

# What AI developers need to know about AI governance

Kacper Łodzikowski  
Vice President, AI Capabilities

ML in PL  
28 Oct 2023



Pearson

# Governance is one of top 3 challenges for AI adoption

- Examples:
  - Trivago fined \$45M (model promoted booking site offers instead of best rates)
  - Dutch tax authorities fined €2.75M (biased fraud detection model)
  - Deliveroo fined €2.5M (blackbox model for rider management)
- Some AI developers lack proper legal support
- Others lack basic legal training
- Landscape evolving over past year



# Health warning

- Educational purpose:
  - not legal advice,
  - not company stance.
- No clear-cut answers. *It depends on your:*
  - case,
  - jurisdiction,
  - risk appetite.
- Find a legal partner to help you decide.

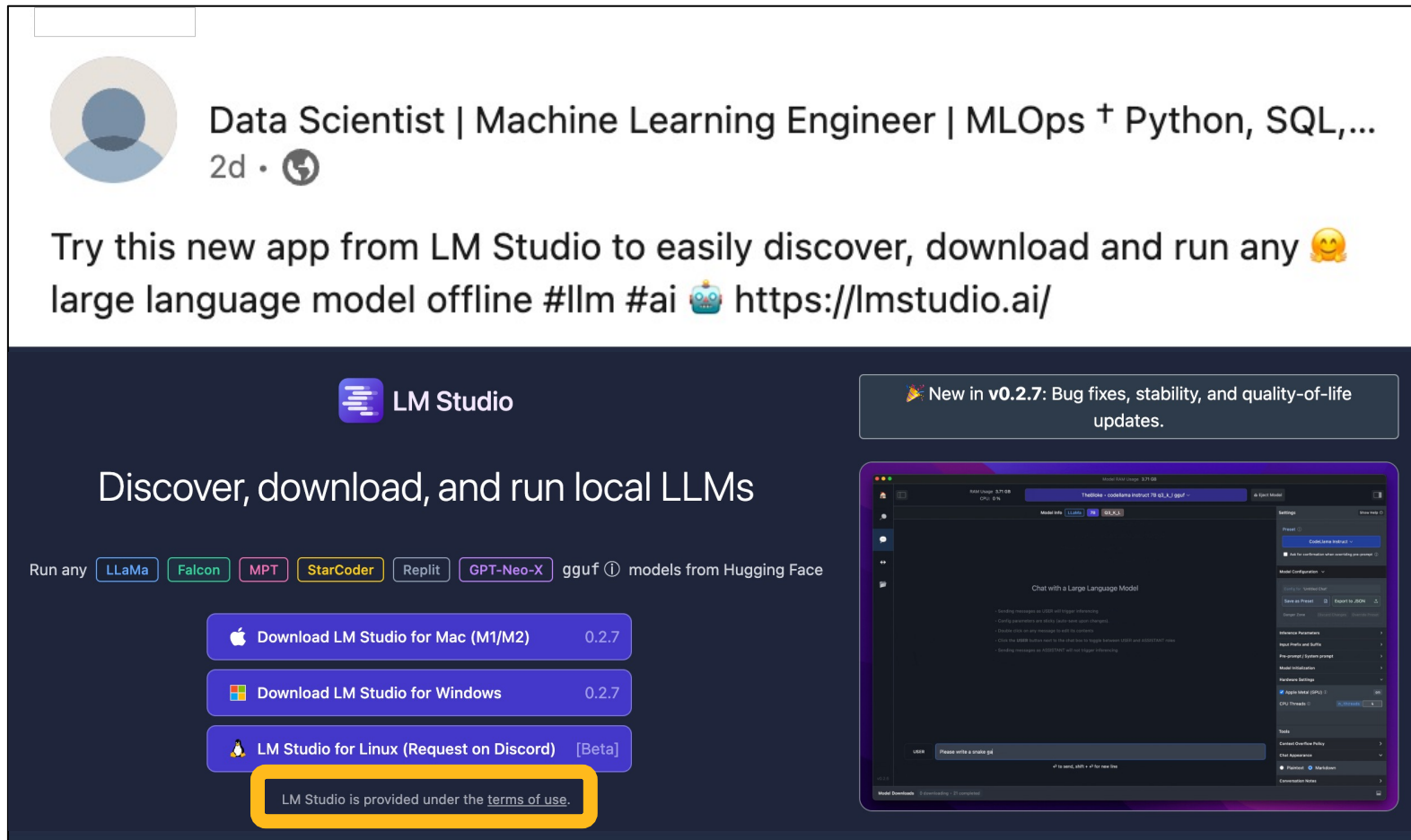




# Licences

Software, models, data

# If it's free to download, it's not always free to use



The image shows a social media post from a user with a profile picture of a blue circle. The user's bio reads "Data Scientist | Machine Learning Engineer | MLOps + Python, SQL,..." and the post is dated "2d". The text of the post says: "Try this new app from LM Studio to easily discover, download and run any large language model offline #llm #ai https://lmstudio.ai/". Below the post is a screenshot of the LM Studio application interface. The