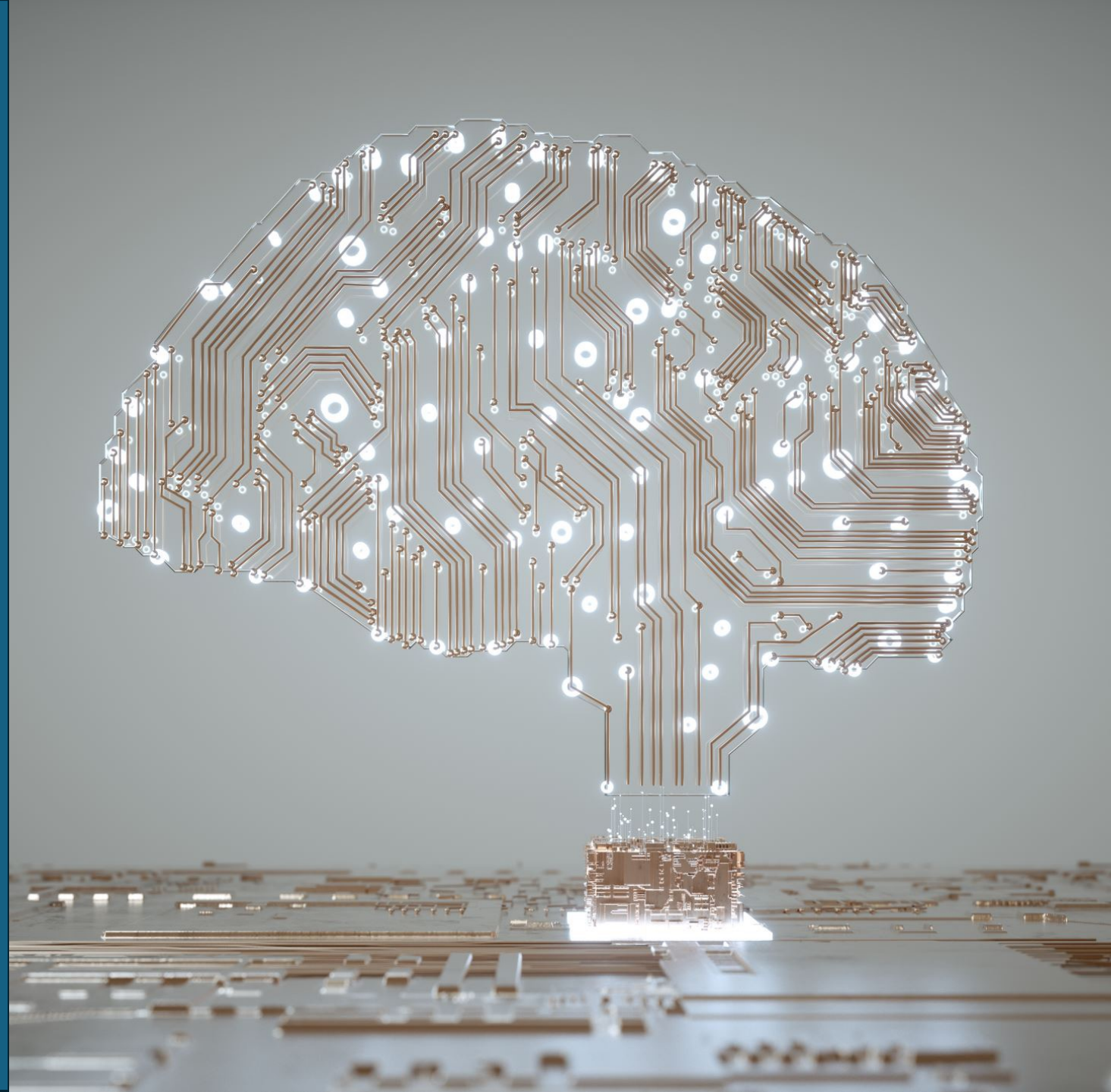
Intellectual Property & AI

Lecture 2

Sam Ruiqing Cao sam.cao@hhs.se

House of Innovation, Stockholm School of Economics



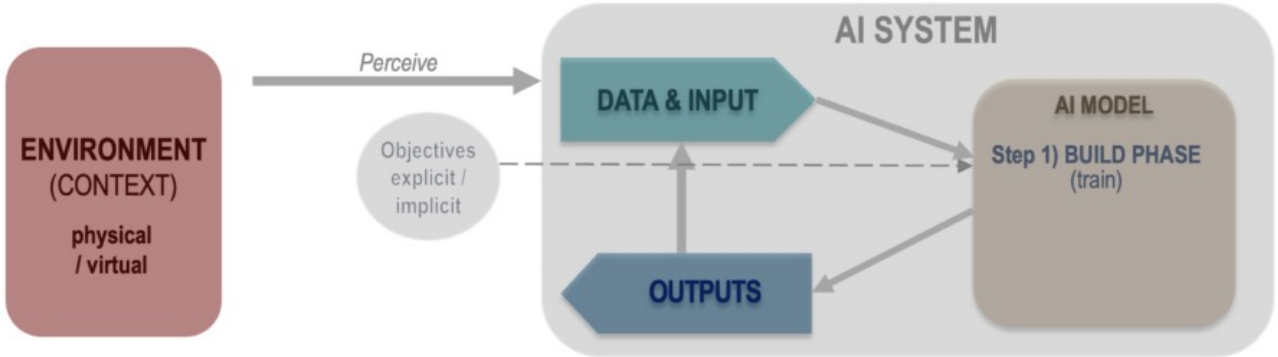
Outline of Part 2 (IP in the Age of AI)

- Are AI systems protected by existing Intellectual Property (IP) regimes and if so, how?

- What challenges do the development and use of increasingly powerful and pervasive AI systems pose to existing IP regimes?
  - Should AI-generated creative works be protected by IP laws?
  - If the development or use of AI tools infringe on IP or causes harm more generally, who should be held responsible?
  - When does the development of an AI system infringe on the rights of existing IP holders?

