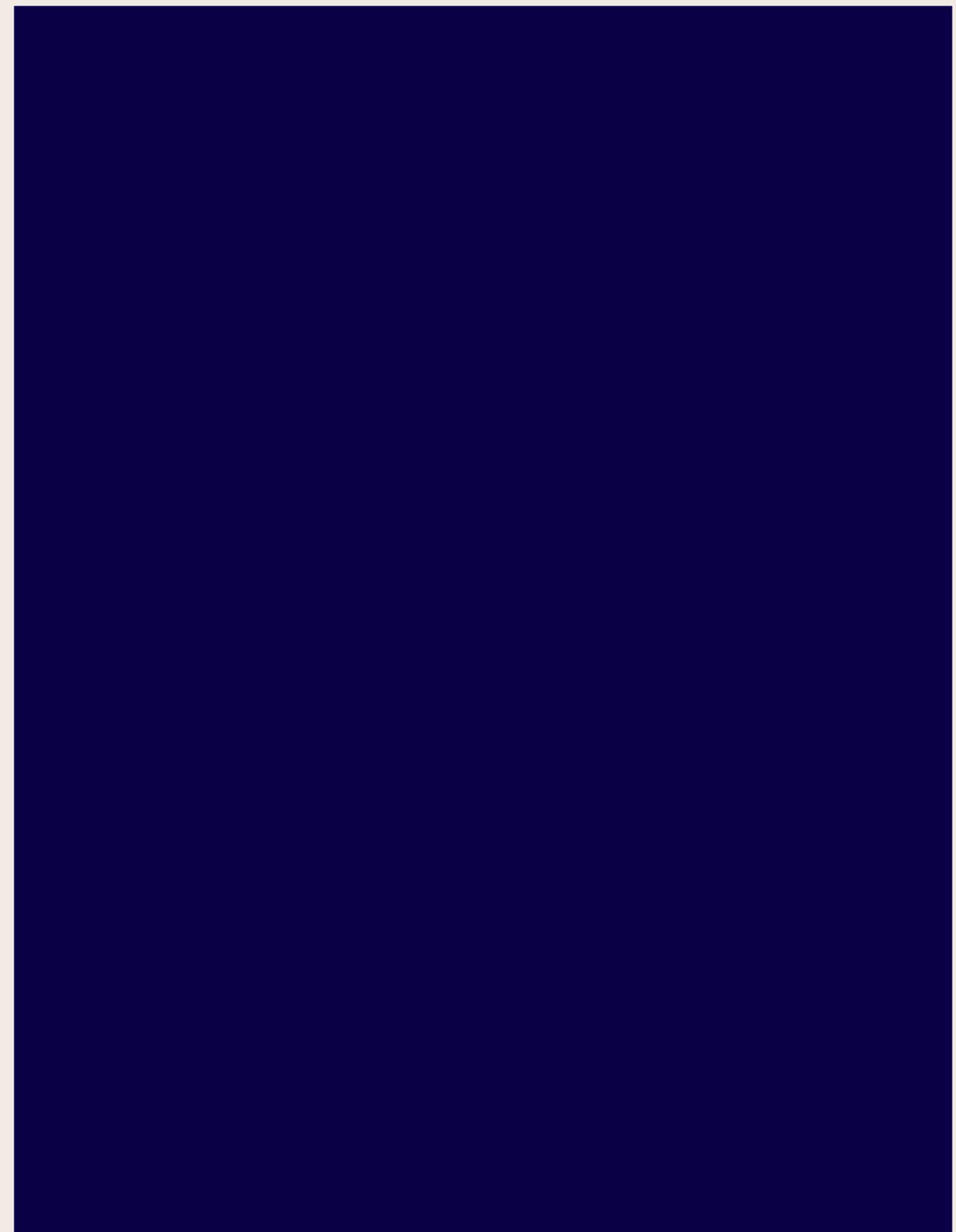


AFTERNOON GROUP 10

LLM TRAINING AND COPYRIGHT

A DIVE INTO POTENTIAL
POLICY



WHAT IS COPYRIGHT INFRINGEMENT

- Extensive use of data creates ethical concerns
- Copyrighted material used without permission from the rights holder
- Boundaries between fair use and violation of personal rights
- Creators ability to retain control over their work
- The trade-off between the benefits and maintaining a safe online environment.



NYT VS OPENAI

- NYT sues OpenAI for using articles to train AI
- Potential threat to journalism
- NYT seeks destruction of AI models trained on its content. What is a fair resolution?
- Balancing copyright protection with technological advancement
- Arguments based on Google Books Case
 - Google posted scanned snippets of books without consent
- AI uses copyrighted information and outputs a completely replaced result. Is that okay?



