



Universiteit
Leiden
The Netherlands

Metaphors we judge (AI) by: a rhetorical analysis of artificial copyright disputes

Smit, M.A.

Citation

Smit, M. A. (2026). Metaphors we judge (AI) by: a rhetorical analysis of artificial copyright disputes. *Journal Of Intellectual Property Law & Practice*, 21(4), 250-262.
doi:10.1093/jiplp/jpag018

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4303925>

Note: To cite this publication please use the final published version (if applicable).

Metaphors we judge (AI) by: a rhetorical analysis of artificial copyright disputes

Michiel A. Smit 

Faculty of Law, Leiden University, Leiden, the Netherlands

Michiel Smit is a jurist and historian working at Leiden University. He is currently writing a doctoral dissertation on copyright, in which he combines his interests in law, history and language.

*Email: m.a.smit@law.leidenuniv.nl

Abstract

- This article is a 'metaphorical' guide to today's most pressing artificial intelligence (AI) copyright questions, focusing in particular on the EU and the USA. Is unauthorized training on copyright-protected works permitted? Can AI models copy? And is AI-generated output itself protected? As this article demonstrates, debates on these questions can all be traced back to a handful of crucial metaphors.
- After all, generative AI is hardly comprehensible without the extensive use of metaphors and analogies. Most notably, AI is systematically conceptualized in human terms such as 'neural networks' that 'learn', 'know' or 'memorize'. This article aims to demonstrate how such metaphors (unconsciously) influence legal evaluations and even judicial decisions in copyright law.
- The resulting analysis is particularly relevant to lawyers, judges and artists interested in copyright and its intersection with AI. Yet, it may also appeal to those interested in AI, legal reasoning and language more generally, as metaphors and their (rhetorical) effects are by no means unique to copyright and may be equally relevant in fields such as privacy law and (legal) philosophy.

I. Introduction

Generative artificial intelligence (GenAI) is too complex to fathom directly in a technological sense. At the very least, that is true for most lawyers, who instead orbit these 'black boxes' *metaphorically* without ever touching their core. As a consequence, language and rhetoric matter even more than they did before in copyright disputes. This article aims to analyse how the key questions concerning copyright and AI, indeed, largely revolve around metaphorical thinking, and how these metaphors affect our legal reasoning and judgements on these matters. In doing so, it provides an overview of European and US approaches to the most essential AI-related copyright questions with a strong emphasis on language.

The first legal question to be addressed is whether unauthorized training on data is allowed under copyright. A debate in which the metaphor 'GenAI learns' plays a crucial role. Next, the 'memorization' question will be discussed: do AI models actually copy works on which they are trained, or do they merely *memorize* or *remember*? Finally, the matter of copyrightability: can AI-generated output be protected by copyright? As will become apparent, this debate hinges on the question whether GenAI is just a 'tool' or 'instrument' or something more. But before addressing these questions, it is important to begin with an introduction to metaphors and why these matter to lawyers, particularly those dealing with AI and its intersection with copyright.

II. Metaphors and artificial intelligence

'And, on the contrary, metaphors, and senseless and ambiguous words are like ignes fatui; and reasoning upon them is wandering amongst innumerable absurdities; and their end, contention and sedition, or contempt'.

Thomas Hobbes¹

The metaphor is often seen as a mere 'figure of speech'; a poetic but *superficial* device with little relevance to everyday life.² However, as most historians are aware, metaphors can be much more than just this and can function as essential parts of a (false) narrative that is actually believed in and acted upon.³ From a linguistic perspective, Lakoff and Johnson, in particular, have demonstrated the omnipresence and importance of metaphors in their influential *Metaphors We Live By* (1980). According to them: 'The essence of metaphor is understanding and experiencing one kind of thing in terms of another'.⁴

¹ T Hobbes, *Leviathan* (1651) para on 'Science'.

² G Lakoff and M Johnson, *Metaphors We Live By* (Chicago 1980) 3.

³ Historical metaphors in particular are notoriously dangerous because of their abuse by, eg autocratic regimes. Think only of the contemporary widespread and reckless use of charged Second World War terms. Luckily, we will not be focusing on these types of metaphors in this article but stick to copyright related issues.

⁴ Lakoff and Johnson, *Metaphors We Live By* 5.

A metaphor, then, is an expression that transfers meaning or experience from one phenomenon to another. After all, the term stems from the Greek ‘*metaphora*’, meaning ‘to transfer’ or ‘transference’.⁵ That said, according to Lakoff and Johnson a metaphor could be much more than just an expression. Sometimes, the expressed metaphor is just *secondary* and may reflect an underlying metaphorical way of *thinking* or *experiencing*. According to Lakoff and Johnson ‘concepts are metaphorically structured in a systematic way’,⁶ and this metaphorical experience and understanding can actually change our perception of reality and thus influence our actions.⁷

Such deeper (conceptual) metaphors could also be seen as ‘systematic’, as they consist of multiple interconnected metaphors. They are often (un)consciously used to structure one complex and abstract phenomenon in terms of another more tangible and simpler phenomenon.⁸ In this way, meaning or understanding is transferred from a simpler source domain to a less accessible target domain. For instance, life could be (structurally) experienced in terms of another phenomenon such as a (rat) race or a river.

Metaphors can also be used to explain a new complex technology. For example, we may clarify generative AI—the ‘target’—metaphorically by using a car or a river as the ‘source’. In legal literature, the terms model and system are often mixed up and used interchangeably, even though they mean different things, and this nuance can be highly relevant for legal analysis (eg think of liability). To illustrate how they differ, we can say that the trained AI model is the mere *engine* of the car. It is very important but useless by itself if you want to actually drive. Therefore, the engine needs to be integrated into a full car, with a chassis, carrosserie, wheels, steering wheel, etc. The AI system is the full car which the user can operate through the steering wheel (API).⁹ In this way, we structure a highly complex and abstract phenomenon (GenAI) in terms of a simpler and more tangible object (the car).

How is this relevant to AI and copyright? First of all, the very origin of AI lies in a ‘metaphorical’ hypothesis: a non-human device that could exhibit (human) *intelligence*.¹⁰ From the very beginning, ‘our’ own *human* intelligence inspired the development of AI. In that sense, it is no coincidence that GenAI is systematically structured and experienced metaphorically as a brain: think of neural networks, neurons, synapses, (brain)layers, learning, and of course the very term AI. Likewise, in the GenAI applications we daily use, we cannot escape anthropomorphic metaphors, for example ChatGPT has ‘think’ and ‘reason’ options. The metaphor ‘AI is a brain’ is widespread among users, lawyers, politicians and AI-technicians alike.

This brings us to a defining feature of the ‘AI is a brain’ metaphor. It is not a one-way flow but a *reciprocal* metaphor.¹¹

We are not just humanizing the computer and AI, but we are also unconsciously computerizing and algorithmizing our own brain (and our view of the world). For example, ever since computers have been a part of most of our lives: human memory is frequently—in contrast with reality—compared to computer storage, as if one could open a PDF-file in his brain.¹² The other way around too, we speak of computer *memory*, as if the computer could remember. The same applies *a fortiori* to GenAI. In short, the metaphor is *double-edged*: AI is a brain and the brain is AI.

The reason that this particular metaphor of the ‘human brain and AI’ is potentially so misleading is that these reciprocal metaphors are made between two phenomena we *both* do not truly understand.¹³ To some extent, this is even true for computer and neuroscientists: like the human brain, ‘artificial neural networks’ at least partly transcend our comprehension (for us lawyers even more so probably).¹⁴

In light of this, this metaphor could potentially confuse people and mislead them. Using a rock or car as the source of a metaphor can be useful and clarifying precisely because most people understand these tangible and simpler domains. However, most jurists are not familiar with neurons, synapses, vectors and so on, let alone the interconnection of these highly abstract and unintelligible phenomena. Because of this—and the systematicity of the brain metaphor—we may fail to identify the metaphor as such and instead take it literally.¹⁵ Once this happens, we could—in a Wittgensteinian way—say that our thought is (unconsciously) shackled by this very metaphor.¹⁶

In conclusion, a metaphor could be much more than just a poetic word or phrase. According to Lakoff and Johnson and the many scholars inspired by their theories, the *expressed* metaphor is sometimes merely secondary and can reflect a deeper *conceptual* metaphor; a way of (unconscious) metaphorical *thinking* or *reasoning*. One does not have to agree with Lakoff and Johnson that this theory applies to most abstract thinking in *general*, to see that it is at least true for (generative) AI. And that is precisely what matters to this article: GenAI is clearly structured systematically in anthropomorphic metaphors and, as we will see, this may affect legal analysis profoundly.

¹² C Greene and G Murphy, *Memory Lane* (Princeton University Press 2025).

¹³ See also: R Brette, ‘Brains as Computers: Metaphor, Analogy, Theory or Fact?’ (2022) 10 Front. Ecol. Evol. 4.

¹⁴ Of course this similarity is no coincidence, as the design of the AI model is made by humans who were inspired by their ‘own’ neural networks. Still this does not mean that any human fully understands any of both phenomena, nor does it mean that the two phenomena are the same.

¹⁵ This is an important aspect in which these AI metaphors differ from most legal ‘as if’ fictions, which they should not be confused with. Legal fictions are normally *explicitly* and *consciously* fictitious. A is treated *legally* as if it were B for practical/policy reasons, but there is no confusion to the fact that A is not actually B. The fiction is clearly not intended to depict reality. Think of corporate authorship (work for hire). The employing corporation is treated as if it is the actual author, even though it is clear that one of its employees actually made the work and not the legal person (which is also a *persona ficta*). The legislator merely communicates that the same norm should nonetheless apply to both A and B. Such legal fictions are not the object of this article. Which instead analyzes how we grasp a new technology metaphorically, and how this affects our perception of reality which then influences legal reasoning and argumentation. In these cases A is actually perceived or experienced as B and the metaphor could easily be taken literally. Certainly, GenAI might someday be regulated in an explicitly (analogous) fictitious manner with regards to copyright, but that is not what is happening right now. H Kelsen ‘Zur Theorie der juristischen Fiktionen. Mit besonderer Berücksichtigung von Vaihingers Philosophie des Als Ob’ (1919) 1 Annalen der Philosophie 638–640; C Kletzer (2015), ‘Kelsen on Vaihinger’ in M Del Mar and W Twining (eds) *Legal Fictions in Theory and Practice. Law and Philosophy Library* (Springer 2015) 25.

¹⁶ ‘Ein Bild hielt uns gefangen. Und heraus konnten wir nicht, denn es lag in unsrer Sprache, und sie schien es uns nur unberrittlich zu wiederholen’. L Wittgenstein, *Philosophische Untersuchungen* (1953) para 115. Available at https://www.wittgensteinproject.org/w/index.php/Philosophische_Untersuchungen#110 (accessed 2 September 2025).