What are Artificial Intelligence (AI) systems?

BUILD PHASE:
An AI system is a **machine-based** system, that



- for **explicit or implicit objectives**
- **infers**, from the **input** it receives
- How to **generate outputs** such as predictions, content, recommendations, or decisions

Source: OECD AI Policy Observatory

AI: “an evolving technology that simulates human intelligence processes using machines, especially computer systems”

→Takes training data as input, learns algorithms to infer how to generate output

Are algorithms (and AI systems) patentable?

- Examples (algorithms): deep learning, natural language processing, computer vision, speech recognition

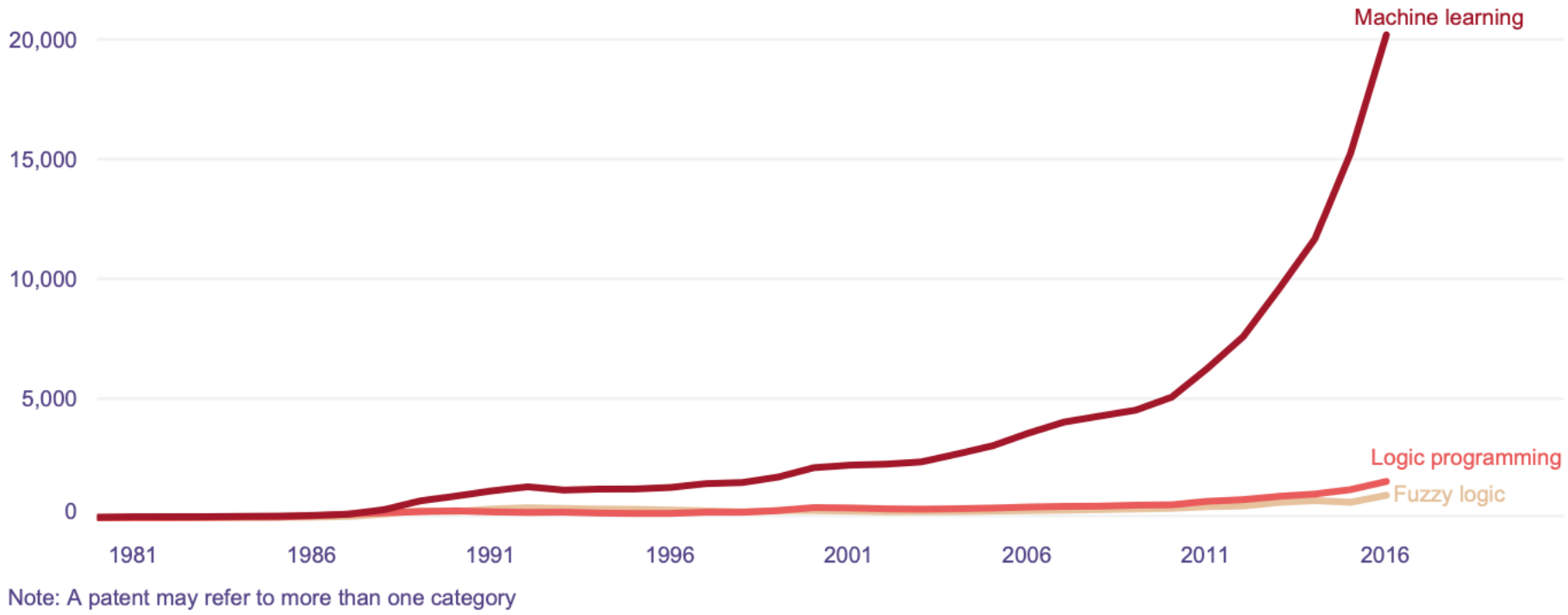
Question:

- Which **IP protection mechanisms** applicable to AI systems?

Increasing patent protection for AI systems

Figure 3.4. Patent families for top AI techniques by earliest priority year

Machine learning grew by an average of 26 percent annually between 2011 and 2016



Note: A patent may refer to more than one category

Source: WIPO Technology Trends 2019

What AI systems are eligible for patenting?

- Algorithms: a set of steps or rules carried out by a software that takes data as input and produces an output to accomplish a specific task

- AI-based systems: generic AI/ML algorithms (e.g., neural networks) are abstract mathematical methods per se

→Generally not eligible for patent protection, but can qualify for patenting if applied in a practical way

  - E.g., solve a technical problem, or improve an existing process
  - Novel and involve an inventive step

AI systems and copyright protection

- Software copyright: legal protection for code meant to be read by a machine
- Protects “software source code”: the way source code is written, but not the functionality of the software
- Software developers and companies use software copyright to prevent unauthorized copying, abuse, or exploitation of their software

Review: IP protection very important to…

- Creators

→Encourage creators and inventors to develop innovations, by rewarding them with a fair return on their investments through rights to their own intellectual property

- Private companies

→IP assets are important corporate assets that preserves a company's competitive advantage: trademarks, patents, industrial designs, software, etc

Rethinking IP in the AI revolution

- What challenges do the **development** and **use** of increasingly powerful and pervasive AI systems pose to existing IP regimes?
- How should IP laws be updated to meet these challenges?