interface has a dark theme and features the LM Studio logo at the top left. A banner at the top right says "New in v0.2.7: Bug fixes, stability, and quality-of-life updates." The main heading is "Discover, download, and run local LLMs". Below this, it says "Run any LLaMa Falcon MPT StarCoder Replit GPT-Neo-X gguf models from Hugging Face". There are three download buttons: "Download LM Studio for Mac (M1/M2) 0.2.7", "Download LM Studio for Windows 0.2.7", and "LM Studio for Linux (Request on Discord) [Beta]". A yellow box highlights the text "LM Studio is provided under the terms of use." at the bottom. On the right side of the interface, there is a preview window showing a chat interface with a "Chat with a Large Language Model" title and a "Settings" panel on the right.

Data Scientist | Machine Learning Engineer | MLOps + Python, SQL,...

2d •

Try this new app from LM Studio to easily discover, download and run any large language model offline #llm #ai <https://lmstudio.ai/>

LM Studio

New in v0.2.7: Bug fixes, stability, and quality-of-life updates.

Discover, download, and run local LLMs

Run any LLaMa Falcon MPT StarCoder Replit GPT-Neo-X gguf models from Hugging Face

Download LM Studio for Mac (M1/M2) 0.2.7

Download LM Studio for Windows 0.2.7

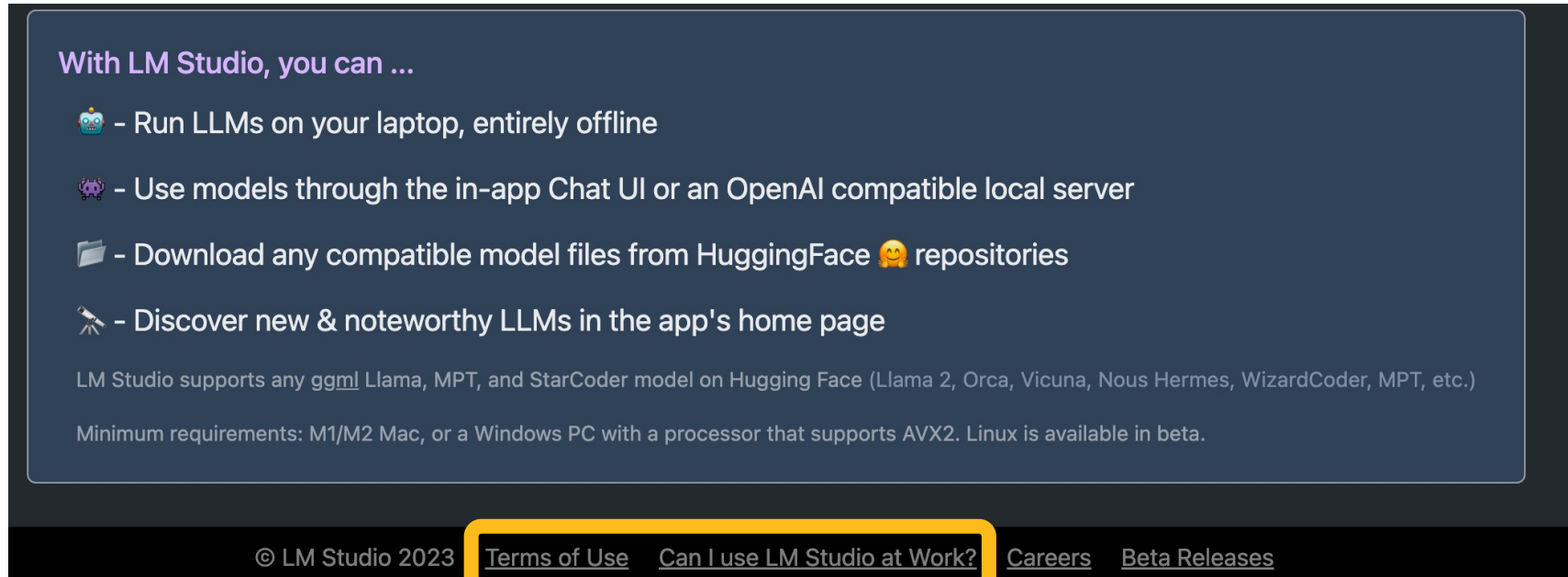
LM Studio for Linux (Request on Discord) [Beta]