⁵ ‘Metaphor’ OED Online (Oxford University Press). Available at https://www.oed.com/dictionary/metaphor_n?tl=true&tab=etymology (accessed 4 December 2025).

⁶ Lakoff and Johnson, *Metaphors We Live By* 52.

⁷ Lakoff and Johnson, *Metaphors We Live By* 145–146. ‘[...] changes in our conceptual system do change what is real for us and affect how we perceive the world and act upon those perceptions’.

⁸ Lakoff and Johnson, *Metaphors We Live By* 115.

⁹ For an elaborate technical explanation, see K Lee, AF Cooper and J Grimmelmann, ‘Talkin’ Bout AI Generation: Copyright and the Generative-AI Supply Chain’ (2023). Available at <https://doi.org/10.48550/arXiv.2309.08133> (accessed 20 November 2025).

¹⁰ S Ullman, ‘Using Neuroscience to Develop Artificial Intelligence’ (2019) 363 Science 692–693; R Brette, ‘Brains as Computers: Metaphor, Analogy, Theory or Fact?’ (2022) 10 Front Ecol Evol 3.

¹¹ R Brette, ‘Brains as Computers: Metaphor, Analogy, Theory or Fact?’ (2022) 10 Front. Ecol. Evol. 3.

III. Is GenAI training allowed under copyright?

A. Introduction

'So when you have something, when you read something and when it goes into this vast intelligence machine, we'll call it, you cannot expect to every time, every single time say, "Oh, let's pay this one that much. Let's pay this one." It just doesn't work that way. Of course, you can't copy or plagiarize an article, but if you read an article and learn from it, we have to allow AI to use that pool of knowledge without going through the complexity of contract negotiations, of which there would be thousands for every time we use AI'.¹⁷

Donald J. Trump (2025)

As set out, the anthropomorphic perception of GenAI is structural and omnipresent, extending from AI users and technicians to lawyers and politicians. The above quote is just one example among many. To say that 'neural networks' 'learn' or 'know' have become a very common expressions. *Metaphorical* expressions to be precise. Certainly, some people will disagree and claim that our brain is *literally* a computer or AI and that our way of learning is actually the same as 'machine learning'. After all, people have always been prone to define their brains and selves in terms of the newest technology (from telegraphy and telephones to Turing machines and modern AI).¹⁸ In this article, such a computationalist perspective is not adhered to. The human brain is a biological organ that is part of a living body. There is infinitely more to say on this topic, but for now I will simply state that, despite similarities with computers and AI—which are after all *brain-inspired*, *profound* differences remain.¹⁹

Therefore, although I realize that my semantic assumptions throughout this article could be rejected depending on one's worldview²⁰ and that these terms are subject to continuous (philosophical) debate²¹, I consider it a metaphor to say that AI *learns*. The following section then, examines what role this 'learning' metaphor plays in the legal analysis of AI training: is it allowed to train ('machine learn') AI models on copyright-protected works without the approval of the authors and rightsholders?

¹⁷ 'Transcript: Donald Trump's Address at 'Winning the AI Race' Event' (24 July 2025). Available at <https://www.techpolicy.press/transcript-donald-trumps-address-at-winning-the-ai-race-event/> (accessed 12 November 2025).

¹⁸ M Cobb, 'A Brief History of Wires in the Brain' (2022) 10 Front Ecol Evol. 1–3.

¹⁹ S Ullman, 'Using Neuroscience to Develop Artificial Intelligence' (2019) 363 Science 692–693; R Brette, 'Brains as Computers: Metaphor, Analogy, Theory or Fact?' (2022) 10 Front Ecol Evol. 3–4.

²⁰ What we consider to be a metaphor is highly dependent on our personal beliefs and convictions. For German and Dutch readers, this is easiest to explain: *beeldspraak* (Bildsprache), *denkbeeld* (Gedankenbild), *wereldbeeld* (Weltbild) and *mensbeeld* (Menschenbild) are deeply intertwined. In that sense it is no coincidence that *beeld/Bild* is present in all these words. For example: the word *schepping* or *Schöpfung* for some people represents a religious concept, reserved for god. Mere people therefore cannot 'schemen' or 'schöpfen': that would be metaphorical. Still, many artists (and lawyers) see this differently and use this term in relation to human artists and their creations. They do not consider this to be metaphorical. Ironically however, once we use this word in relation to generative AI these people are likely to object. According to many human artists only people 'schemen' (Create with a capital C), while AI merely 'generates' or 'produces'. This example illustrates that the same word can be experienced differently by different people: to some it is 'anthromonopolized' and to others it is not.

²¹ These discussions are hardly new, think only of the Turing test and the Chinese room argument: The Chinese Room Argument, *Stanford Encyclopedia of Philosophy* (Winter 2008 edn, rev 9 April 2014). Available at <https://plato.stanford.edu/archives/fall2018/entries/chinese-room/> (accessed 2 December 2025).

B. AI learns (and is fair use)

The first judicial decision in the USA and the world on commercial AI training was the result of a class action by Bartz and other authors against the AI company Anthropic, which trained its model Claude on the works of these authors without their permission.²² One of the arguments advanced by Bartz's lawyer was 'that using works to train Claude's underlying LLMs was like using works to train any person to read and write, so, authors should be able to exclude Anthropic from this use'.²³ To say that this argument backfired would be an understatement. Not only did judge Alsup refute this argument: he also kept on using it against its proponents. First of all:

Authors cannot rightly exclude anyone from using their works for training or learning as such. Everyone reads texts, too, then writes new texts. They may need to pay for getting their hands on a text in the first instance. But to make anyone pay specifically for the use of a book each time they read it, each time they recall it from memory, each time they later draw upon it when writing new things in new ways would be unthinkable. For centuries, we have read and re-read books. We have admired, memorized, and internalized their sweeping themes, their substantive points, and their stylistic solutions to recurring writing problems.²⁴

In explaining why he considered AI training to be 'quintessentially transformative' Alsup similarly resorted to the learning metaphor: 'Like any reader aspiring to be a writer, Anthropic's LLMs trained upon works not to race ahead and replicate or supplant them—but to turn a hard corner and create something different'.²⁵

With regard to the important fourth fair use factor—the effect of the use upon the potential market—Alsup's primary motivation again relied on the learning metaphor. In response to the 'generic' argument of the authors that LLMs will compete with them, even if no copies in the output emerge, the judge did not rule out this possibility but still invalidated their claim:

Authors' complaint is no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works. This is not the kind of competitive or creative displacement that concerns the Copyright Act. The Act seeks to advance original works of authorship, not to protect authors against competition.²⁶

Concludingly, teaching children to write well is out of copyright's reach—as long as the copies are legally obtained—and therefore teaching AI how to write is permitted as well. Such competition does not 'count' under the Copyright Act.²⁷ So the authors clearly lost the summary judgement in this aspect. As long as the training does not result in *direct* substitution—copies in the output—training is in principle fair use.

However, in this specific case, Anthropic also downloaded 7 million copies illegally to create a permanent library—inter alia for training purposes—and, according to Alsup, these illegal

²² United States District Court ND California, *Bartz v. Anthropic* (6/23/2025, No C 24-05417 WHA).

²³ United States District Court ND California, *Bartz v. Anthropic* 12.

²⁴ United States District Court ND California, *Bartz v. Anthropic* 12.

²⁵ United States District Court ND California, *Bartz v. Anthropic* 13–14.

²⁶ United States District Court ND California, *Bartz v. Anthropic* 28.

²⁷ United States District Court ND California, *Bartz v. Anthropic* 28.

downloads can never be redeemed by any subsequent fair use.²⁸ He treats these copies separately from those used for training the model. After the summary judgement, the plaintiffs therefore chose a different strategy and focused on the illegally obtained library that contained those copies even after training the AI model. The eventual USD 1.5 billion settlement is the result of this particular strategy and is primarily focused on the illegal database, instead of the trained AI model.²⁹

C. AI does not learn (and is unfair use)

Two days after this first judgement on generative AI and fair use, a second decision was delivered by Judge Chhabria in the case between Kadrey and Meta Platforms. This lawsuit too concerned authors of mostly books, on the one hand, and the provider of an LLM, on the other (Meta and its Llama-model).³⁰

Right at the beginning of his decision Chhabria challenged Alsup's ruling in *Anthropic v. Bartz*:

Speaking of which, in a recent ruling on this topic, Judge Alsup focused heavily on the transformative nature of generative AI while brushing aside concerns about the harm it can inflict on the market for the works it gets trained on. Such harm would be no different, he reasoned, than the harm caused by using the works for "training schoolchildren to write well," which could "result in an explosion of competing works." [...] According to Judge Alsup, this "is not the kind of competitive or creative displacement that concerns the Copyright Act." *Id.* But when it comes to market effects, using books to teach children to write is not remotely like using books to create a product that a single individual could employ to generate countless competing works with a miniscule fraction of the time and creativity it would otherwise take. This inapt analogy is not a basis for blowing off the most important factor in the fair use analysis.³¹

It is clear that Chhabria refuted Alsup's ruling primarily because of the learning metaphor (and analogy): teaching children is 'not remotely' the same as creating a product that can be used to generate countless competing works.

With regard to the first factor, Judge Chhabria—just like Alsup—considers the training of AI models transformative. Interestingly, *amici* law professors of Kadrey argued the opposite by comparing training AI with people who are reading. In their view, AI training has the same purpose as learning humans because 'we human beings are always learning from works, and are always learning patterns that are spread across multiple works'.³² And because: 'Our human use of copyrighted works to gain knowledge or to acquire skills has never been sufficient for a finding of fair use'.³³ If training AI is the same as human

learning, then training AI cannot be seen as a different purpose and is not transformative, so the argument went.

Chhabria rejected this learning metaphor as well because an LLM's 'consumption' of a book differs from human consumption: an LLM merely 'learns' 'statistical patterns' by a predictive process that is repeated billions or trillions of times.³⁴ Secondly, Meta copied books in order to create a tool which can then be used by anyone to generate any sort of text, and this is incomparable to teaching an individual person how to write. In conclusion, Chhabria consistently refuted the learning metaphor and ruled the training of GenAI to be transformative.³⁵

With regard to the fourth factor, Chhabria clarified that, in principle, training on copyrighted material will not be fair use, as long as the plaintiffs prove market substitution, which they failed to do in this particular case.³⁶ Relevant for future cases is that Chhabria also allowed the 'generic' approach that was rejected by Alsup, which is the so-called indirect substitution. Training on unauthorized copies is not just unfair in case copies or adaptations appear in the output (direct substitution) but also when the output itself is not infringing (substantially dissimilar).³⁷ According to him, such indirect substitution counts under the fourth factor because it dilutes the market of the original authors. In conclusion, Chhabria is not persuaded by the anthropomorphic argument that proved so decisive in the *Anthropic v. Bartz* case: training AI is not the same as a learning human. Relatedly, indirect substitution as a consequence of training *does* count under the Copyright Act and unauthorized training will in principle be unfair use.