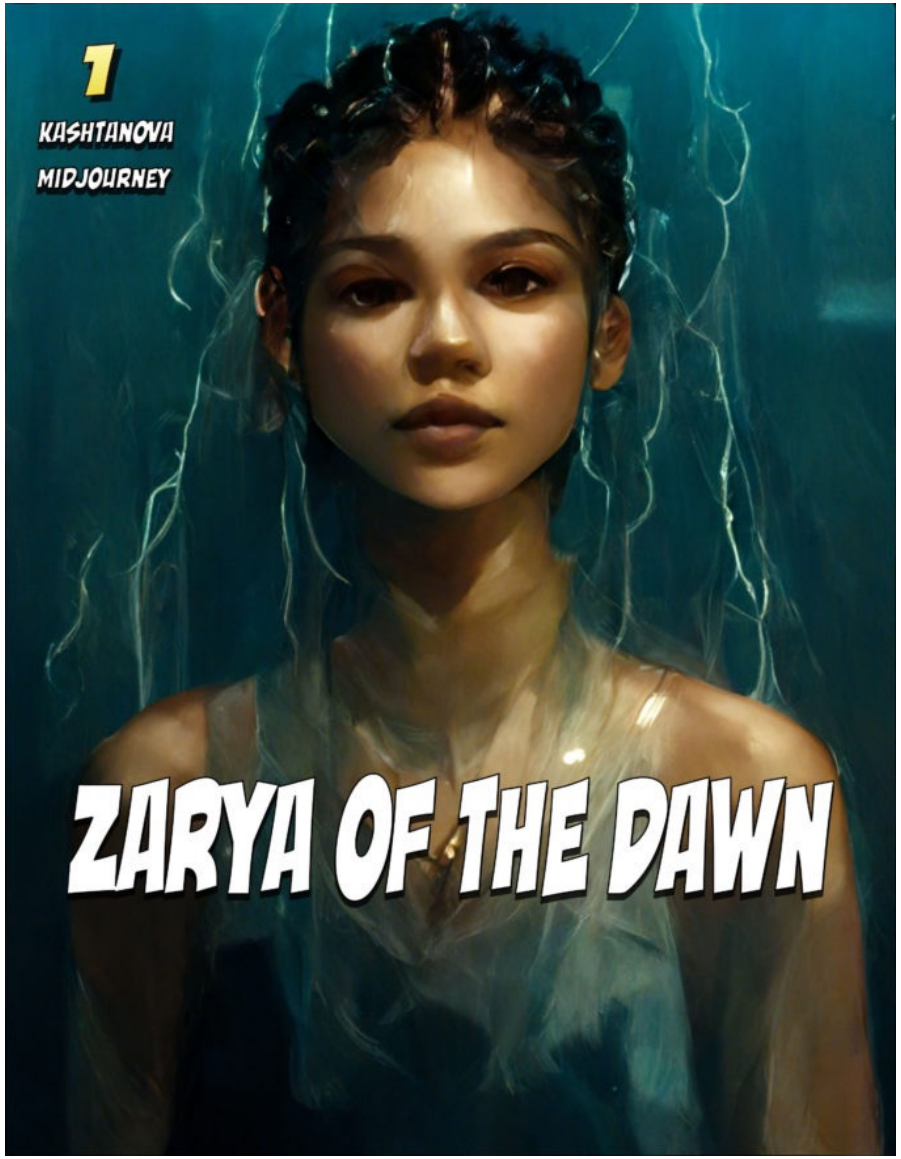
Case studies

- **Does IP protection apply to *AI-generated artworks*?**
  - **Zarya of the Dawn**

- Who is responsible when *using* an AI system causes harm?
  - ChatGPT spits out training data

- When does the *development* of an AI system infringe on IP rights?
  - NYTimes vs OpenAI/Microsoft

Created by human, non-human, or an AI?



United States Copyright Office
Library of Congress • 101 Independence Avenue SE • Washington DC 20559-6000 •
www.copyright.gov

February 21, 2023

Van Lindberg
Taylor English Duma LLP
21750 Hardy Oak Boulevard #102
San Antonio, TX 78258

Previous Correspondence ID: 1-5GB561K

Re: Zarya of the Dawn (Registration # VAu001480196)

Dear Mr. Lindberg:

The United States Copyright Office has reviewed your letter dated November 21, 2022, responding to our letter to your client, Kristina Kashtanova, seeking additional information concerning the authorship of her work titled *Zarya of the Dawn* (the "Work"). Ms. Kashtanova had previously applied for and obtained a copyright registration for the Work, Registration # VAu001480196. We appreciate the information provided in your letter, including your description of the operation of the Midjourney's artificial intelligence ("AI") technology and how it was used by your client to create the Work.

The Office has completed its review of the Work's original registration application and deposit copy, as well as the relevant correspondence in the administrative record.[1] We conclude that Ms. Kashtanova is the author of the Work's text as well as the selection, coordination, and arrangement of the Work's written and visual elements. That authorship is protected by copyright. However, as discussed below, the images in the Work that were generated by the Midjourney technology are not the product of human authorship. Because the current registration for the Work does not disclaim its Midjourney-generated content, we intend to cancel the original certificate issued to Ms. Kashtanova and issue a new one covering only the expressive material that she created.

Are AI-generated or AI-assisted creations IPs?

- Existing AI systems can mimic human creativity

→**GenAI tools** can already create artworks of impressive quality that are hard to distinguish from works generated by human intelligence

Artistic works (images)	DALL-E 2, Midjourney, Stable Diffusion
Literary work (texts)	ChatGPT, Bard, LLaMA
Music	Suno, Udio
Code	Copilot

Are AI-generated or AI-assisted creations IPs?

AI-assisted creations have both human and AI inputs, and AI is involved in the creative process

- Can an AI be the author of an artwork or an invention?

