

Attribution and Moral Rights in Generative AI

david mertz, ph.d.

mertz@gnosis.cx

licenseai.substack.com

Utopian Visions

medical testing
scientific research
knowledge access

Partisans of Utopia

freedom of information
capitalist profits
efficiency of intellectual labor

Dystopian Fears

artificial general intelligence

the alignment problem

rule by robot overlords

Grounded Dangers

bias and toxicity
overreliance & reflexive trust
rule by data

Bluwashing

the frontier model forum
(limited) indemnification
copilot *code referencing*

Attribution and Forgery

simulacra of humanity

hallucinations

authorship

Watermarking

red/green tokens
small pixel deltas

Entropy and Perplexity

Obama	Hussein	<u>Adama</u>	<u>Ehud</u>	Moshe	<u>brandy</u>
0.85	0.1	0.01	0.01	0.003	0.002

Michael	<u>James</u>	<u>Mary</u>	Richard	<u>Jennifer</u>	...many...
0.043	0.041	0.035	0.034	0.031	< 0.03

Copyrights and Royalties

creative works as training corpora

protection of *publishers*

laborers in the machine of capital

Representing and Refraining

right to publicity
moral rights

Right to Be Forgotten

eu general data protection regulation §17

hold: india, s. korea, argentina, philippines

oppose: china and usa (mostly)

Rights of Creators

authors guild:

royalties • consent • dignity

Compensation

compensate authors who wish
to allow their works to be used
in training of generative AI

Consent

require permission for the use of
writers' works in generative AI

Transparency

**create transparency obligations for AI
developers to disclose what works
they use to train their AI**

Label AI Content

require authors, publishers, platforms & marketplaces to identify when a significant portion is AI generated

Authors Guild v. Google

how transformative is aggregation?

fair use and fair dealing

Availability

appropriation and unconscious use

statistical removal of training

copyright, patents, trademarks & trade secrets

“As-If” Attribution

gradients per source

“semantic” similarity

in-style-of prompts

Redaction Agnosticism

gradient ascent

least-squares concept erasure

google machine unlearning challenge

Attribution Description Framework

<https://github.com>

[/anaconda-ai](https://github.com/anaconda-ai)

[/attribution-description-framework](https://github.com/anaconda-ai/attribution-description-framework)

Anaconda
Machine Learning
Public License
(AMPL)

What Next?

david mertz, ph.d.

mertz@gnosis.cx

licenseai.substack.com