D. AI-training is text and data mining (?) (EU)

The legal system in the European Union is quite different from the US: the EU lacks a flexible fair use framework and instead must abide by the system of exceptions as *inter alia* laid down in the Copyright Directive 2001.³⁸ Within this system, the only real possible exception for training commercially exploited AI models is the *text and data mining* (TDM) exception (art. 4 DSM-Directive³⁹).

The fiercely debated question in Europe, however, is whether this *mining* is the same as *machine learning* or *training*. Notice how all three could be considered metaphors. In European literature, compelling cases have been made that training with the aim of making a *generative* AI model is not TDM. These scholars stress that the AI 'learns' from the protected elements itself—the *expression*—and moreover that this results in a machine that will compete with the authors of those very works it was trained on.⁴⁰ In these important aspects training *generative* AI is different than traditional *analytical* TDM.

Nevertheless, early indications tell us that training GenAI should, in fact, be covered by the TDM exception. First, the AI Act⁴¹ states that the development and training of generative AI models should be in line with current copyright regimes and exceptions,

²⁸ United States District Court ND California, *Bartz v. Anthropic* 18–19.

²⁹ For their purposes, subsequent copies inside the model or output did not necessarily matter because statutory damages are awarded per work infringed upon; hence proving that the same work is copied multiple times does not increase compensation. E Lee, 'Is case v. OpenAI now Bartz II?' (28 October 2025). Available at <https://chatgptiseatingtheworld.com/2025/10/28/is-case-v-openai-now-bartz-ii-judge-stein-denies-openais-motion-to-strike-download-claim-from-shadow-library/> (accessed 1 December 2025).

³⁰ United States District Court ND California, *Kadrey v. Meta* (6/25/2025, Case No. 23-cv-03417-VC).

³¹ United States District Court ND California, *Kadrey v. Meta* 3.

³² Amicus Curiae Brief of Copyright Law Professors, in *Kadrey et al v. Meta Platforms, Inc.* (US District Court, Northern District of California, No 3:23-cv-03417-VC, filed 11 April 2025) 7.

³³ Amicus Curiae Brief of Copyright Law Professors, in *Kadrey et al v. Meta Platforms, Inc.* (US District Court, Northern District of California, No 3:23-cv-03417-VC, filed 11 April 2025) 7.

³⁴ United States District Court ND California, *Kadrey v. Meta* 17.

³⁵ United States District Court ND California, *Kadrey v. Meta* 17.

³⁶ United States District Court ND California, *Kadrey v. Meta* 39–40.

³⁷ United States District Court ND California, *Kadrey v. Meta* 26.

³⁸ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

³⁹ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

⁴⁰ TW Dornis and S Stober, *Urheberrecht und Training generativer KI-Modelle* (Nomos 2024); TW Dornis & N Lucchi, 'Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25' (2025) SSRN. Available at <http://dx.doi.org/10.2139/ssrn.5391439>.

⁴¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act).

the TDM exception in particular.⁴² Secondly, partly inspired by the AI Act, the first rulings on the training of Generative AI—*Kneschke v LAION* and *GEMA v OpenAI*—have stated that training GenAI is indeed covered by the TDM exception as long as all the criteria are met.⁴³ Most notably - in the case of commercial TDM - these criteria are lawful access and respecting the opt-out mechanism. Also, training may not result in permanent copies inside the trained model (see part IV).⁴⁴

The resulting status quo may be summarized as follows: TDM is machine learning, which in its turn is the same as GenAI training. In short: TDM = machine learning = training. Fully comprehending the (supposed) logic of this equation might require its own separate analysis, which I will not embark upon. Instead, I will focus on the legal effects of the equation of TDM and GenAI training. Which is after all what we care most about as jurists. Even if the equation of machine learning and TDM would be uncontested: discussion and litigation will probably still arise with regard to the question of whether the TDM exception aligns with the three-step test of the Berne Convention, which is also stipulated in, inter alia, the Copyright Directive.⁴⁵ After all, the limitations on copyright can be limited by this three-step test.

It is in this specific context of the TDM exception and the three-step test that we might see similar metaphorical arguments as used in the USA. Especially the second criterion of normal exploitation is relevant. Is indirect substitution—without proven copies in the output—relevant to the normal exploitation? It seems that a causality problem could easily arise for the individual copyright holder. How to prove that the normal exploitation of a work or oeuvre is threatened as a consequence of the training on this particular work if the model has also been trained on millions or billions of other works?⁴⁶ In other words: the scale advantage of AI companies—by virtue of the sheer size of their training datasets—creates a causality problem for the individual copyright holder.⁴⁷

Finally, an important question remains: how far along the 'AI supply chain'⁴⁸ does this exception 'go'? Does the TDM exception only justify the (temporary) copies before and during the training process, or might compressed copies inside the trained model itself too be justified? This matter will be discussed in the next section on memorization (part IV).

E. Rhetorical analysis

First, as Judge Chhabria speaks of analogy instead of metaphor, one might ask why we are talking about *metaphors* to begin with. In practice, the distinction between metaphor, analogy and other literary figures⁴⁹ is often difficult to make because they may

largely overlap (and there are different competing definitions).⁵⁰ In the aforementioned examples—concerning artificial neural networks—this is even more so: here, the relation of analogy and metaphor could be thought of as a Droste effect or a Gordian knot that cannot be untangled. Also, for the purposes of this legal article, it is trivial what exact rhetorical figures or devices are used to humanize AI. What matters foremost is the fact that AI is anthropomorphized, because this is what may affect legal analysis and decisions. Still, metaphors are the most logical angle for this article on AI. Because many (supposed) AI analogies are actually structured in metaphorical (human) terms, 'analogical' reasoning is far from immune to these metaphors and may have the exact same effect as a 'normal' metaphor. Even in deductive reasoning—or that what is presented as such—tenacious metaphors like the neural network may resonate and affect decisions profoundly.⁵¹

Now let us have a look at this 'AI is a brain' metaphor and its role in the *Anthropic v. Bartz* ruling. The first premise is that learning in itself is free: out of copyright's reach. The second premise is that the training of AI is learning. Which leads to the conclusion that AI training is out of copyright's reach too. Rightsholders cannot oppose to AI training, just like they cannot oppose learning itself. However, the second premise is a metaphor: it is not true, and therefore the conclusion cannot be valid. Still, apparently, the judge sees this differently and to him the second premise is not metaphorical but something literal he builds his judgement on.⁵² And metaphorical 'reasoning' does not end here: after this human-to-AI metaphor—AI learns—the judge turns the metaphor around to uphold his judgement in a rather circular manner.

If one were to apply copyright to *teaching* (and learning) AI, then by the same reasoning *schoolchildren* too would have to pay in order to learn how to read and write. And of course, it would be outrageous if humans had to pay for each time they read a book or learn a thing. In this way, the reverse metaphor (AI-to-human) is used to facilitate a *reductio ad absurdum*. Arguments like this—let's call it a 'reverse reductio'—are related to the double-edged nature of the brain metaphor (see part II), which also explains their recurrence in legal AI debates.

In *Meta v. Kadrey*, the judge explicitly refutes the metaphors (and follow-up analogies) used by Alsup but also by other parties in the lawsuit. Chhabria does not reason metaphorically from human to AI (or vice versa) and this affects his legal judgement: training AI is *not* learning and, in principle, *not* fair use.

⁵⁰ For example, in a simile or analogy one might say that A is like B or A is as B, however these comparisons are often not expressing literal similarities either, but rather consist of one or more metaphorical terms. With regard to GenAI 'analogies' are often structured metaphorically (in anthropomorphic terms). Saying that AI training is 'like' or 'as' a child learning from reading books or 'no different than', is not a literal comparison and still transfers meaning from the human domain to the AI domain (just like a 'normal' metaphor: 'A is B'). The comparison is after all structured in human terms, and one could even argue that these metaphorical words in themselves contain an implicit analogy (the ancient definition of a metaphor). In its turn the analogy may be part of the larger conceptual 'AI is a brain' metaphor. On the inseparability of analogy and metaphors in cases like this: F van Besien, 'Metafoor: wiens begrip is het eigenlijk?' *De Nieuwe Taalgids* (1993) 199–201.

⁵¹ An (unconsciously) metaphorically structured *analogy* may have the exact same effect as a metaphor. Also deductive reasoning may be based on a metaphorical premise: AI learns, knows or memorizes, eg. So, in all these cases what matters most is the structural metaphor of AI as a neural network which then strongly affects legal reasoning, regardless of the nomenclature and label that is given to this category of reasoning.

⁵² It might be so that this judge does not actually believe machine and human learning to be exactly the same, but that they are similar enough to be treated the same *legally*. But, in that case, he should have motivated this more clearly. In this case, he did not and apparently relied mostly on this metaphor.

⁴² AI Act Art. 53–1 sub c. See also recital 105.

⁴³ Landgericht Hamburg 27 September 2024 *Kneschke v. LAION* para 79–80. This case concerned the TDM exception for scientific research (art. 3 of the DSM-Directive).

⁴⁴ Landgericht München I 11 November 2025 *Gema v. OpenAI*.

⁴⁵ Berne Convention art. 9 (2); Directive 2001/29/EC art. 5 (5). For example, see: T Dornis and S Stober, Urheberrecht und Training generativer KI-Modelle, (Nomos 2024) 143–148; M Senftleben, 'TDM, GenAI and the Copyright Three-Step Test' SSRN 2025 1–3; DJG Visser, 'AI en Auteursrecht medio 2025' (2025) *Bedrijfsjuridische berichten* 5.

⁴⁶ See: M Senftleben, 'TDM, GenAI and the Copyright Three-Step Test' SSRN 2025 14–16.

⁴⁷ MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' *IER* 2025/6 footnote 100.

⁴⁸ On the supply chain, see K Lee, AF Cooper and J Grimmelmann, 'Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain' (2023) *Arxiv*. Available at <https://doi.org/10.48550/arXiv.2309.08133> (accessed 20 November 2025).

⁴⁹ Think of: catathesis, paraphemia, hypocatastasis, antonomasia, allegory, parable, etc.

In Europe, no judgement has yet been rendered on the commercial training of a GenAI model specifically. However, it is possible that in future judgements, we will see similar arguments and rhetoric with relation to TDM and the three-step test.⁵³

IV. Do GenAI models copy?

A. Introduction

The previous section concerned the mere training of AI, not the resulting trained *model* itself. After all, it is often assumed that AI training does not result in any permanent copies inside this trained (and deployed) AI model. The training (machine learning) phase is like a distillation process: only semantic relations, abstractions and other unprotected elements are distilled from the works. The works as such or copyright protected parts on the other hand are not retained after training, hence such 'clean' training does not result in any copies inside the model. The trained model is therefore out of copyright's reach and cannot carry copies of the training data.