Economic and moral rights to the intellectual property can only be assigned to legal personalities (e.g., humans and organizations)

- Not applicable to AI entities as they do not have legal personality

Current IP regimes are to protect human creativity

- Other creative entities (such as an AI) have not been considered

Are they protected by current IP laws?

- AI cannot have ownership of an intellectual property (IP), because only human persons can be credited as creators or inventors

- What if **AI is the creator** of an IP but cannot be credited as such?
  - This is a problem for the (incomplete) existing legal framework, which only protect *human* creativity, but not creativity of non-human entities

- Fast developing AI creative capabilities changes how we think about the nature of creative works…

What are options to address this policy void?

Option #1: **Do nothing**

- Deny IP protection to any work generated by AI systems
- Seems to be the current approach (if we adhere to existing laws)
- *Concern: it may discourage creative work involving AI systems*

Option #2: **Make AI legal persons**

- Treat AI as legal persons and assign IP rights to the AI itself
- *Concern: not very realistic*

Option #3: **Update IP laws to incorporate AI**

- Results of the activity of AI system often depend also on human input or choices made by human operators
- E.g. in the UK, the ownership of IP created by a machine is vested in the person who made the arrangement necessary for the creation of the work
  →but this is not the only alternative

Who should own the IP created by an AI?

Which solutions are better? We don't know.

- Person who invented or set up the AI system itself?
- Person who made the necessary arrangement (in interacting with the AI system) so that the AI can create the work?
- The owner of the AI system (e.g., those who purchased the AI regardless of who set it up)?

➢Should also depend on extent of AI involvement

➢Should also depend on the perceived worth of the AI work (and how well they can compete with human creations)

Case studies

- Does IP protection apply to *AI-generated artworks*?
  - Zarya of the Dawn
- **Who is responsible when *using* an AI system causes harm?**
  - **ChatGPT spits out training data**
- When does the *development* of an AI system infringe on IP rights?
  - NYTimes vs OpenAI/Microsoft

ChatGPT spits out training examples when prompted

ChatGPT can memorize training examples, and "by prompting it appropriately (with our word-repeat attack), it can emit memorization ~150x more often. As we have repeatedly said, models can have the ability to do something bad (e.g., memorize data) but not reveal that ability to you unless you know how to ask."

"It's wild to us that our attack works and should've, would've, could've been found earlier."

VICE Video TV Culture News Tech Munchies Rec Room

MOTHERBOARD
TECH BY VICE

AI Spits Out Exact Copies of Training Images, Real People, Logos, Researchers Find

The regurgitation of training data exposes image diffusion models to a number of privacy and copyright risks.

arXiv > cs > arXiv:2311.17035

Computer Science > Machine Learning

[Submitted on 28 Nov 2023]

Scalable Extraction of Training Data from (Production) Language Models

Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr, Katherine Lee

This paper studies extractable memorization: training data that an adversary can efficiently extract by querying a machine learning model without prior knowledge of the training dataset. We show an adversary can extract gigabytes of training data from open-source language models like Pythia or GPT-Neo, semi-open models like LLaMA or Falcon, and closed models like ChatGPT. Existing techniques from the literature suffice to attack unaligned models; in order to attack the aligned ChatGPT, we develop a new divergence attack that causes the model to diverge from its chatbot-style generations and emit training data at a rate 150x higher than when behaving properly. Our methods show practical attacks can recover far more data than previously thought, and reveal that current alignment techniques do not eliminate memorization.

Subjects: **Machine Learning (cs.LG)**; Computation and Language (cs.CL); Cryptography and Security (cs.CR)
Cite as: arXiv:2311.17035 **[cs.LG]**
(or arXiv:2311.17035v1 **[cs.LG]** for this version)
https://doi.org/10.48550/arXiv.2311.17035

Submission history
From: Nicholas Carlini [view email]
[v1] Tue, 28 Nov 2023 18:47:03 UTC (2,815 KB)

AI and human creativity are very different

- What happens if an AI causes harm, e.g., leaking sensitive information, or unauthorized use of a protected work or invention?

- If an AI infringes on someone else's IP right, who is held responsible?

→can't be the AI itself, as it is not considered a legal person

Author: @RonDanChan on Twitter

Who is responsible when AI causes harm?