LM Studio is provided under the terms of use.

Chat with a Large Language Model

Settings

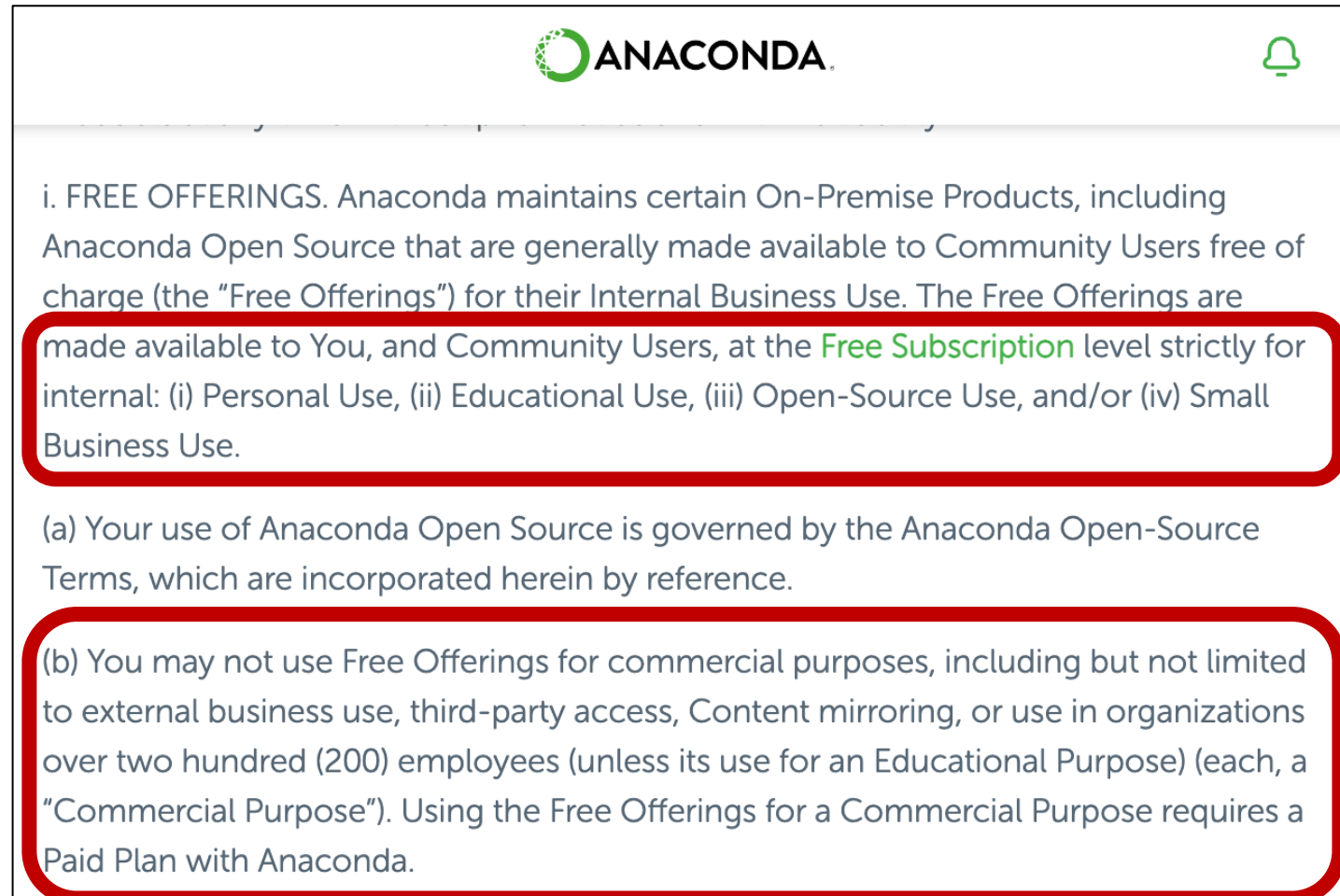
# First click on Terms of Use & search for commercial





