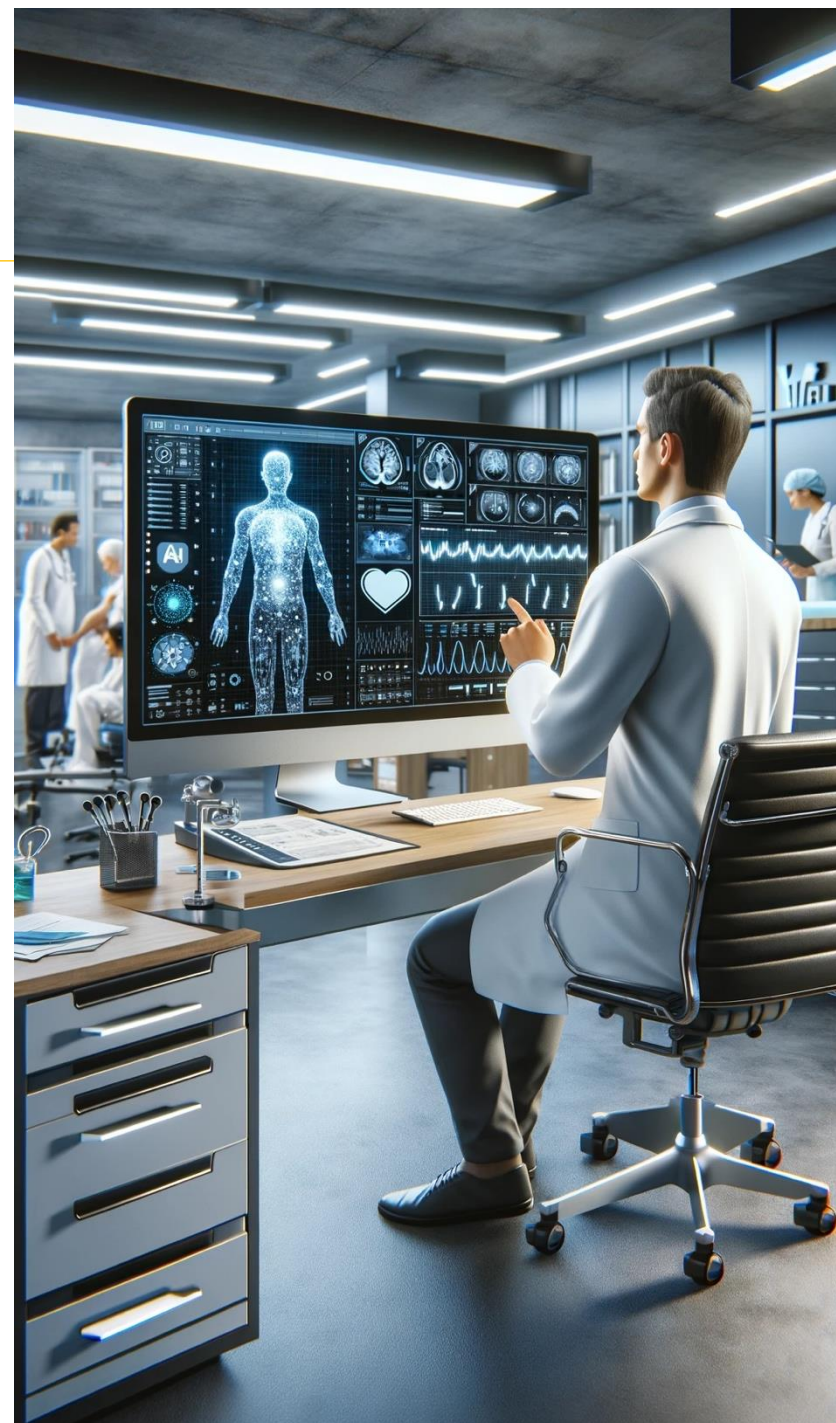
- Entity who
 - **develops** or
 - **has developed** [i.e., lets a third party develop]
- an AI system or a general-purpose AI (GPAI) model and
- places it on the market or puts the AI system **into service** under its **own** name or **trademark**



Rules for deployers

Art. 3(4): deployer

- Entity who **uses an AI system** under its authority
 - Exception: **consumers**



GPAIs - The tiered approach

All GPAIs, Art. 53

- Transparency & copyright

A
I

S
A
F
E
T
Y

GPAIs with Systemic Risks

In addition, Art. 55:

- Risk management
- Red Teaming
- Incident Reporting
- Cybersecurity

Systemic risks:

Assumption: 10^{25} FLOPs

→ approx. GPT-4

Or: Commission Designation

- Compute
- Data
- Effects

Fine-tuning?

