



Generative AI and copyright: between fair use and theft

Alexander Emery Naumann

Suggested citation:

Naumann, Alexander Emery. 2025. "Generative AI and copyright: between fair use and theft." Hannover: Hochschule Hannover. <https://doi.org/10.25968/opus-3650>.

Abstract

This paper takes a closer look at the question of whether an AI-generated artwork infringes on copyrights of human creators. To this purpose, a number of relevant terms are defined, and the legal framework of copyright in the EU and the USA outlined. Afterwards, an explanation of how generative AI is trained follows before the paper examines the process with which an AI generates content. The conclusion analyses ethical aspects of using copyrighted works, especially without the creators' permissions, to train a generative AI that then produces similar works that could directly compete with human creators on whose data it was trained.

Terms of use

CC BY-NC-ND 4.0

This document is made available under these conditions:
Creative Commons - CC BY-NC-ND - Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International
For more information see:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>



Hochschule Hannover

Fakultät III - Medien, Information, Design

Abteilung Information und Kommunikation

Generative AI and copyright: between fair use and theft

Term paper for Information Ethics,
a module of the
Information Management (extra occupational) programme.

Submitted by Alexander Emery Naumann.

Prüferin: Prof. Dr. Jutta Bertram

Dresden, 19.06.2025



This document is licensed under the Creative Commons [Attribution-NonCommercial-NoDerivatives 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CC BY-NC-ND 4.0) license, with the exception of the images used in the work, which are taken from other works and for which individual image rights and license apply. The respective sources of these illustrations are indicated in the captions.

DOI: <https://doi.org/10.25968/opus-3650>

Table of Contents

- 1. Introduction.....2
- 2. Definitions.....2
- 3. Legal Framework.....2
 - 3.1. Data Protection.....3
 - 3.2. Copyright.....3
- 4. Training GenAI.....4
- 5. How does GenAI generate content?.....6
- 6. Conclusion.....9
- Sources.....11

1. Introduction

Artificial intelligence (AI) is advanced technology that has great potential to help humans with their scientific and creative work, but first an AI needs to be trained on large quantities of data. This paper takes a closer look at the question of whether an AI-generated artwork infringes on copyrights of human creators. To this purpose, a number of relevant terms are defined, and the legal framework of copyright in the EU and the USA outlined. Afterwards, an explanation of how generative AI (GenAI) is trained follows before the paper examines the process with which an AI generates content. The conclusion analyses ethical aspects of using copyrighted works, especially without the creators' permissions, to train a GenAI that then produces similar works that could directly compete with human creators on whose data it was trained.

2. Definitions

A **derivative work** is a work that is created based on another person's work or idea. This includes but is not limited to translations into other languages, adaptations into other media like film, music, visual arts, etc., as well as any other modification to the pre-existing work.¹

Data mining, including its subset of **text data mining**, is "any automated analytical technique aimed at analysing text and data in digital form."² In this way, information is generated which can then be used for various purposes, e.g. marketing, training algorithms, etc.

Fair use and **transformative works** describe a practice wherein a limited use of copyrighted material is allowed under USA copyright law without the requirement to obtain consent from the copyright holder. This includes reviews, satire, and educational purposes. Two important prerequisites are the non-profit use of this material, and the amount of the original work that is used in the creation of the transformative work.³

3. Legal Framework

Many common GenAI tools are created, programmed, trained and hosted in the USA, but find applications in the EU as well. It is therefore necessary to look at copyright law and other laws that may regulate the use of AI in the USA and the EU.

1 U.S. Copyright Office, 2020, p. 1 ; Berner Übereinkunft, 1971, Art. 2, 3)

2 European Parliament, 2019, Article 2, (2)

3 U.S. Copyright Office, 2024, § 107 ; U.S. Copyright Office, 2025

3.1. Data Protection

The EU has two of the most comprehensive data protection laws in the world, the General Data Protection Regulation (GDPR), as well as the Data Protection Regulation for EU institutions, bodies, offices and agencies (EUDPR). These two regulations clarify consumer rights and protection for individuals in the EU. Even if the used service is based outside of the EU, as long as a citizen of the EU is able to use a service, the company providing it must comply with the GDPR to operate within the EU.⁴

The USA's privacy and data protection laws are much less strict than those of the EU and do not offer sufficient protection to individuals in terms of data security, consumer rights and corporate accountability.⁵ Individual laws such as the California Consumer Privacy Act exist to regulate specific aspects of the storage of personal information. However, in text and data mining, vast quantities of personal information are not only saved but also analysed and cross-referenced, making it all the more important to protect the privacy of the individual.