EUROPEAN UNION POLICY FOR COPYRIGHT PROTECTED CONTENT FOR ML



- UK first allowed unauthorized reproduction of copyrighted works for the purpose of non-commercial Text and Data Mining (TDM).
 - France, Germany, and Estonia later followed suit.
- European policymakers started proposing changes to EU copyright rules that would bind every Member State to adopt corresponding TDM exceptions.
- Research institutions will have the unlimited right to mine copyrighted content, while other actors still must respect the opt-out choice of the rightsholder.

RIGHTS OF DATA CREATORS OR OWNERS

- Right to file a lawsuit against the company who created the model
- Right to stop the use of their copyrighted data by the GenAI companies
- Right to claim any kind of profits (direct or indirect) GenAI gains from using their product
- Right to set terms and conditions as to when, how and where they want AI systems to use their content
- Right to exercise their control over any output that has been derived

HANDS OFF!
It's mine.

Regulation for Gen AI Developers: Protection of Copyrighted Data Act

Ban the training of large language models (LLMs) on copyrighted data in the US without explicit consent.

Establish a licensing agreement for all copyrighted data, allowing owners to set usage guidelines and requiring proper credit for their data in generated outputs.



Implications of Stringent Copyright Laws on AI Industry

01. Surge in lawsuits against LLM creators.

02. Increased costs for compliance and licensing.

03. Reduced capabilities of leading AI models.

04. Ethical AI at competitive disadvantage.

05. AI companies may relocate overseas.

Trade-offs

Duty to protect individual
copyright holders

Potentially decimate the AI
industry

Providing data creators with
due credit and financial
compensation

Increased cost for GenAI
users

Protecting US Data creators

Impede the pace of AI progress
internationally

Alternative Scenario: Regulating Data Holders

- Regulations to ensure that Copyrighted material is only accessed with expressed consent
- Data Poisoning Mandate: Mandate that data holders (Social media, Data storage) utilize data poisoning to fight against unethical training

PROJECTIONS & TRADEOFFS

HIGH COST
TO DATA
HOLDERS

USABILITY
COULD BE
REDUCED

US DATA
COSTLIER TO
TRAIN ON

MODELS LESS
EFFECTIVE IN
US

GEN AI
DEVELOPERS
STAY IN US

US GOV
MAINTAINS
CONTROL
OVER GEN AI



Comparison of Regulatory Outcomes

REGULATE GEN AI DEVELOPERS

- Reduction in Gen AI efficiency
- Would result in defection of developers to more lenient countries
- Duty to defend rights of copyright holders

REGULATE DATA HOLDERS

- Reduction in Gen AI efficiency
- Would reduce defection from developers
- Likely a better outcome
 - Less copyright protections, but more efficient and safe AI

STAKEHOLDERS PERSPECTIVE ON COPYRIGHT GUIDELINES

Creators (Artists, Musicians etc.)

Concerned
about fair
compensation,
recognition by
AI and losing
control over
their work