**4.2 Commercial Activities.** You agree that you will not (except to the extent expressly authorized by the Agreement or under a separate license agreement with the Company):

- (a) Reproduce, duplicate, copy, sell, trade, resell or exploit for any **commercial** purpose any portion of Company Properties, or access to or use of Company Properties;
- (b) Use Company Properties or any part thereof for any **commercial** or business purpose; or

# If it's open-source, it doesn't mean commercial use is free

A screenshot of the Anaconda website's terms of service page. The Anaconda logo is at the top center, and a notification bell icon is at the top right. The text is organized into sections with red rounded rectangular highlights. The first highlighted section is item (i) under 'FREE OFFERINGS', which states that free offerings are available to users at the 'Free Subscription' level for internal use only. The second highlighted section is item (b) under a list, which states that free offerings cannot be used for commercial purposes, such as external business use or third-party access, and that such use requires a paid plan.

i. FREE OFFERINGS. Anaconda maintains certain On-Premise Products, including Anaconda Open Source that are generally made available to Community Users free of charge (the "Free Offerings") for their Internal Business Use. The Free Offerings are made available to You, and Community Users, at the **Free Subscription** level strictly for internal: (i) Personal Use, (ii) Educational Use, (iii) Open-Source Use, and/or (iv) Small Business Use.

(a) Your use of Anaconda Open Source is governed by the Anaconda Open-Source Terms, which are incorporated herein by reference.

(b) You may not use Free Offerings for commercial purposes, including but not limited to external business use, third-party access, Content mirroring, or use in organizations over two hundred (200) employees (unless its use for an Educational Purpose) (each, a "Commercial Purpose"). Using the Free Offerings for a Commercial Purpose requires a Paid Plan with Anaconda.

# License agreement breaches are real

- Most businesses audited at least once per year by a vendor
- **Additionally:** whistleblowers, competitors, public disclosures, code leaks, etc.
- Breaches are rarely high-profile, but costly:
  - outstanding license fees,
  - audit cost,
  - retroactive payments,
  - penalties.





But it's for *non-commercial research purposes*, and I do R&D.  
Can I build a **prototype** for client & present it at a conference?

## LLaMA: Open and Efficient Foundation Language Models

arXiv

### Abstract

We introduce LLaMA, a collection of foundation language models ranging from 7B to 65B parameters. We train our models on trillions of tokens, and show that it is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to proprietary and inaccessible datasets. In particular, LLaMA-13B outperforms GPT-3 (175B) on most benchmarks, and LLaMA-65B is competitive with the best models, Chinchilla70B and PaLM-540B. We release all our models to the **research** community.