However, this narrative is challenged seriously by scientific research on so-called memorization. An AI model 'memorizes' when it 'encodes details from its training data, such [that] it is capable of generating outputs that closely resemble its training data'.⁵⁴ 'Memorization' is deemed to exist when the (datafied) work can be discovered (indirectly) inside an AI model, usually by means of a prompt.⁵⁵ For example by extracting a poem from the model into the output by using a simple prompt (such as the author of the poem). At this moment, there is plenty of scientific evidence of such 'memorization'.⁵⁶ Still, the question that remains is whether this phenomenon qualifies as *copying*. This section will deal with this highly relevant question which is to be separated conceptually from the question in the previous section, that is whether 'clean' training—without copies inside the model—is allowed.⁵⁷ After all, this memorization debate challenges the very premise that AI training is 'clean' to begin with.

B. AI models memorize

According to prominent AI companies, trained models do not contain copies of works but merely consist of 'weights' or 'parameters' which in turn are nothing but numbers and statistical patterns.⁵⁸ For instance, Google wrote to the US Copyright Office:

The interconnections between these neurons, also referred to as nodes, are numerical weights that essentially represent the importance of the contribution of that neuron to the final output. It is also relevant to note that there is no copy of the training data—whether text, images, or other formats—present in the model itself. Deep neural networks themselves determine the attributes of the data that they use to recognise patterns, as opposed to a human coder setting those attributes manually.⁵⁹

Other big AI companies have made highly similar arguments. For example, OpenAI claims that their model only consists of a 'large string of numbers' that reflects 'statistical relationships'.⁶⁰ It can therefore not copy: 'Much like a person who has read a book and sets it down, our models do not have access to training information after they have learnt from it'.⁶¹

If, for some reason, a work *seems* to be copied inside the model because a work appears in the output after a simple prompt, then this is referred to as an unintended glitch and the phenomenon is still not called copying or storing but 'memorization'.⁶² This anthropomorphic metaphor conveniently and neatly fits in the broader systematic metaphor of the AI model as a neural network (part II).

Both *atomizing* the AI model into unprotected parts and abstractions and *anthropomorphizing* the model into a 'neural network' that 'memorizes' and 'learns' have proven effective in deterring lawyers from using language such as storage or copying.⁶³ Especially, because these two rhetorical routes are sometimes combined: just like *human* neural networks artificial neural networks would merely memorize patterns without copying.

So, instead of 'copying', lawyers too speak of remembering, memorizing or knowing. Take for example Guadamuz's comment on the 'memorization' issue in *Disney v. Midjourney*: 'But Midjourney does not store those images and serves them, image models know what a stormtrooper looks like and can reproduce it. Is it storage? I don't think so'.⁶⁴ And in response to the *Getty v. Stability* case: 'That is indeed the case, models can memorize, but as I have been reminding people endlessly, memorisation is not an exclusive right of the author (otherwise memorizing a poem could land you in court). Reproduction is an exclusive right of the author'.⁶⁵

⁵³ The *Kneschke v. Laion* case dealt with the scientific TDM exception (although this was controversial) and the *GEMA v. OpenAI* case mostly concentrated on copies inside the trained model ('memorisation') and infringement in output.

⁵⁴ AF Cooper, K Lee, J Grimmelmann and others, *Report of the 1st Workshop on Generative AI and Law* (2023) Yale Law & Economics Research Paper 5.

⁵⁵ GenLaw, 'Memorization' (GenLaw Glossary). Available at <https://blog.genlaw.org/glossary.html#memorization> (accessed 1 November 2025); AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 Chi-Kent L Rev 159–161.

⁵⁶ AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 Chi-Kent L Rev; N Carlini, 'What my privacy papers (don't) have to say about copyright and generative AI' (11 March 2025). Available at <https://nicholas.carlini.com/writing/2025/privacy-copyright-and-generative-models.html> (accessed 1 October 2025); A. F Cooper and others, 'Extracting memorized pieces of (copyrighted) books from open-weight language models' (2025) Arxiv. Available at <https://arxiv.org/abs/2505.12546> (accessed 2 December 2025) 2: 'Successful extraction of a piece of training data in a model's generation is evidence of memorization of training data inside the model. This is uncontroversial in machine learning (ML). As Carlini explains, when a sufficiently large and unique piece of training data is extracted, "the only possible explanation is that the model has somewhere internally stored [that piece of training data]"'.⁵⁷

⁵⁷ It might very well be the case that memorization/overfitting and generalization are not really technically distinguishable in the training process. Still, for a legal analysis, I deem it useful to separate these questions conceptually.

⁵⁸ USCO, *Part 3: Generative AI Training* (pre-publication) 19; AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 Chi-Kent L Rev 186–187.

⁵⁹ Google Comments submitted to USCO, 1 November 2023. Available at <https://www.regulations.gov/comment/COLC-2023-0006-9003> (accessed 10 August 2025) 3–5.

⁶⁰ OpenAI Comment submitted in Docket No COLC-2023-0006 on [Regulations.gov](https://www.regulations.gov): Artificial Intelligence and Copyright (comment ID COLC-2023-0006-8906). Available at <https://www.regulations.gov/comment/COLC-2023-0006-8906> (accessed 1 September 2025). This is highly similar to OpenAI's arguments in: *Landgericht München I* 11 November 2025 *Gema v. OpenAI*.

⁶¹ OpenAI Comment submitted in Docket No COLC-2023-0006 on [Regulations.gov](https://www.regulations.gov): Artificial Intelligence and Copyright (comment ID COLC-2023-0006-8906). Available at <https://www.regulations.gov/comment/COLC-2023-0006-8906> (accessed 1 September 2025) 6.

⁶² AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 Chi-Kent L Rev. 216–217.

⁶³ On 'atomizing' the model, see: MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6. On atomizing previous copyright relevant technologies, see: T Dreier and G Schulze, *Kommentar zum Urhebergesetz* 2025 art. 16–23; G Schulze, 'Werkgenuss und Werknutzung in Zeiten des Internets', NJW 2014 722.

⁶⁴ A Guadamuz, 'Disney and Universal sue Midjourney for copyright infringement' TechnolLama (12 June 2025). Available at <https://www.technollama.co.uk/disney-and-universal-sue-midjourney-for-copyright-infringement> (accessed 1 September 2025).

⁶⁵ A Guadamuz, 'Getty Images v Stability AI: A landmark High Court ruling on AI, copyright, and trade marks' TechnolLama (6 November 2025). Available at <https://www.technollama.co.uk/getty-images-v-stability-ai-a-landmark-high-court-ruling-on-ai-copyright-and-trade-marks> (accessed 10 November 2025).

Here is another subtle but clear example of a reverse metaphor, facilitating a *reductio ad absurdum*: memorization cannot be infringement, because if it was then memorizing people too would be infringing. Meaning our human brain would be infringing, which of course would be absurd.

This seems to be part of a broader reluctance among certain lawyers to apply 'mechanical' language—like copying—on neural networks.⁶⁶ The 'neural networks' are so systematically humanized that mechanical terms such as storing and copying are replaced with mental terms like knowing, remembering and memorizing.⁶⁷

All of this raises the question if artificial neural networks indeed 'remember', 'memorize' and 'know' or whether it simply copies and stores information. After all: copying is something copyright can forbid, but *human* remembering is (luckily) out of its reach. In the following, I will try to shed some light on this question.

C. AI models do not memorize

First of all, artificial 'neural networks' are not actually neural networks. To some this may seem obvious but, because of the omnipresence of anthropomorphic metaphors, it is necessary to emphasize certain differences between biological and artificial neural networks. They are, after all, relevant to our copyright analysis of *inter alia* 'memorization'.⁶⁸

The human brain is a *living* organ that *continuously* learns and changes. It consists of approximately 86 billion neurons interconnected by more than one hundred trillion constantly adapting (synaptic) connections.⁶⁹ Plasticity is an essential feature of the brain, and concomitantly, human memory is inherently fallible and incomparable to the reliable mechanical *storage* of works on a hard drive or computer.⁷⁰ In short, the brain does not function like a (compressed) database or storing device. We may read *Harry Potter and the Sorcerer's Stone* quickly or thoroughly, but that will not result in fixed numbers in our brain permanently encoding the entire book (near) verbatim.⁷¹ In fact, if our brain had worked like that copyright would probably never have come into existence or at least in a radically different form.⁷²

In this sense, a trained artificial neural network that 'memorizes' and a human brain differ in a way that is relevant to copyright law.⁷³ Unlike human beings, AI models do not 'learn' and

change throughout their 'deployment'. During its training phase, the weights may be conditioned on vast quantities of data. However, once that training phase is completed, the model is static.⁷⁴ The numerical values of its parameters are fixed and stable until and unless the model is retrained.⁷⁵ Such (full) retraining is relatively rare as it is both expensive and risky.⁷⁶ After all, new training ('learning') carries the serious risk that previously acquired information may be (completely) overwritten—commonly referred to as *catastrophic interference*.⁷⁷

Furthermore, it is clear that the possible 'deployments' of an AI model differ drastically from a human brain. Unlike a biological neural network, an artificial neural network can be deployed, published and exploited commercially on a global scale. Open-source AI models like Stable Diffusion have been downloaded innumerable times, and Llama-models have even surpassed 1 billion downloads.⁷⁸ The most common alternative is its deployment in a software system like Co-Pilot or ChatGPT. The user then interacts with the system directly—and indirectly with the model—while normally paying for the usage of both (either with money or data).

D. AI models copy

There is a lot we do not know about 'memorization', but what we do know is that trained GenAI models, in some way, 'encode' *some*⁷⁹ specific data they are trained on.⁸⁰ We also know that AI 'memorization' is not the same as memorizing *humans*. It is a metaphor used in the absence of a more accurate description (our language simply falls short).⁸¹ So what should we call it? This section considers the case for qualifying 'memorization' as copying.

According to Cooper and Grimmelman the 'patterns' inside a model—fixed in the parameters/weights—can be so specific and

human-like than computers or classic storing devices, but that still does not mean that it is the same as human memorisation. As a matter of fact, it is relevant to realize that questions like these can be answered on different 'levels of analysis' (hardware, processes and function) and that once again (unconscious) (philosophical) assumptions about the mind and brain play an important role. RP Cooper & D Peebles (2026) 'Levels of Analysis' in MC Frank & A Majid (eds), *Open Encyclopedia of Cognitive Science* MIT Press. Available at <https://doi.org/10.21428/e2759450.f61797a0> (accessed 15 January 2026).

⁷⁴ Jack Bell and others, 'The Future of Continual Learning in the Era of Foundation Models: Three Key Directions' (2025) Arxiv. Available at <https://arxiv.org/abs/2506.03320> (accessed 3 December 2025).

⁷⁵ USCO, Part 3: *Generative AI Training* (pre-publication) 29; MS Bennett, A *Brief History of Intelligence*, London 2023 132.

⁷⁶ This also explains 'knowledge/data cutoffs', the model does not 'learn' anymore after its training. Thereby underlining the point that AI actually never learns or knows to begin with. My focus on the AI model might evoke a 'system reply'—somewhat reminiscent of the Chinese Room Argument discussion—that the AI system as a whole does in fact memorize. I still find that unconvincing though: the model is an essential part of the system and system level techniques like RAG do not change or retrain the underlying model.