EU AI Act – High-Risk AI Obligations

Provider rules:

- Pre-market assessment & testing
- Ongoing risk management
- Data quality & documentation
- Conformity assessment & CE marking

Deployer rules:

- Monitoring

High-Risk AI Systems

(Annexes I A, III AI Act):

- Facial recognition/biometrics
- Employment
- Medical AI
- Credit scoring
 - **Except fraud recognition**
- Insurance (life, health)

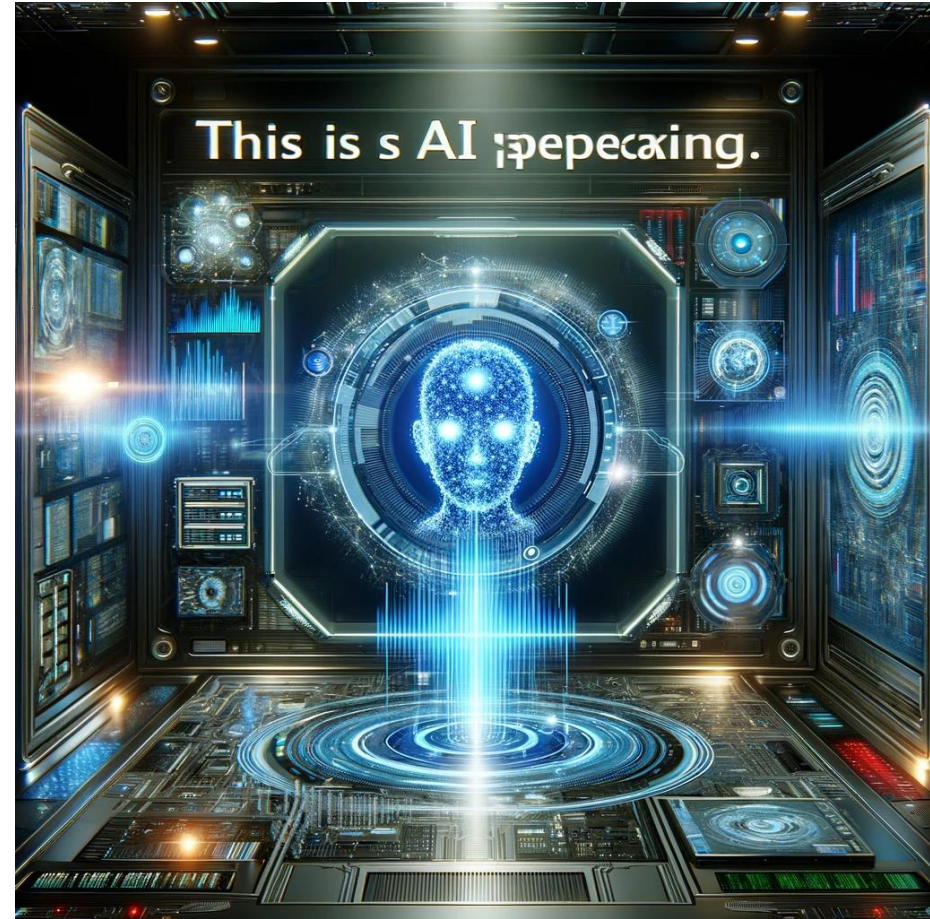
Not included:

- **Generative AI**
 - E-commerce
(recommender systems)
 - Search engines
- Partially in the Digital Markets Act (DMA) & Digital Services Act (DSA)