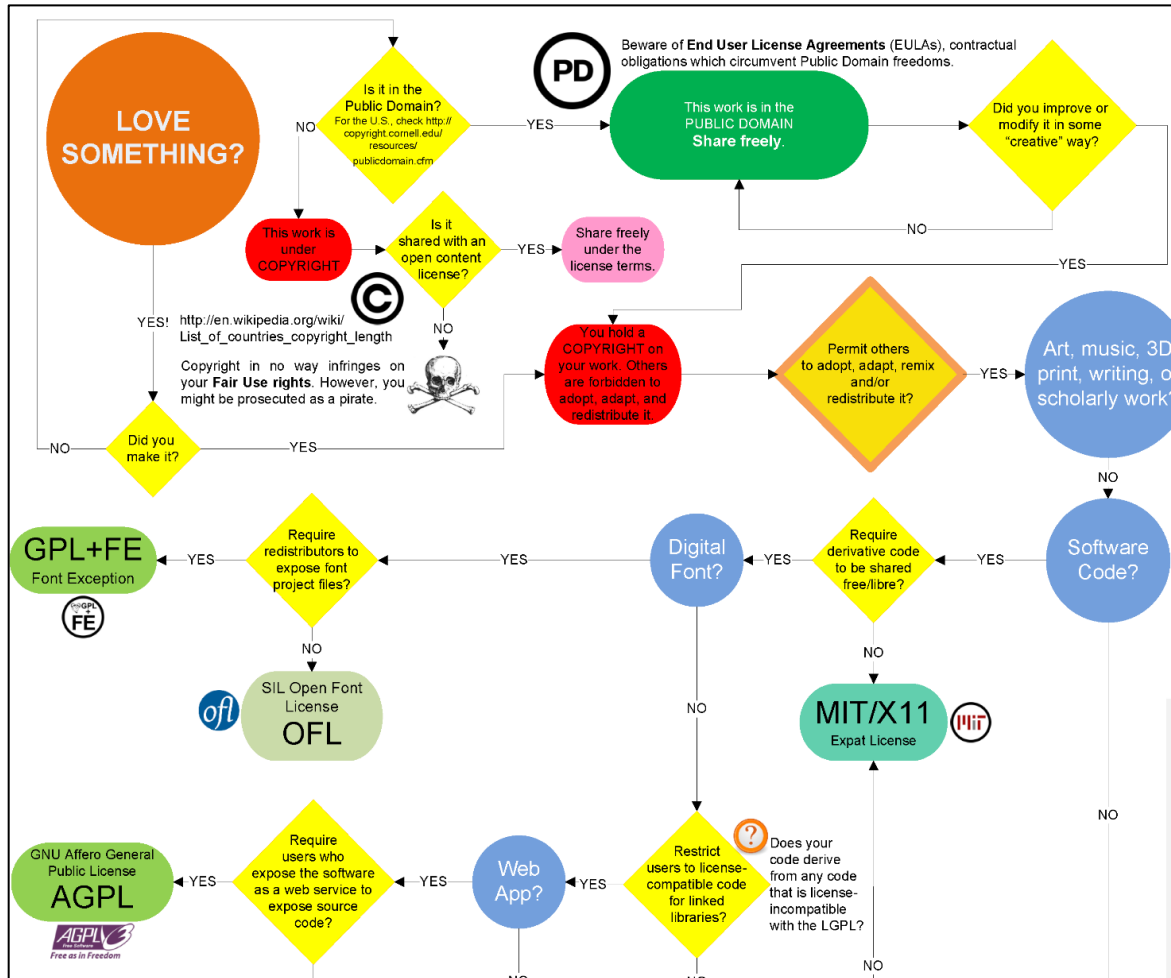
- No clear definition of commercial use
- Consider context & purpose
- **Rule of thumb:** monetary compensation or commercial advantage
- Applies to **academic research**, e.g. if partially supported by private funding (even w/o intention to monetise findings)

# The model/data is *commercially* licensed

Before you commit, **what else** do you need to do **apart from implementing** it?

- **Derivatives:** do you want to be able to modify it?
- **Sharealike:** are you okay with disclosing your work under the same licence?
- **IP protection:** do you want to be able to patent your work? (*e.g. Apache vs MIT*)
- **Other use cases,** *e.g. standalone model hosting* may be not allowed (*e.g. Falcon*)

# Choose wisely when open-sourcing *your own work*



- Use one of many decision trees
- If you aim to support open science, use a commercial licence, possibly attribution + copyleft (e.g. GNU GPL)
- Note: different licences for data, e.g. CDLA-Sharing-1.0



# Copyright

Model input / output

Can I use copyrighted data to train a model?

*The Atlantic*

REVEALED: THE AUTHORS  
WHOSE PIRATED BOOKS ARE  
POWERING GENERATIVE AI

Stephen King, Zadie Smith, and Michael Pollan are among thousands of writers whose copyrighted works are being used to train large language models.