⁷⁷ And of course there is a more anthropomorphic synonym: catastrophic forgetting. MS Bennett, A *Brief History of Intelligence*, London 2023 131–132.

⁷⁸ 'Celebrating 1 Billion Downloads of Llama' (18 March 2025) Meta. Available at <https://about.fb.com/news/2025/03/celebrating-1-billion-downloads-llama/> (accessed 9 December 2025).

⁷⁹ Estimates range from 0.1 percent to 10 percent of the training data. Even a small percentage could mean an enormous amount of works though. AF Cooper and J Grimmelman, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 *Chi-Kent L Rev.* 202–203.

⁸⁰ N Carlini, 'What my privacy papers (don't) have to say about copyright and generative AI' (11 March 2025). Available at <https://nicholas.carlini.com/writing/2025/privacy-copyright-and-generative-models.html> (accessed 1 October 2025); AF Cooper and others, 'Extracting memorized pieces of (copyrighted) books from open-weight language models' (2025) Arxiv. Available at <https://arxiv.org/abs/2505.12546> (accessed 2 December 2025) 1.

⁸¹ One might argue that memorisation has evolved into a *homonym*. However this is not very convincing. After all, for a homonym it should be clear from the grammatical context that the same syntactic form has a different semantic meaning. In this case, this is not clear at all: in fact many people seem to think memorizing is (effectively) the same as humans remembering. Which is precisely the case because it is part of the broader and structural 'AI is brain' metaphor.

⁶⁶ Eg LG Hamburg 27 september 2024 Kneschke versus Laion, para 72; J Torenbosch, 'Memorization en generatieve AI' (2025–2) Auteursrecht.

⁶⁷ See the *Anthropic v. Bartz* case as well for 'memorization': United States District Court ND California, *Bartz v. Anthropic* 12–13.

⁶⁸ Note: this part does not deal with 'generalization', see part V for that.

⁶⁹ MS Bennet, A *Brief History of Intelligence* (London 2023) 5; 'Lifelong learning and the plastic brain' (2024). Available at <https://www.cam.ac.uk/research/features/lifelong-learning-and-the-plastic-brain> (accessed 10 September 2025); MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6 436.

⁷⁰ Human memories are likely to be (unconsciously) modified with each reactivation and may even be completely forgotten or fabricated. C Greene and G Murphy, *Memory Lane* (Princeton 2025) 21 & 26–28. On neuroplasticity, also see S Miller, 'Brains are Made of Plastic' (2013) Brainfacts. Available at <https://www.brainfacts.org/brain-anatomy-and-function/anatomy/2013/brains-are-made-of-plastic> (accessed 5 August 2025); 'Lifelong learning and the plastic brain' (2024). Available at <https://www.cam.ac.uk/research/features/lifelong-learning-and-the-plastic-brain> (accessed 10 September 2025).

⁷¹ This entire book of JK Rowling is 'memorized' 'near-verbatim' by a Llama model. AF Cooper and others, 'Extracting memorized pieces of (copyrighted) books from open-weight language models' (2025) Arxiv. Available at <https://arxiv.org/abs/2505.12546> (accessed 2 December 2025) 165.

⁷² In such a world no one would need (physical) copies instead you would enter a library and quickly scan all the books you might ever need. In this scenario, one would only have to pay for accessing a library.

⁷³ S Ullman, 'Using Neuroscience to Develop Artificial Intelligence' (2019) 363 *Science* 692–693; A Salles and others, 'Anthropomorphism in AI' (2020) 11 *AJOB Neuroscience* 92. Certainly such AI models store information much more

detailed that they contain protected elements and works in a compromised and latent manner.⁸² At least in such a way that it is possible to reconstruct the training data in the output. As long as the user of the AI system does not provide or assemble the protected expression himself, and it is not provided by a non-parametric source such as RAG either.⁸³ then the only logical explanation is that the copy in the output is an offshoot of the 'statistical' copy inside the model, which functions as a master copy. For example, when I merely mention the name 'Uderzo' as a text prompt, the AI system Sora immediately and repeatedly presents me with perfect images of Asterix and Obelix. This would mean that the model encoded the protected expression of these characters inside the parameters of the model, which the system can then reconstruct in the output, making the copy available and perceivable to its user.⁸⁴

At this moment, it is worth learning from the past. In the early 20th century, in both the USA and Germany, manufacturers of mechanical music reproductions like the piano roll claimed successfully that their products—the rolls—could not qualify as reproductions of compositions. After all, the work could not be perceived directly in the roll itself: it was merely a roll with perforations which—unlike sheet music—was on its own incomprehensible to humans. A technical device like the piano player was necessary to hear the music and consume the work. Using these arguments, producers of mechanical music reproductions managed to initially evade composers' copyrights so that they did not have to pay royalties.⁸⁵

To correct this discrepancy in copyright protection—music composers could oppose old school music sheet copies but not the more popular mechanic reproductions—the Berne Convention in 1908 drafted a new Article 13,⁸⁶ stating: 'Les auteurs d'œuvres musicales ont le droit exclusif d'autoriser: 1. l'adaptation de ces œuvres a des instruments servant a les reproduire mécaniquement'.⁸⁷

Thus, the Berne Convention took an important step towards a technologically neutral reproduction right; any instrument that could mechanically function as a reproduction of the musical work was a copy. The work did not have to be directly perceivable to human senses: indirect perceptibility—by using a technical aid like a gramophone for instance—could fulfil the definition of reproduction. To this day, the article survives in modernized form in, inter alia, the Dutch Copyright Act.⁸⁸

More importantly, these principles of technological neutrality have been absorbed in the legislation and case law of jurisdictions in, inter alia, the USA, Germany and the Netherlands.⁸⁹ In fact, this goes for all EU Member States as the reproduction right is generally harmonized in the Copyright Directive as follows: 'Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part'.⁹⁰ This is also what prompted the Regional Court in Munich to rule that OpenAI's language model contains copies of several German song texts.⁹¹ Nevertheless, for a final answer on whether AI models can indeed contain compressed copies of training data, we will have to wait for a ruling of the CJEU.⁹²

If the CJEU indeed confirmed this approach, such copies could not be remedied by the TDM exception (or any other). After all, the TDM exception only amounts to the training process of models. Therefore, it seems highly unlikely that (latent) permanent copies inside the trained model can be justified by this exception.⁹³ The TDM exception can only be applicable to those preparatory reproductions before and during the training, not those that are retained thereafter in the model's parameters (albeit in a different (mathematical) language).⁹⁴ The possible subsequent copy in the output—also a making available to the public—is even less likely to be justified by this exception.⁹⁵

Finally, it is worth briefly mentioning the UK decision in *Getty v. Stability AI*, which has led to some confusion. Although the judge ruled that the AI model Stable Diffusion did not contain copies, and was not a copy in itself, it is important to note that Getty eventually abandoned several claims and *did not contend* that memorization had occurred: in fact, it did not claim that any specific work was copied inside the model.⁹⁶ Relatedly, the fact that the model of Stability AI was not trained in the UK but in the USA affected the outcome.⁹⁷ For these reasons, instead of lingering on this specific UK ruling, it is probably more useful to

never meant 'intended' or purposefully; the object should merely function as a reproduction, just as in the French original text of the Berne Convention ('servant'). Hence, the reproduction could even be accidental. As an AI model is an object—and not actually a living organ like the brain—and can apparently make copyright protected works visible in the output, the model—or parameters in it—can qualify as reproduction in the sense of art. 14. At least that is what I recently argued in a Dutch publication. MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6.

⁸⁹ This principle was stipulated in art. 13 (1) of the Berne Convention until 1967, after which it was transferred to the new reproduction provision, see art. 9 (1) and (3).

⁹⁰ Directive 2001/29/EC art. 2.

⁹¹ Landgericht München I 11 November 2025 *Gema v. OpenAI* para b.aa (p. 36).

⁹² The *Like Company*-case is especially relevant: CJEU, case C-250/25, *Like Company v. Google Ireland*, preliminary reference lodged on 3 April 2025.

⁹³ See: Directive 2019/790 art. 4 (2); 2001/29/EC art. 2.

⁹⁴ I have argued this in: MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6440. This was confirmed in the first memorization case: Landgericht München I 11 November 2025 *Gema v. OpenAI* 42–44.

⁹⁵ See: E Rosati, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' (2025) *European Journal of Risk Regulation* 16 (2) 619; TW Dornis & N Lucchi, 'Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25' (14 August 2025) SSRN. Available at <http://dx.doi.org/10.2139/ssrn.5391439> (accessed 20 October 2025) 35–40.

⁹⁶ *Getty Images v Stability AI* [2025] EWHC 2863 (Ch), para 559: '[...] it is important to be absolutely clear that Getty Images do not assert that the various versions of Stable Diffusion (or more accurately, the relevant model weights) include or comprise a reproduction of any Copyright Work and nor do they suggest that any particular Copyright Work has been prioritised in the training of the Model'. And para 560: 'Getty Images' case is that they do not need to show any of this; they say it is enough that the making of the model weights (had it been carried out within the UK) would have constituted an infringement of copyright.

⁹⁷ *Getty Images v Stability AI* [2025] EWHC 2863 (Ch), para 755.

⁸² AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 *Chi-Kent L Rev.* 168–169 & 186–187.

⁸³ Retrieval-augmented generation.

⁸⁴ The USCO and Munich Regional Court also acknowledge this possibility of internal copies inside the model. MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6; USCO, *Part 3: Generative AI Training* 2025 (pre-publication) 19–21 and 28–29.

⁸⁵ *White-Smith Music Pub. Co. v. Apollo Co.*, 209 U.S. 1 (1908), 'A "copy" of a musical composition within the meaning of the copyright statute is a written or printed record of it in intelligible notation, and this does not include perforated rolls which, when duly applied and properly operated in connection with musical instruments to which they are adapted, produce the same musical tones as are represented by the signs and figures on the copy in staff notation of the composition filed by the composer or copyright'; K Raustiala and CJ Sprigman, 'How a terrible Supreme Court decision about player pianos made the cover song what it is today'. Available at <https://slate.com/technology/2014/05/white-smith-music-case-a-terrible-1908-supreme-court-decision-on-player-pianos.html> (accessed 20 August 2025); M de Cock Buning, *Auteursrecht en Informatietechnologie* (Amsterdam: Otto Cramwinckel, 1998) 115–117; MA Smit 'Pianola's, platen en patronen: de betekenis van art. 14 Aw in het kunstmatige tijdperk' IER 2025/6 430–431.

⁸⁶ S Ricketson and JC Ginsburg, (3rd edn Oxford 2022) *International Copyright and Neighbouring Rights: The Berne Convention and Beyond* para 3.17.

⁸⁷ Berne Convention (1908), art. 13.