AI Act: Chatbots

Art. 50: **Disclosure** when AI interacts with humans

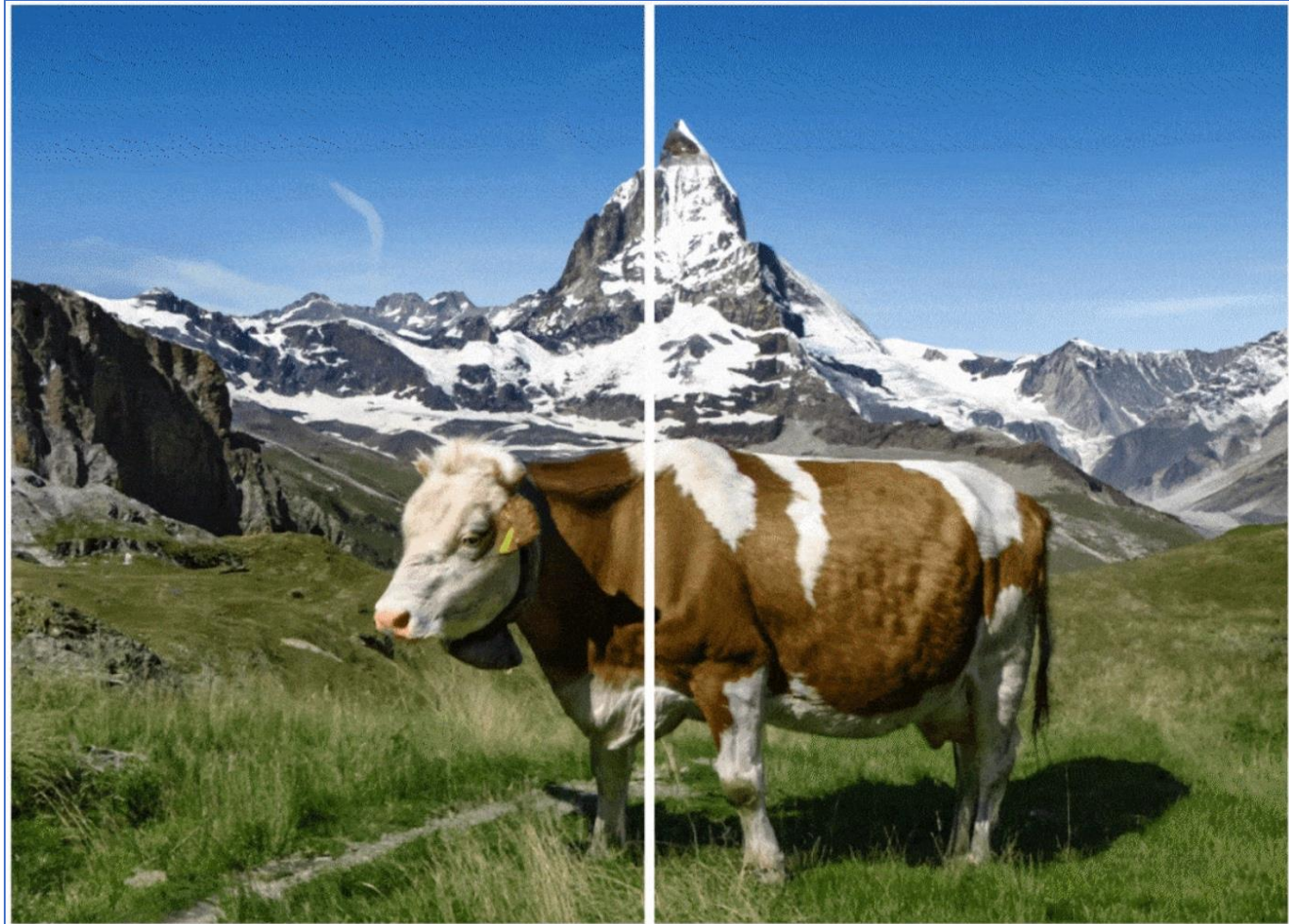
- Also for **deep fakes**
 - **AI-generated or manipulated** image, audio or video **content**
 - that **resembles** existing persons, objects, places, entities or events
 - and would **falsely appear** to a person to be **authentic** or truthful
- For **AI text production**: only for text to inform the public on matters of public interest
 - Exception: **editorial review** and assumption of responsibility