By Alex Reisner

*Editor's note: This article is part of The Atlantic's series on Books3.*

# Training on copyrighted data depends on jurisdiction

- Some countries (e.g. Japan, South Korea, Israel) allow a copyright exception for **text & data mining (TDM)**
- EU: *somewhat* allows TDM under 2019 Copyright Directive:
  1. for the purpose of **scientific research** by institutions,
  2. by any entity for **any purpose**, but allows rightsholders to **opt-out**.
- **Caveats**: lawful access, proving opt-out, etc.
- No Polish implementation yet



# Training on copyrighted data depends on jurisdiction

- USA: fair use interpretation of Copyright Act
  - 2015: Google wins vs Authors Guild. Using copyrighted books to create a searchable database is **transformative**
  - Currently: Big tech using this defence
- Both sides hoping to set a precedent
- Plans to adapt copyright laws to AI age

This recognition of the law's history in comparative context should compel policymakers, who might be inclined to discount this public interest dimension in current copyright debates, to actively distinguish today's circumstances from that of the past. They, along with any interest groups or lobbyists that promote a particular copyright agenda, would have to clearly and cogently 'make their case' for derogating from this policy dimension or any other 'transcendent' principle, at least in those jurisdictions that have been the subject of close historical scrutiny. The more that scholars engage in uncovering the many copyright 'origin stories' around the world, the more a global picture will emerge that will provide greater depth of understanding on the comparative aspects of the law, including any areas of convergence or divergence.

Pages 46 to 128 are not shown in this preview.



# Regardless of who's liable, big tech indemnifies you

## Microsoft announces new Copilot Copyright Commitment for customers

Sep 7, 2023 | Brad Smith, Vice Chair and President, Hossein Nowbar, CVP and Chief Legal Officer

Microsoft's AI-powered Copilots are changing the way we work, making customers more efficient while unlocking new levels of creativity. While these transformative tools open doors to new possibilities, they are also raising new questions. Some customers are concerned about the risk of IP infringement claims if they use the output produced by generative AI. This is understandable, given recent public inquiries by authors and artists regarding how their own work is being used in conjunction with AI models and services.

## Shared fate: Protecting customers with generative AI indemnification

October 13, 2023

**Neal Suggs**  
VP Legal, Google Cloud

**Phil Venables**  
VP, TI Security & CISO, Google  
Cloud

### To our customers:

At Google Cloud, we put your interests first. This means that when you choose to work with us, we become partners on a journey of shared innovation, shared support, and shared fate.



# Can I use a model's output to train a distilled model?

That depends on T&Cs:

- Most vendors (e.g. OpenAI, Google, Anthropic) explicitly forbid to use their models' outputs to develop (competing) models
- Some use permissive licences inspired by Apache 2.0, e.g. Falcon

Bonus points: avoiding double trouble such as Alpaca (*Llama 1 tuned on GPT 3.5*)

# Evolving approach to web scraping

- Data protection regulators:
  - Public data is still subject to privacy laws
  - Platforms & customers urged to protect data, e.g. captchas, rate limits
- OpenAI sued for stealing private information
- Scraping to be treated as data breach?
- Companies aim to prevent scraping (BBC) or monetise API-based data access (Reddit)

UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA	
PLAINTIFFS P.M., K.S., B.B., S.J., N.G., C.B., S.N., J.P., S.A., L.M., D.C., C.L., C.G, R.F., N.J., and R.R., individually, and on behalf of all others similarly situated,	Case No.:
Plaintiffs,	<b><u>CLASS ACTION COMPLAINT</u></b>
vs.	1. VIOLATION OF ELECTRONIC COMMUNICATIONS PRIVACY ACT, 18 U.S.C. §§ 2510, <i>et seq.</i>
OPENAI LP, OPENAI INCORPORATED, OPENAI GP, LLC, OPENAI STARTUP FUND I, LP, OPENAI STARTUP FUND GP I, LLC, OPENAI STARTUP FUND MANAGEMENT LLC, MICROSOFT CORPORATION and DOES 1 through 20, inclusive,	2. VIOLATION OF THE COMPUTER FRAUD AND ABUSE ACT, 18 U.S.C. § 1030
Defendants.	3. VIOLATION OF THE CALIFORNIA INVASION OF PRIVACY ACT (“CIPA”), CAL. PENAL CODE § 631
	4. VIOLATION OF CALIFORNIA UNFAIR COMPETITION LAW, BUSINESS AND PROFESSIONS CODE §§ 17200, <i>et seq.</i>