3.2. Copyright

The EU *Directive 2019/790* regulates the copyright on the internet within the EU since April 2019. Since June 2024 the *Regulation (EU) 2024/1689* governs the training and use of GenAI within the EU. According to both these regulations, text and data mining is permissible in scientific research environments, provided that the source material is lawfully accessible. It further constrains the copyrights of certain groups of creators, such as authors or performing artists, etc., to make text and data mining possible outside of scientific research. Creators must explicitly deny consent if they wish for their works to be excluded from text and data mining for any purpose other than scientific research.⁶

Furthermore, *Regulation 2024/1689* specifically states that a GenAI is not considered high risk (i.e. in danger of being grossly misused) but is required to comply with the EU copyright law, especially the transparency clauses. For example, it will be necessary to disclose if content was AI-generated, and developers have to disclose summaries of the copyrighted data that was used to train the AI. Developers also must ensure that the AI cannot generate illegal content.⁷

Fair use and transformative works are not as broadly regulated in EU copyright law as in USA copyright law. However, the directive 2001/29/EC includes various uses of copyrighted

4 European Parliament, 2016, Art. 3

5 Aidun, 2025, II B

6 comp. European Parliament, 2019, Art. 3, Art. 4 ; European Parliament, 2024, Art. 53

7 European Parliament, 2025

material that do not require the copyright holder's permission, such as using excerpts for reviews, caricature, and satire. The directive is not legally binding for member states but rather a guideline of permissible exceptions from stricter copyright rules.

According to the Code of Federal Regulations of the USA, text and data mining is permissible for scientific research, teaching, and higher education purposes, similar to European law.⁸

Additionally, the Copyright Office of the USA conducted a study on Artificial Intelligence from August 2023 that has nearly concluded as of May 2025. A report was published in three parts, the third of which, at time of writing this paper, is a pre-publication that awaits finalisation.

Only the third part is relevant to this paper as it discusses the training of AI. For the question of whether it is lawful to use copyrighted material without consent of the creators to train AI, the report argues that the use can be classified as fair use if there is no commercial gain and if the training data has been obtained legally.⁹

4. Training GenAI

GenAI is trained to generate human-like content.¹⁰ To understand how GenAI might be infringing on creators' copyrights, it is crucial to grasp how an AI "learns." Data mining and text data mining are employed to scrape information from a source, which then will be fed into the AI so it can be trained on that information. It can include patterns, trends, and correlations as well as other types of information relating to statistical data. An AI is trained by using machine learning processes that use statistical learning optimisation methods with which a computer can analyse datasets, and find and reproduce patterns.¹¹

To train an AI to "create" images, first digital images are tagged with relevant keywords by a human before introducing the data to the AI training database, or, if the AI is already sophisticated enough, the AI will tag the images itself. A third and more common option is to take the keywords from the website from where the image was copied.¹²

As of writing this paper, it is nearly impossible to avoid having your own information or work scraped for data mining, especially if it is publicly available on the internet. Companies such as Meta, Apple, and Google have stipulations in their Terms of Service that allow them the use of an individual's information to train their respective AI models¹³ and even forward that

8 U.S. Copyright Office, 2025a, b (4), (5)

9 U.S. Copyright Office, 2025b, p. 107

10 Miao, Holmes, 2023, p. 8

11 Sipes, 2006, p. 126 ; UC Berkeley, 2020

12 Getty v. Stability AI, 2023a, IV, 25. VIII, D.89.

13 Tan, 2024

information to third parties¹⁴, making it impossible for a user to discover what their mined information is used for. In theory it is possible, at least in the EU, to opt out of these stipulations without losing the use of these services entirely.¹⁵ However, there is no guarantee that anyone's objection is honoured, nor is there a law or rule that would hinder the provider from making their service unusable without the use of AI or the permission to mine for data from the user. Furthermore, a provider may change their terms of service at any point and require an additional opt-out.¹⁶

Tech companies' data mining doesn't stop at their own databases either. While arguments can be made for a service provider to utilise the data of their users if consent is given, large language model developers have not stopped there. Earlier this year it was revealed that Meta had accessed an illegal library of pirated books for text data mining. While Meta themselves did not pirate these books, they used LibGen, one of the largest online libraries of pirated books, to train their AI model Llama 3 on large quantities of literature without the copyright holders' permissions.¹⁷ In this case, evidence has been given that Meta employees have not only openly discussed the illegality of this method¹⁸, they have also made efforts to remove copyright identifying markers (CIM) like ISBNs from the works, or sought out data sets from LibGen that already removed CIMs before redistributing the pirated works.¹⁹

Other lawsuits have been issued by the New York Times²⁰ and Getty Images²¹ against other GenAI developers on similar grounds. In the case of Getty Images v Stability AI, the AI has reproduced copyrighted photographs by Getty Images that were minimally altered, including the now distorted watermark of the Getty Trademark (s. Ill. 1 on p. 6).²²

14 Meta, 2025, 3.

15 Verbraucherzentrale, 2025

16 Belanger, 2025

17 Reisner, 2025

18 Kadrey v. Meta, 2023a, pp. 3-7 ; Kadrey v. Meta, 2023b, pp. 2-3 ; Kadrey v. Meta, 2023c, pp. 3-7

19 Kadrey v. Meta, 2023d, E ; Kadrey v. Meta, 2023e, B

20 Allyn, 2025

21 Getty Images, 2023

22 Getty v. Stability AI, 2023b, pp. 24-25, 77



Ill. 1: Examples from the Getty lawsuit showing an original Getty photograph on the left, and its imitation generated by Stability AI on the right.²²