⁸⁸ Dutch Copyright Act art. 14: 'The reproduction of a literary, scientific or artistic work includes the fixation of this work or a part thereof on any object destined to make the work visible or audible'. The word destined has

shift our attention to future cases in the USA, which is where most actual ‘memorization’ cases are pending.⁹⁸

E. Rhetorical analysis

‘Atomizing’ the model into unprotected strings, weights, vectors and other incomprehensible technical jargon may be neutralized by appealing to the technological neutrality of the reproduction right. If the reproduction right is indeed as radically technological neutral as contended in the literature and *GEMA v. OpenAI*, then there is only one more rhetorical way out for AI companies: bringing AI to life. At least in the minds of lawyers who might believe the metaphors that AI indeed memorizes, remembers or knows.⁹⁹

This may result in similar reciprocal metaphorical reasoning as seen in *Anthropic v. Bartz*. If the judge actually believes the neural network to be (essentially) the same as the human brain, then this metaphorical premise may lead the judge to the conclusion that the AI model (neural network) like humans merely remembers and by definition does not copy or store works.

After all, as we would all agree upon: we cannot speak of copies inside the *human* brain. At least not in any manner that is copyright-relevant: a brain cannot be infringing. This *reverse reductio*—memorizing AI cannot be infringing because if it did, then so would memorizing humans—may be rhetorically convincing but scientifically it is not. In fact, it reflects a double misunderstanding of both the human brain and AI models.¹⁰⁰ Whether ‘memorization’ counts as copying is debatable, but what is clear is that AI models cannot memorize/remember like humans. And equally important, people do not store information or works like overfitting AI models.

The underlying copyright question is what the supposed ‘technological neutrality’ of the reproduction right actually means in relation to GenAI. Is it necessary to *understand* the exact technological procedure in which information is (allegedly) retained inside a model before we can speak of copies? Or does a ‘functional’ presence and a convincing causal relationship between the model and infringing output suffice? In that case, we could say that the technological neutrality of the reproduction right is *radical* and knows no *technological* boundaries. If, on the other hand, a judge demands that the supposedly copied work is actually pinpointed *inside the model*: demanding *direct* proof of its presence and a technological explanation, then copies cannot *legally* be present inside the model. After all: exactly pinpointing the work is—for now—probably impossible because of the brain-inspired and distributed way information is stored across different parameters. In such a scenario, the opacity and complexity of AI models is ‘rewarded’ with immunity against liability.¹⁰¹

⁹⁸ For example: *Disney v. Midjourney* and *NYT v. OpenAI*.

⁹⁹ Another possibility is the futuristic scenario in which the metaphorical hypothesis of AI is actually ‘fulfilled’, meaning that anthropomorphic metaphors have become reality and are therefore no longer metaphorical. *Organoid Intelligence* (OI) is an interesting field of research in this aspect. ‘Organoid Intelligence: A New Biocomputing Frontier (Explainer)’ (2023) *Frontiers in Science*. Available at [https://www.frontiersin.org/journals/science/article-hubs/organoid-intelligence-a-new-biocomputing-frontier/explainer?delimiter=026E30F\\$](https://www.frontiersin.org/journals/science/article-hubs/organoid-intelligence-a-new-biocomputing-frontier/explainer?delimiter=026E30F$) (accessed 24 August 2025).

¹⁰⁰ H Smit and PMS Hacker ‘Seven Misconceptions About the Mereological Fallacy: A Compilation for the Perplexed’ (2014) *Erkenn.* 79, 1085; R Brette, ‘Brains as Computers: Metaphor, Analogy, Theory or Fact?’ (2022) 10 *Front Ecol Evol.* 3.

¹⁰¹ At least with regard to the model itself.

V. Is AI-generated output copyrightable?

A. Introduction

The two previous questions both concerned the infringing side of copyright and AI. This section in contrast will deal with the ‘creative’ side: the question of whether the output of GenAI can be copyright-protected. It is important to distinguish this question from the previous section which dealt with the technical feature known as ‘memorization’. In such cases ‘generative’ AI is actually not purely ‘generating’ new content, but (partly) reproducing training works (‘MemAI’ instead of ‘GenAI’). This section, in contrast, will deal with the more common ‘creative’ or ‘generative’ feature of AI models (*generalization* instead of *memorization*).¹⁰² This could also be seen as the more human-like feature of GenAI or even ‘synthetic creativity’.¹⁰³

B. GenAI is a tool

One common school of thought is that generative AI is merely the latest *tool* or *instrument* an artist may use to make creative and copyright-protected works with.¹⁰⁴ This is most clearly visible in Chinese case law. The first and most important ruling concerned a dispute between Li and Liu (2023) involving an image ‘created’ with Stable Diffusion using only a textual prompt.¹⁰⁵

According to the Beijing Internet Court, the AI-generated output was protected, for ‘in essence, it is still humans that use tools to create’.¹⁰⁶ Therefore the court asserted that the output constituted an ‘original’ and ‘intellectual achievement’ of the human user who they regarded as the legitimate author. Of course, by stating this, the court presumed that the output actually reflected the human creative choices, which could only be the case if one indeed considers GenAI a passive tool that does not generate independently from the user.

The court elaborated on this as follows: ‘Before the advent of cameras, people needed superb painting skills to reproduce an object perfectly; then the cameras made it easier to record the image of an object. Nowadays, the camera of smartphones is getting better and easier to use. However, as long as the photos taken with a smartphone reflect the photographer’s original intellectual investment, they will constitute photographic works and are protected by the Copyright Law’.¹⁰⁷

In subsequent cases, Chinese judges have followed the same reasoning and granted copyright to visual AI output that was generated in response to mere textual prompts.¹⁰⁸ Only in one particular case copyright was denied; however, this was mostly due to the fact that the AI user had not saved her prompts.¹⁰⁹ In

¹⁰² From a technical point of view these two phenomena are not (always) separable from one another. Nor are they exact opposites of each other. F Chollet, *Deep Learning with Python* (New York 2021) 122; ‘Generalization’ *GenLaw Glossary*. Available at <https://blog.genlaw.org/glossary.html#generalization> (accessed 12 September 2025); USCO, Part 3: *Generative AI Training* (pre-publication) 20.

¹⁰³ On synthetic creativity, see D Burk, ‘Cheap Creativity and What It Will Do’ (2023) *Georgia Law Review* 1669.

¹⁰⁴ These scholars also might speak of ‘technological neutrality’ on the ‘creative’ side—not to be confused with the technological neutrality of the reproduction right. In contrast to the latter, the foundations of this principle on the creative side are more obscure (at least in Europe).

¹⁰⁵ Beijing Internet Court Civil Judgement of 27 November 2023 Jing 0491 Min Chu No. 11279.

¹⁰⁶ ‘Copyright Protection for “AI-Generated” Images’ (2024) 73 *GRUR International*.

¹⁰⁷ ‘Copyright Protection for “AI-Generated” Images’ (2024) 73 *GRUR International*.

¹⁰⁸ ‘Another Chinese court finds that AI-generated images can be protected by copyright’. Available at <https://www.scl.org/another-chinese-court-finds-that-ai-generated-images-can-be-protected-by-copyright-the-changshu-peoples-court-and-the-half-heart-case/> (accessed 20 September 2025).

¹⁰⁹ ‘Distinguishing Copyrightable from Non-Copyrightable AI-Generated Content’, (2025) 74 *GRUR International*.

general, Chinese courts remain very willing to grant copyright to AI-generated output because they merely consider GenAI a tool.

In Europe, there has only been one judgement so far, resulting in the denial of copyright. But, this was a test case that did not teach us a lot.¹¹⁰ More interesting is the approach of Hugenholtz. In 2020, Hugenholtz et al. published a report in which they proposed a three-tier approach to the question of whether AI output could meet the human creativity requirement. Even though the generation of output is performed mostly by AI (in the 'execution' phase), the human user can still make enough creative choices in the preceding 'conception' phase and in the final 'redaction' phase.¹¹¹ By virtue of the human choices in those three different phases, output can be protected just like 'works of photography or traditional visual art',¹¹² thereby emphasizing the artistic iterative process and warning not to conflate the work with the AI output. Hugenholtz compares GenAI particularly to photography: 'Does the Copyright Office require photographers to disclaim the "camera-generated" features in a photographic work? I don't think so'.¹¹³ As a matter of fact, the equation of GenAI and photography is posited by virtually all scholars and judges who believe copyright protection should be granted more leniently to AI-generated output.¹¹⁴

Interestingly, this three-phase approach builds on the *Painer* case in which the CJEU decided that—despite the creative possibilities being somewhat limited in comparison to other photographers—portrait photographers still have enough creative options and choices to justify copyright (those choices being distributed across three phases).¹¹⁵ Nevertheless, the *Painer* ruling only concerned (classic) cameras, and the CJEU clearly did not have generative AI in mind (ruling back in 2011). In fact, AG Trstenjak concluded that 'only human creations are therefore protected, which can also include those for which the person employs a technical aid, such as a camera'.¹¹⁶ In other words: an assisting/aiding tool or instrument.

In conclusion, although Hugenholtz's approach is more nuanced than the Chinese case law, it still very similarly relies on the assumption that GenAI is (just) a technical aid or tool.¹¹⁷ In short: conflating GenAI with a tool—like a camera or brush—justifies transferring the same legal approach and effects to GenAI, entailing that AI-generated output should be granted protection just as easily.

C. GenAI is autonomous (or wild)

It is clear that to these scholars and judges, the phrase 'GenAI is a tool' is not metaphorical: it is what they believe to be a literal and true statement and consequently reason upon. Why then do I consider this inaccurate (and arguably a metaphor)? To begin with, let us take a closer look at the very sentence: GenAI is a tool or a technical aid. To be clear: we are talking about *generative* AI: all these discussions concern systems like Midjourney, ChatGPT and Suno. So what is contended is that *generative* AI is merely a tool. And that statement contains a contradiction. After all: AI can only be generative if it has the ability to generate—'create'—content, which implies a form of 'autonomy'. The adjective generative describes the AI: it is the AI that generates and not the user. Note at this moment that generative is just another word for creative but one that in some languages feels slightly more mechanical and is thus experienced as less 'threatening' by artists than the clearly 'anthromonopolized'¹¹⁸ word 'creative'. Even though the words mean essentially the same and one could even argue that 'generative' is more (cosmologically) 'creative' as it is etymologically related to the word genesis.

Saying that generative AI is (merely) a tool precisely overlooks this most defining feature of *generative* AI.¹¹⁹ It can for instance generate ('create') content such as a song, article or image in response to a single-worded text prompt.¹²⁰ Therefore, saying that GenAI is just a technical aid or instrument is at least incomplete and potentially misleading. In light of this, metaphors and related analogies with cameras, brushes and other passive instruments—by themselves—fall short in an analysis of generative AI.