# Model output copyright: uncharted territory?



## Images:

- US copyright law applies to humans, **not animals/machines**
- Midjourney not considered **tool** that can be **used** by author
- Sufficient **human authorship**: retouched images *no*, book *yes*
- Copyright Office **won lawsuits**, but **court admitted challenges**



## Code & text:

- **As above**, verbatim LLM output could go into public domain
- Copyright only after heavy edits or additions, e.g. **dev labels AI code snippets**, so that the entire application can be copyrighted

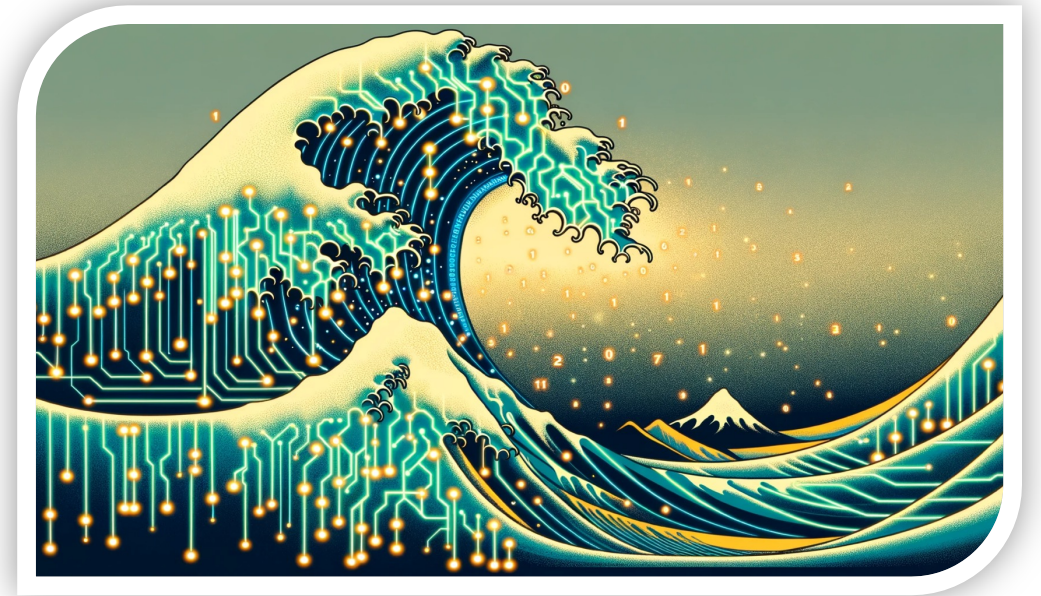


# Regulations

Model deployment & operations

# Tidal wave of AI regulations

- China leads on strict regulations, e.g. devs liable for model output & training data copyright infringement
- EU: GDPR already addresses algorithmic transparency, profiling consent & safeguards, data minimisation
- EU: AI Act in final consultation
- US & UK to follow (Brussels effect)
- Spain sets up 1<sup>st</sup> AI agency in EU
- Brazil looking to lead in S. America
- etc...