Tech / GenAI Companies

Face challenges
in adapting
business models
and increasing
operational
costs

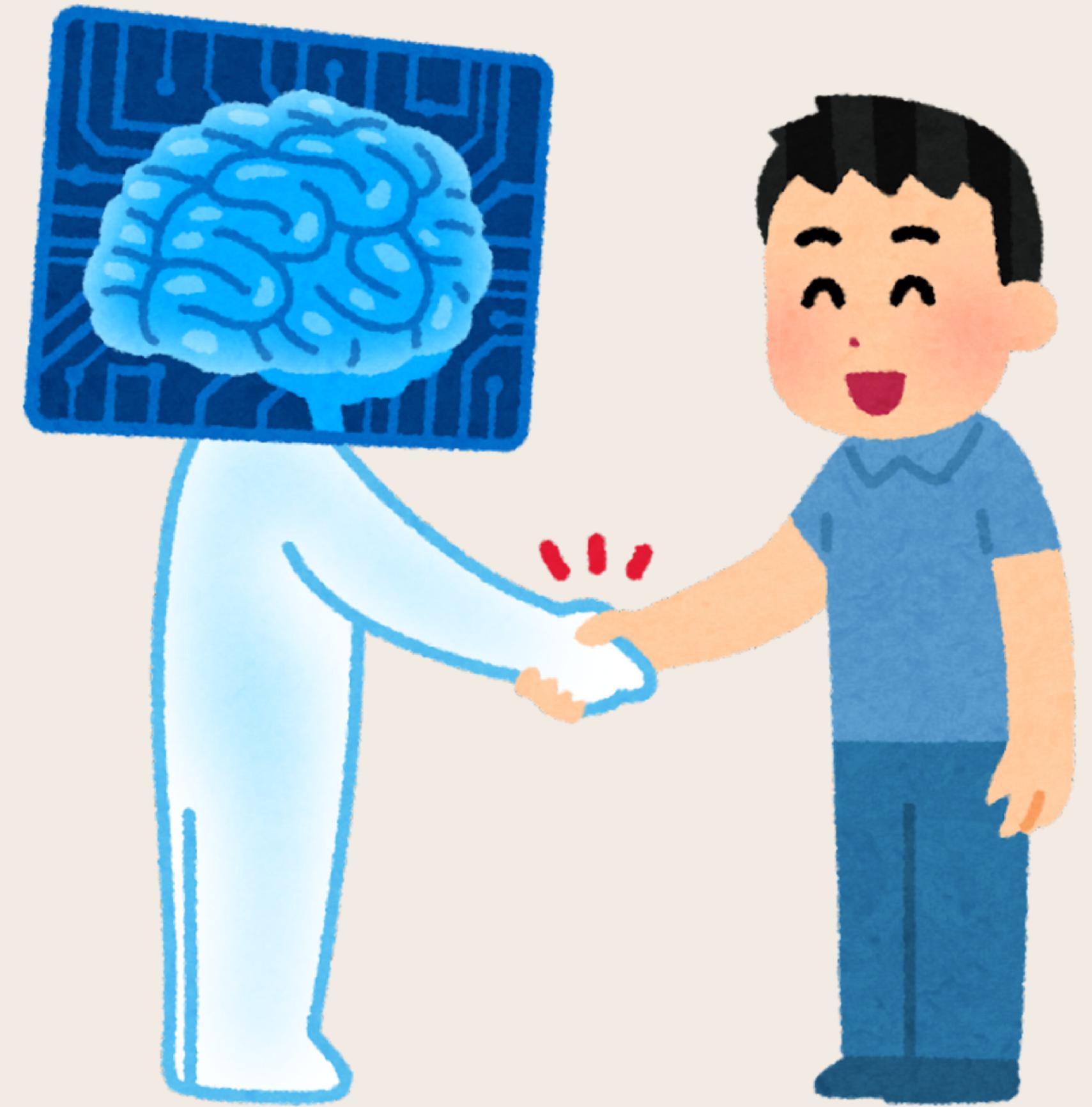
Legal Experts

Worry about
enforcing the
laws and
international
legal alignment

General Public

Mixed feelings
about AI
generated
content and
concerns about
data privacy
might still exist

**THANK
YOU VERY
MUCH!**



Citations

K&L Gates. "Federal Court Rules Work Generated by Artificial Intelligence Alone Is Not Eligible for Copyright Protection." *K&L Gates*, 30 Aug. 2023, www.klgates.com/Federal-Court-Rules-Work-Generated-by-Artificial-Intelligence-Alone-Is-Not-Eligible-for-Copyright-Protection-8-30-2023#:~:text=1%3A22%2Dcv%2D01564,protection%20under%20the%20Copyright%20Act.

Authors Alliance. "Fair Use Week 2023: Looking Back at Google Books Eight Years Later." *Authors Alliance*, 24 Feb. 2023, www.authorsalliance.org/2023/02/24/fair-use-week-2023-looking-back-at-google-books-eight-years-later/.

Capoot, Ashley. "AI Startup Stability Lays Off 10% of Employees After CEO Exit." *CNBC*, 18 Apr. 2024, www.cnbc.com/2024/04/18/ai-startup-stability-lays-off-10percent-of-employees-after-ceo-exit.html.

Simonite, Tom. "Midjourney Copyright Battle Shows How Messy Generative AI Laws Are." *IEEE Spectrum*, 23 Jan. 2024, www.spectrum.ieee.org/midjourney-copyright.

Turton, William. "Midjourney and OpenAI Defend Themselves in Copyright Battle." *The Register*, 8 Jan. 2024, www.theregister.com/2024/01/08/midjourney_openai_copyright/#:~:text=OpenAI%20defends%20itself%20to%20Lords&text=The%20AI%20biz%20said%20it,to%20support%20and%20empower%20creators.%22.

Lee, Timothy B. "Why the New York Times Might Win Its Copyright Lawsuit Against OpenAI." *Ars Technica*, 15 Feb. 2024, www.arsTechnica.com/tech-policy/2024/02/why-the-new-york-times-might-win-its-copyright-lawsuit-against-openai/.

Harvard Law Review. "NYT v. OpenAI: The Times's About-Face." *Harvard Law Review*, 24 Apr. 2024, www.harvardlawreview.org/blog/2024/04/nyt-v-openai-the-times-about-face/.

"What Is Copyright?" *U.S. Copyright Office*, www.copyright.gov/what-is-copyright/.