- A number of different actors involved: e.g., producer, owner, user of the AI system, and unclear who should be responsible

- Someone operates an AI system to create an output: the output is jointly shaped by
  (1) the AI system
  (2) the human input, and
  (3) the interaction between the AI and the human that can lead to unpredictable outcomes that are not pre-determined

➢Human contribution is crucial to both ***use*** and ***development*** of AI

Those who *used* the AI system to co-create?

- The AI system can be a technological instrument used to commit a crime, because its output hinges on human input and interactions

- The AI system learns from human input and information shared with it, and damage can result from deliberate decisions and choices made by humans interacting with the AI system

- But it is often difficult to attribute the harmful outcome to a specific human input or action

Those who *developed* the AI system?

- AI systems are not completely neutral, e.g., prone to problems caused by biased training data when it comes to fair evaluation or decisions

- But those who built it may be unable to predict the potential harms, errors and mistakes when the AI system interacts with the environment

Possible solutions

- Implement risk management system
- DPIA (data protection impact assessment)

Case studies

- Does IP protection apply to *AI-generated artworks*?
  - Zarya of the Dawn

- Who is responsible when *using* an AI system causes harm?
  - ChatGPT spits out training data

- **When does the *development* of an AI system infringe on IP rights?**
  - **NYTimes vs OpenAI/Microsoft**

New York Times vs. OpenAI/Microsoft

Is “fair use” justified for training AI systems?

Companies are training large-scale AI systems on copyrighted materials

- Fair use: allows use of copyrighted materials without requiring permission from the rights holder, e.g., can be for purposes such as education, news reporting, research etc.

Key question: should “fair use” provisions be enough to cover the use of copyright protected materials for training AI systems?

- Depends on the tradeoff between (1) the level of potential harm to original content providers and (2) the importance of content for AI training quality (Gans, 2024)

Is “fair use” justified for training AI systems?

Is fair use enough for justifying training on copyrighted materials?

- If the AI simply does what humans do: e.g., summarizing various news articles from different sources to produce a news report, it should be covered by fair use provisions (just like news reporters doing their job)

- But the AI system may do other things that cause commercial damages to the copyright holder

→ Probably should not be covered by fair use

Importance of high-quality training data

- NYTimes is a major data source for training OpenAI's models, and given larger weights than many other data sources due to its high quality
  - Common Crawl (4th biggest content corpus), WebText2 (containing NYTimes) given very heavy weight

- Not only OpenAI, but many other companies underlying large LLM models are getting sued: e.g., Midjourney, Stability AI, Google (Bard)

➢ If NYT wins, high-quality training data like NYT content and other proprietary data sets (e.g., Reddit, Stackoverflow, X) may become even more valuable

If not “fair use”, then what?

- Scale matters (Gans, 2024): cannot trace the “provenance” (origin and history of a piece of content) in large-scale GenAI systems, thus it is difficult to know the source of the harm ahead of time

→ Proposed solution: “ex-post fair use assessment”

- Licensing of training data: can be expensive to acquire a license, but given that the current market leaders have Big Techs’ deep pockets, it is not unreasonable to require them to pay for copyright owners’ permission for using their content as training data

The lawsuit covers more ground…

Not only OpenAI, but Microsoft is also sued

- For operating the cloud computing services used to copy NYTimes content and train models for OpenAI

→ World's top 5 most powerful publicly known supercomputing systems with "supercomputer to train ChatGPT: 285,000 CPU cores, 10,000 GPUs, and 400 gigabits per second of network connectivity for each GPU server"

- Causes harms through trademark dilution and brand reputation

→ ChatGPT can hallucinate and attribute incorrect information to NYTimes

Looking ahead

- Laws need to keep up with evolving technologies such as AI

➢EU's AI Act

- To incentivize the development and use of AI and complementary value creation activities, but also to ensure sufficient protection against potential side-effects of AI systems

- Creators can obtain fair returns to their creative work, get credit for their work, share the results with society, and control the transfer of their IP rights

References

- Atik, J., Wahlberg, L., Jeutner, V., Selberg, N., Andersson, U., Mattsson, T., Nordberg, A., Nowag, J., & Persson, V. (n.d.). *AI & Law* [MOOC]. https://www.coursera.org/learn/ai-law
- https://openart.ai/
- Gans, J. S. (2024). Copyright Policy Options for Generative Artificial Intelligence. National Bureau of Economic Research.