# EU AI Act proposes a *risk-based* safety framework

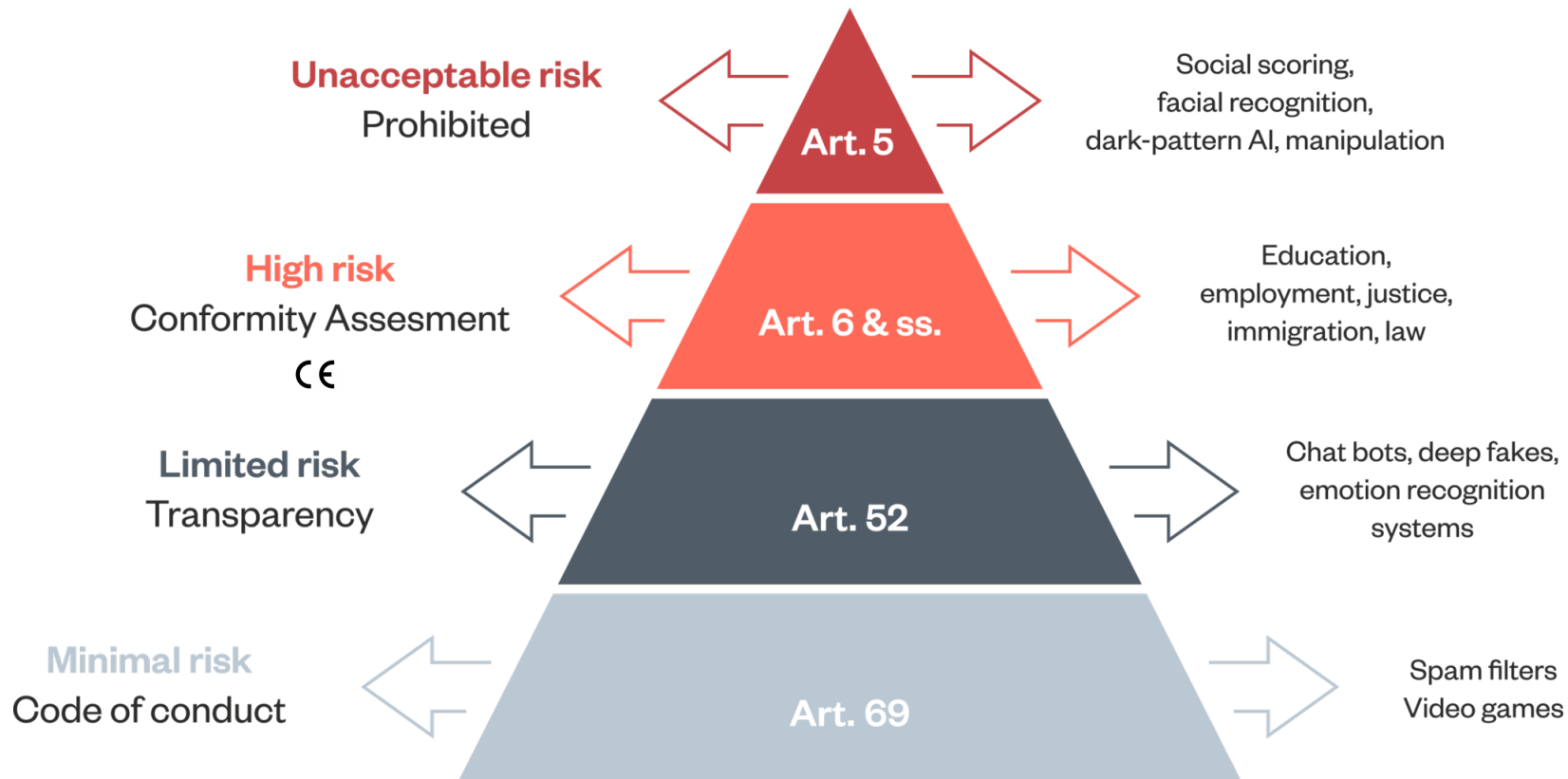


Image: Lilian Edwards (Ada Lovelace Institute)

Pre/post-market requirements for high-risk system providers:

- training data governance,
- technical documentation,
- risk management,
- human oversight,
- accidents reported to authorities.

Significant model finetuning can make you the provider.

Open-source models more likely to meet obligations?

# Other things we won't cover today


- Other risks, e.g. international data transfers, esp. from EU to US
- Other solutions, e.g. regulatory sandboxes (e.g. US autonomous vehicles)
- Proposals for international interoperability (e.g. UN PNAI)




# Let's continue the discussion online & in Poznań



Hosted By  
Gosia and Krzysztof J.





**AI & Tech Talks #8**  
20.09.2023, g. 18:00  
Poznań, Nobel Tower, I piętro


 Pearson

**Details**

Zapraszamy na pierwszy po wakacjach meetup z cyklu AI & Tech Talks. Jak zawsze chcielibyśmy się podzielić z Wami wiedzą i doświadczeniami jakie zdobyliśmy w pracy nad naszymi systemami e-learningowymi. Tym razem na naszej scenie wystąpią speaker i speakerka, którzy przedstawią najnowsze trendy i rozwiązania z obszaru Large Language Models.

 20.09.2023

 godz. 18:00

 Poznań, Nobel Tower, I piętro

[meetup.com/pearson-ai-tech](https://meetup.com/pearson-ai-tech)