When I first used the term 'autonomous'¹²¹ in a draft article to distinguish GenAI from simpler instruments the editorial board of the legal journal advised against using it because AI merely consists of algorithms and lacks spirit.¹²² Of course I cannot read minds but I presume that autonomy for some people—depending on their worldview—is an 'anthromonopolized' concept. Only

¹¹⁰ Decision of the Municipal Court of Prague of 11 October 2023, č.j. 10 C 13/2023-16 (Taubel).

¹¹¹ PB Hugenholtz, 'Copyright and the Expression Engine: Idea and Expression in AI-Assisted Creations' (2025) 100 Chi-Kent L Rev 253.

¹¹² PB Hugenholtz, 'Copyright and the Expression Engine: Idea and Expression in AI-Assisted Creations' (2025) 100 Chi-Kent L Rev 262.

¹¹³ PB Hugenholtz, 'Copyright and the Expression Engine: Idea and Expression in AI-Assisted Creations' (2025) 100 Chi-Kent L Rev 264.

¹¹⁴ Eg also: E Lee, 'Prompting Progress: Authorship in the Age of AI' (2024) 76 Fla. L. Rev. 1445. Available at: <https://scholarship.law.ufl.edu/flr/vol76/iss5/5/>.

¹¹⁵ CJEU Case C-145/10 Eva-Maria Painer (2011) ECLI:EU:C:2011:798 para 91: 'In the preparation phase, the photographer can choose the background, the subject's pose and the lighting. When taking a portrait photograph, he can choose the framing, the angle of view and the atmosphere created. Finally, when selecting the snapshot, the photographer may choose from a variety of developing techniques the one he wishes to adopt or, where appropriate, use computer software'.

¹¹⁶ Opinion AG Trstenjak in Case C-145/10—Painer (2011) ECLI:EU:C:2011:239, para 121 (my italics).

¹¹⁷ In fact this is the explicit point of departure in the IViR-report, stating (in 2020) that 'fully autonomous creation or invention by AI does not exist' and that 'We therefore view AI systems primarily as tools in the hands of human operators'. From a copyright and practical point of view the autonomy analysis in 2026 could be different though (see para 5.C). IViR & JIIP *Trends and Developments in Artificial Intelligence, Challenges to the Intellectual Property Rights Framework: Final Report* (2020) 6. It is of course possible that this IViR-report has influenced Chinese courts, as it was already published in 2020.

¹¹⁸ By this term I mean that it is experienced as an exclusively human word/concept. After all 'identifying' a certain term as a 'metaphor' tells us something about the identifier in the first place: about his or her beliefs. Saying that 'AI creates art' can only be an anthropomorphic metaphor if someone first anthromonopolizes 'create' and 'art'.

¹¹⁹ AF Cooper and J Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' (2025) 100 Chi-Kent L Rev 149–150 & 210. This part of the article is all about the 'generalizing'—not the memorizing—features of GenAI. Despite the fact that such expressive data emerge from algorithmic computation and prediction one could still speak of a system that functions autonomously (once trained and deployed) from a copyright perspective.

¹²⁰ This is exactly why *generative* AI is competing with human jobs, in for example design and translation industries. E Creamer, 'Survey finds generative AI proving major threat to the work of translators' *The Guardian* 16 April 2024; WEF, *The Future of Jobs Report 2025*.

¹²¹ Autonomous from a copyright perspective.

¹²² In the article *Prompteursrecht*, I explained in a more technical manner why GenAI systems should be deemed autonomous from a copyright point of view. An important difference with tools (like a camera) is that generative AI typically relies on randomness embedded within the AI system. In the case of visual AI systems (diffusion-based models), this is done mostly by means of the so-called seed number. This number initiates a unique 'big bang' (admittedly a metaphor) of random noise that underlies the production and expression of each artificial creation in diffusion programmes. Each seed number is entered into an algorithm, after which the neural network generates 'pixels' in a unique and stochastic (random) manner. Think of a coloured static television screen. Once this random noise has been generated, the neural network processes it in a stochastic process, iteratively and predictively refining it into a form that ideally conforms to the prompt. The reason that output varies—despite identical input (prompts)—is this seed number. The corresponding random noise determines the expression in the output, while the prompt often merely functions as a theme or abstract idea. The amount of seed numbers is colossal. For example, most diffusion programmes have 4.294.967.296 numbers. This means that there are equally many output possibilities for one and the same prompt. If the user would generate all those different outputs with the same identical input: would his copyright stretch out to all the different outputs? That would indeed practically be the same as protection of ideas. MA Smit, 'Prompteursrecht', *Ars Aequi* 2025/1.

humans—and depending on one's beliefs perhaps (certain) animals—can be autonomous. Applying it to a machine like AI is therefore a metaphorical exercise.

And yet, different scholars describe generative AI as (partly) autonomous.¹²³ Qian Wang, for example, was one of the first to criticize the decision of the Beijing Court. According to him, GenAI is not like a passive tool; it is in fact 'autonomous' which in this case entails that the user does not decide the eventual expression (output) but merely the idea.¹²⁴ Ideas are however uncopyrightable. According to Wang, granting copyright to the GenAI user is like granting copyright to an art teacher for all the different paintings that have been made by his 30 students in response to his assignment.¹²⁵

In this way, Wang and other scholars counter the 'tool' metaphor with anthropomorphic metaphors and analogies. Like the user of GenAI the art instructor does not control the neural networks of his students and their creative expression. The autonomy of GenAI and the concomitant lack of human control and predictability therefore legitimize a different approach than 'merely passive tools'.¹²⁶

This manner of legal reasoning is comparable to the US Copyright Office (USCO) and its first refusals to grant copyright to AI-generated images. For example, in the *Zarya of the Dawn* case the supposed author (AI user) claimed that the AI-generated image was protected because the user 'authored every aspect of the work, with Midjourney serving merely as an assistive tool'.¹²⁷ The USCO however did not agree with this portrayal of GenAI as a simple tool, which proved to be decisive: 'Rather than a tool that Ms. Kashtanova controlled and guided to reach her desired image, Midjourney generates images in an unpredictable way. Accordingly, Midjourney users are not the "authors" for copyright purposes of the images the technology generates'.¹²⁸

In other cases like *Opéra Spatial* where the user—Allen—allegedly used 624 prompts to produce the output, the USCO decided similarly because it (still) did not deem GenAI just a tool and because the text prompts do not function as 'direct instructions' and the eventual expression cannot be said to originate in the author's creativity.¹²⁹ Importantly, Allen repeatedly refused to show what prompts he used. This fact—combined with the unpredictable nature of GenAI—prompted the USCO to deny copyright.

¹²³ D Gervais, 'The Machine As Author' (2020) SSRN 2106; N Maamar, 'Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen' (2023) ZUM 490; J Fritz, 'Understanding authorship in Artificial Intelligence-assisted works' (2025) JIPLP Issue 5.

¹²⁴ Q Wang, 'Creation Is Not Like a Box of Chocolates: Why Is the First Judgment Recognizing Copyrightability of AI-Generated Content Wrong?' (2024) GRUR 773.

¹²⁵ Q Wang, 'Creation Is Not Like a Box of Chocolates: Why Is the First Judgment Recognizing Copyrightability of AI-Generated Content Wrong?' (2024) GRUR 773–776.

¹²⁶ Q Wang, 'Creation Is Not Like a Box of Chocolates: Why Is the First Judgment Recognizing Copyrightability of AI-Generated Content Wrong?' (2024) GRUR 773.

¹²⁷ US Copyright Office, *Cancellation Decision re: Zarya of the Dawn* (21 February 2023) 3.

¹²⁸ US Copyright Office, *Cancellation Decision re: Zarya of the Dawn* (21 February 2023) 9. This reflected the USCO guideline in which the USCO explained that the Office will research 'whether the "work" is basically one of human authorship, with the computer [or other device] merely being an assisting instrument, or whether the traditional elements of authorship in the work (literary, artistic, or musical expression or elements of selection, arrangement, etc.) were actually conceived and executed not by man but by a machine'. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed Reg 16 192 (16 March 2023) ('AI Registration Guidance').

¹²⁹ US Copyright Office, *Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial* (5 September 2023) 7.

To rebut the 'GenAI is a tool' claim the USCO also drew an analogy with the *Kelley v Chicago Park District* case where the designer of a vast 'living garden' could not claim copyright on the garden due to a lack of control: for years the garden was subject to the forces of nature such as wind, animals, diseases etc.¹³⁰ Nature's indomitability thus hindered creative control and concomitantly copyright protection.¹³¹ Allen subsequently complained by using a—at this point familiar—rhetorical reversal: 'Presumably painters are not required to submit a video of themselves painting to prove they created the painting. The same should be true for the use of other tools'.¹³²

A metaphor that is closely related to these themes of 'autonomy' and 'wild nature' is that of an animal tamer. In fact, the newly coined Dutch term 'prompteur'—the human user of an AI system—contains this very metaphor. This neologism—introduced by Dirk Visser—is a contraction of the words 'prompt' and 'dompteur' while it is also reminiscent of the Dutch word for author 'auteur'.¹³³ Dompteur is the Dutch (and French) word for a tamer of wild animals, lions in particular. Thus, it aptly illustrates the 'control' issue of copyright and AI: if the system—like a wild animal—does not obey the prompteur (sufficiently), the output will not be protected.

D. GenAI is an artificial apprentice

As the previous section showed, the copyrightability question revolves around creative control and causality. If the author does not sufficiently control the AI, then a causality problem arises. The expression in the output is not the result of the free and creative choices of the user, but of the 'algorithms' inside the model. Entailing that the output does not reflect the personality of the human user.¹³⁴ As copyright has always been and still is anthropocentric this is clearly an obstacle to copyrightability.

A useful exercise that may, to a certain extent, reconcile the previously described opposing metaphors and perspectives is to consider AI an artificial apprentice or pupil. After all, the ways in which AI can be used are just as myriad and diverse—yet more accessible—as the ways people can 'use' or help other people in creating art.

For instance, painters like Peter Paul Rubens employed apprentices (pupils) who assisted them in making large paintings. Such assisting painters possibly had some (slight) degree of autonomy: they might have filled in some details on their own, but as long as they in general faithfully filled in the design or sketch of Rubens: Rubens would still be considered the author. From a causal point of view the pupil—like a middleman—drops away and the copyright solely belongs to the intellectual author (Rubens).¹³⁵ That is, of course, provided that Rubens did not merely control the idea but the expression itself. In fact, the Dutch Copyright Act still stipulates this principle in article 6, which is inspired by the very example of Rubens and his apprentices.¹³⁶ Although

¹³⁰ US Copyright Office, *Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial* (5 September 2023) 7; *Chapman Kelley v. Chicago Park District*, No. 08-3701 (7th Cir. 2011) 7.

¹³¹ 'At any given moment in time, a garden owes most of its form and appearance to natural forces, though the gardener who plants and tends it obviously assists'. *Chapman Kelley v. Chicago Park District*, No. 08-3701 (7th Cir. 2011) 32.