DALL-E 3,
"This is AI
speaking",
11/13/23

AI Act: Chatbots

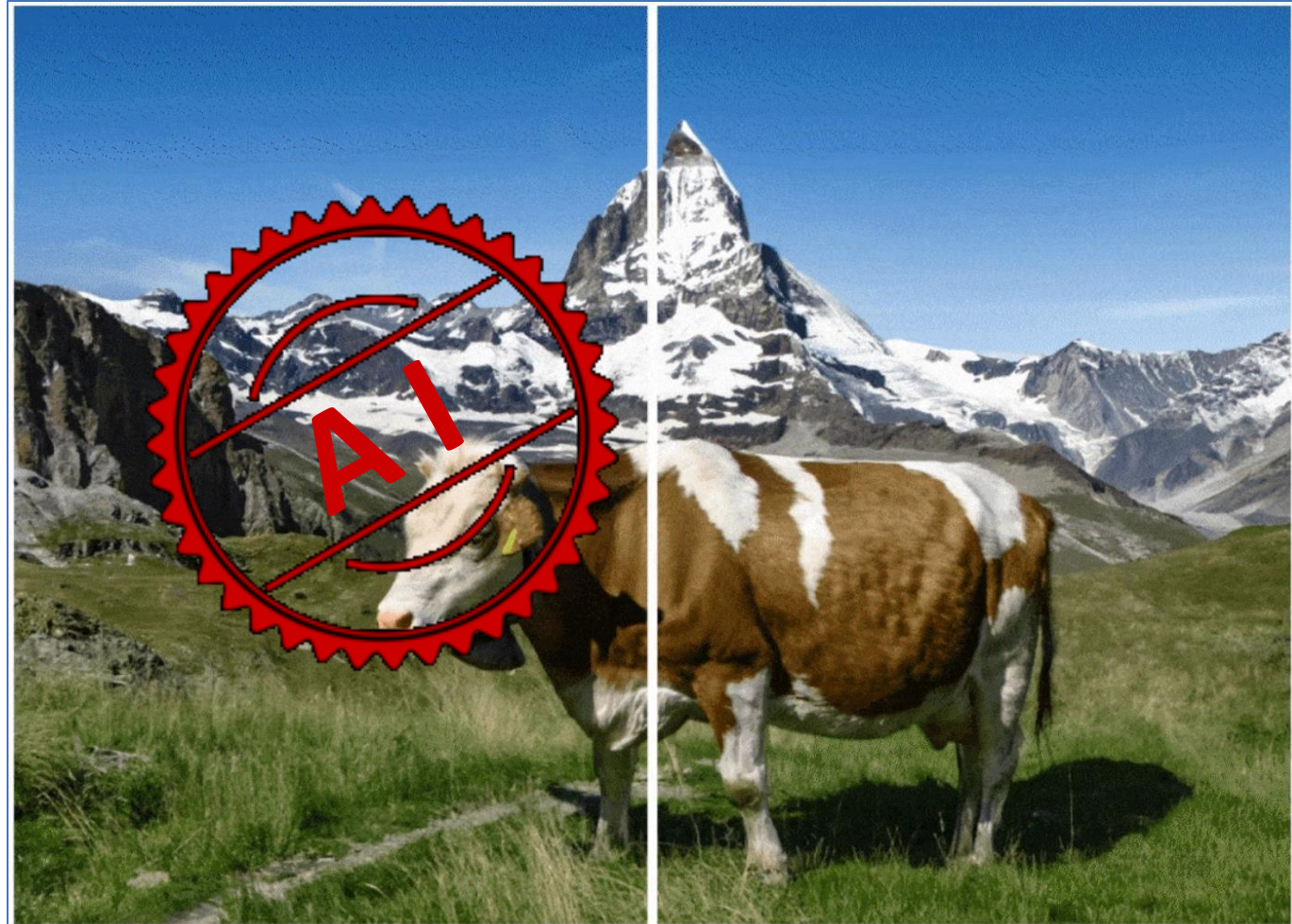
- Art. 50: **Labeling** of AI-generated content (e.g. watermarks)
→ Fraud detection



Melissa Heikkilä, Google DeepMind has launched a watermarking tool for AI-generated images, MIT Technology Review (Aug. 29, 2023)

AI Act: Chatbots

- Art. 50: **Labeling** of AI-generated content (e.g. watermarks)
→ Fraud detection



Watermarked

Non-watermarked

Melissa Heikkilä, Google DeepMind has launched a watermarking tool for AI-generated images, MIT Technology Review (Aug. 29, 2023)

Part V.

Compliance

Standards and Code of Practice

Technical standards for High-Risk AI:

- Presumption of conformity
- But published only in 2026

Code of Practice for GPAI:

- Presumption of conformity
- Published in July 2025

Managing Regulatory Risk – Best Practices

- Governance or ethics councils & oversight
- Tool classification & control
- Documentation & transparency
- Human-in-the-loop & audits

Conclusion

The future will be exciting.

Thanks!



Philipp Hacker

Professor of Law & Technology at ENS
| Speaker | Consultant