¹³² District Court, D. Colorado, *Allen v. Perlmutter* (1:24-cv-02665) 35–36.

¹³³ MA Smit, 'Prompteursrecht' *Ars Aequi* 2025/1.

¹³⁴ The EU criterion: Mio (Joined Cases C-580/23 and C-795/23) EU:C:2025:941 (4 December 2025) para 82.

¹³⁵ MA Smit, 'Leiding en toezicht over de kunstmatige helper' *Auteursrecht* 2025/3.

¹³⁶ Article 6 of the Dutch Copyright Act: 'If a work has been made after the design by and under the direction and supervision of another person, that

such an apprentice could have had a slightly higher degree of autonomy, the legal effects are identical to the usage of a human *amanuensis*.¹³⁷

Now, we could reason similarly when it comes to GenAI. Provided that GenAI actually follows the sketch and design of the user (which begs the question of whether it is still ‘generative’). The problem, however, is that most GenAI systems—especially textual prompt based ones—by their basic architecture do not work like this and are more comparable to *outsourcing* art to a commissioned artist or to a stubborn pupil who determines the expression for himself.¹³⁸ The difference between these situations being that if the *human* apprentice ‘intervenes’ by not following the instructions, he himself will be the author.¹³⁹ Whereas the *artificial* apprentice is no legal subject and cannot carry rights: an ‘artificial’ interruption in the chain of causation will therefore cast the (unedited) output into the public domain.¹⁴⁰

These examples and analogies show that this is all a matter of autonomy and control, but also how creating categories such as ‘generative’ versus ‘assisting’ AI is of limited value exactly because the categories are not clearly delineated. For example, a large language model—usually and rightfully so, associated with highly autonomous AI systems such as Copilot, in which they are integrated—could theoretically also be used in such a manner that human control over it is, in fact, (nearly) complete.¹⁴¹

E. Rhetorical analysis

The presumption that GenAI is a tool and (essentially) functions like a camera or brush greatly affects the legal analysis of the copyrightability question. After all, this is essentially a

person is taken to be the maker of the work’. This article was drafted for the first modern Dutch Copyright Act of 1912 and has remained unchanged ever since. The underlying principle of the article is that the copyright belongs to the *intellectual* (‘geestelijke’) author, which means that he can be author even if he did not physically manufacture the first concrete copy of the work. In the official explanatory memorandum, the legislator mentions famous painter Peter Paul Rubens who like other painters at the time employed apprentice painters (pupils) who assisted him in making large paintings. Because of article 6, the ‘hand’ of the apprentice, operating faithfully under direction and supervision of the master painter, is effectively disregarded from a copyright and causal perspective. Consequently, the intellectual work is attributed directly to the ‘brain’ of the master painter (Rubens), leaving the pupil copyrightless. Although some minor interpretation of the apprentice himself would probably not affect this mechanism, the masters design should influence the actual expression and not just the idea or theme. MA Smit, ‘Leiding en toezicht over de kunstmatige helper’ Auteursrecht 2025/3.

¹³⁷ For centuries writers have relied on *amanuensis*: people who functioned as the (writing) hand (*manu*) of the creative person, so effectively what we would now call a typist. If an *amanuensis* wrote what the artist delivered then no one should claim that the *amanuensis* is the author, he is merely the manufacturer of the first physical copy of someone else’s creative work. Not the author. Of course nowadays AI applications exist for this and we can merely dictate to this app what we want written down. Despite the *amanuensis* now being an AI application; no one should deny the copyright of the human deliverer. From a causal copyright perspective the *amanuensis*, typist and such an obeying AI application are all used as a mere tool and therefore not relevant to the analysis. On the *amanuensis* and copyright see: JC Ginsburg and LA Budiardjo, ‘Authors and Machines’ (2019) 34 Berkeley Tech L J 358–359.

¹³⁸ See footnote 120.

¹³⁹ On ‘interventions’ and ‘disruptions’ in the chain of causation, see DL Burk, ‘Thirty-Six Views of Copyright Authorship, By Jackson Pollock’ (2020) Houston Law Review UC Irvine School of Law Research Paper No. 2020-40.

¹⁴⁰ On causation and copyright, see DL Burk, ‘Thirty-Six Views of Copyright Authorship, By Jackson Pollock’ (2020) Houston Law Review UC Irvine School of Law Research Paper No. 2020-40; DJ Gervais, ‘Second-Degree Intellectual Property’ (2024) SSRN.

¹⁴¹ For instance, there have been experiments in which LLMs are integrated into a *brain-computer interface* (BCI) allowing the patient to write or talk from thought and effectively restoring or healing the central nervous system. Of course in such a situation, it would be wrong to call the AI ‘generative’. This technique uses a large language model; but in such a manner that the human control over the AI-augmented BCI is complete. Therefore, from a causal and copyright point of view, there is no *generative* AI. There is merely AI that helps restoring the autonomy of the human such as the ability to talk or write. J Levý e.a., ‘Brain-to-Text Decoding: A Non-invasive Approach via Typing’ (2025) Arxiv. Available at <https://arxiv.org/abs/2502.17480>; F.R. Willet e.a., ‘High-performance brain-to-text communication via handwriting’. (2021) Nature (593).

causality question: does the expression in the output reflect human creative choices? Because Chinese courts depict GenAI as a tool, there is no causality issue, and the ‘AI generated’ output is deemed a human creation and therefore protected.

On the other hand, stating that GenAI is autonomous and humanizing GenAI with anthropomorphic metaphors creates (or exposes) a causality disruption because humans similarly have the potential of creating something on their own. If the human user merely controlled the idea or theme he should not be the author. Viewing AI as either a tool or something *autonomous*—both of which could be argued to be metaphors—is thus very decisive for the legal analysis and judgement. This essentially explains the USCO approach and how this differs dramatically from the Chinese approach, even though, in both jurisdictions, copyright is anthropocentric and requires a *human* intellectual creation.

With regard to the categorization of AI, we can conclude that copyright theory manifests itself *prototypically* on the matter of copyrightability. Scholars across the world have tried to distinguish *assisting* or *aiding* AI from *generative* AI, but to my knowledge, no one has succeeded in actually defining those categories/terms and clearly providing a boundary between the two. The reason for this is that we simply cannot in *abstracto* say where the exact border lies between sufficient and insufficient human control or autonomy.¹⁴²

If the human user is autonomous enough and controls the expression sufficiently, we speak of *assisting* AI. If they do not, we call it *generative*. Thus, we actually define or categorize AI by the level of human control over it, at least from a copyright perspective. However, how much creative control is sufficient to justify copyright is one of the holy grails in copyright.¹⁴³ Nobody knows, at least not in *abstracto*. The question can only be answered in relation to specific facts of a case, and even then, two equally equipped judges might both come to a different conclusion.

And so, when scholars want to ‘define’ assisting AI from generative and distinguish them from one another, all they essentially do is mentioning *prototypes* of those perceived categories.¹⁴⁴ For example, no one will deny copyright for someone using Microsoft Word Spelling Check. This is the prototype of *assisting* AI: the autonomy of the human is undisputed. Vice versa, no copyright scholar will argue that a book chapter generated by ChatGPT—in response to a short textual prompt—will be copyright-protected. Text-to-output chatbots like ChatGPT are the prototype of *generative* AI. However, these examples are the easy and largely uncontested cases. They do not shed light on the large twilight zone between these prototypes and no one exactly knows where the border lies. To complicate matters even more, some judges and scholars simply combine the two and speak of ‘Generative AI-assisted output’ without even noticing the contradiction. All in all, attempts to theoretically categorize AI (use) into ‘assisting’ and ‘generative’ may be valiant, but possibly also slightly futile.

VI. Conclusion

‘Though questions of truth do arise for new metaphors, the more important questions are those of appropriate action. In most cases, what is at issue is not the truth or falsity of a

¹⁴² This is also increasingly difficult because of AI applications and integration in, eg cameras, as noticed by Hugenholtz: PB Hugenholtz, ‘Copyright and the Expression Engine: Idea and Expression in AI-Assisted Creations’ (2025) 100 Chi-Kent L Rev 264.

¹⁴³ This is an ongoing quest (and debate) since—at least—the rise of Dadaism.

¹⁴⁴ Lakoff and Johnson, *Metaphors We Live By* 164–166.

metaphor but the perceptions and inferences that follow from it and the actions that are sanctioned by it’.

Lakoff and Johnson (1980)¹⁴⁵

This quote cuts to the core of this paper, which main purpose was not to resolve semantic and philosophical debates, but rather to analyze how lawyers reason and argue metaphorically when it comes to GenAI and how this results in legal consequences and specific rulings. *Metaphors we judge by*.

It is hard to deny the influential role metaphors and (follow-up) analogies play in the first judgements and legal literature around the world, regarding these three core copyright questions. Because GenAI is such an inaccessible and abstract phenomenon, lawyers—like other human beings—are bound to conceptualize and analyze GenAI in terms that actually relate to another phenomenon. The AI-is-brain metaphor is particularly tenacious and omnipresent. Moreover, its double-edged nature easily creates a maelstrom of fallacies leading to misunderstandings within and outside the realm of copyright.

When AI is considered a human brain, then it can learn. And learning in itself should not be bothered by copyright; otherwise, even learning schoolchildren would be affected. Therefore, training is allowed and cannot be opposed by copyright holders. A metaphor that is already judged by. Likewise, we cannot possibly speak of copies inside a brain but merely of memorizing or knowing, and those phenomena are similarly out of copyright’s reach. If we consider AI a mere tool, devoid of any (algorithmic) autonomy or own expressive potential, then we should treat it the same as a paint brush or a (classic) camera. Therefore,

AI-generated output should be copyright protected. A metaphor that is already judged by.

In all these cases, the best explanation for the contrasting judgements and analyses is the (unconscious) acceptance or refusal of the dominant metaphor. But all three questions have something else in common; they all pivot around a causality problem. In the first case (AI learns): the central question is whether there is a causal link between training a model (machine ‘learning’) and alleged market substitution. The second issue focuses on the causal chain between an infringing copy in the output and the alleged master copy *inside* the AI model from which the secondary copy in the output supposedly originates. In the third case, a direct causal effect needs to be proven between the creative choices of the *human* user and the specific expression in the output (otherwise it is likely that the AI itself determined the expression). And as causality often is a malleable concept to begin with, it is ‘proven’ or ‘disproven’ through the aforementioned metaphors.

Finally, this article is certainly not a case against metaphors in general; that would be futile, as we all unavoidably use them in relation to GenAI. Still, that is precisely why we should pay close attention to them. Even more when such a metaphor is part of a broader structural metaphor or narrative strategically benefitting one party or industry in the ongoing copyright disputes and debates.

Acknowledgements

The author is grateful to Prof. S.T. Nieuwenhuis for his comments on an earlier draft of this article.

¹⁴⁵ Lakoff and Johnson, *Metaphors We Live By* 158.