5. How does GenAI generate content?

GenAI, as it exists now, does not think for itself and is therefore incorrectly named “artificial intelligence”. All current AI are statistic machines that can only be trained to predict, with varying degrees of accuracy, what is the statistically most likely correct response to a prompt. AI can be compared to predictive text input that we’ve been using for years on our smart devices. Show-runner and TV writer Vince Gilligan gave statement in response to the USA Copyright Office’s study on GenAI:

“Someday, true generative artificial intelligence may come to exist. But right now, what companies such as OpenAI, Google and Meta are hawking are only elaborate guessing machines. As I understand it, statistical probability is at the heart of this process, which enables the software to predict what particular word is most likely to follow any other specific word.

I’ll admit, it’s an impressive trick. But it’s not creation. It’s not storytelling. My toaster oven makes great toast, but that doesn’t mean it’s a chef. Nor is ChatGPT a writer.”²³

A recent example is the case of a journalist who submitted an AI-generated list of recommendations for summer reading to the Chicago Sun-Times. The list contained 15 books,

²³ Gilligan, 2023, p. 1

five of which exist. The remaining ten were fabrications by the AI that the author used to compile this list.²⁴

Once a sufficiently large database has been created, an AI can respond to a text prompt by merging relevantly tagged images. This new image is not a new and original work, but a remix or blend of existing images and thus by definition a derivative work.

An example is given in the Getty v Stability AI lawsuit documents.²⁵ The prompt given to the Stable Diffusion AI was “a dog wearing a baseball cap while eating ice cream”. An AI can only blend elements of images that are saved in its database, and therefore only generate by pure chance (which is increased by generating a large quantity of images with these elements) the correct image, or fail entirely. (s. III. 2)



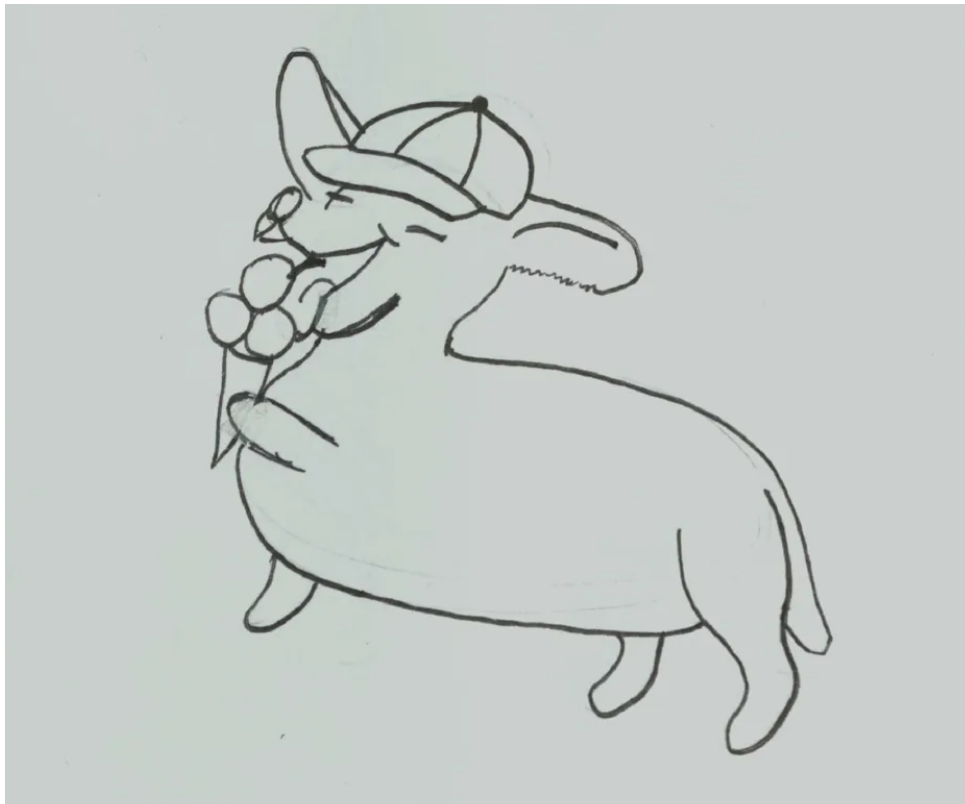
*Ill. 2: Stable Diffusion AI’s image results to the prompt “a dog wearing a baseball cap while eating ice cream”.*²⁵

24 Koebler, 2025 ; Blair, 2025
25 comp. Getty v. Stability AI, 2023, IV, 25. VIII, D.91-98

All of the dogs in illustration 2 are wearing something on their head, and only three of them are eating anything. Ice cream is present in three of the pictures, but only one of these dogs (bottom left) appears to be eating it.

Even though these are likely new images that aren't exact copies of any of the images in the training data, the AI is incapable of adding any original elements to them, therefore only blending and remixing existing images with varying degrees on how much any individual piece of the original training data is used.

For any human individual on the other hand, it would be easy enough to create a visual representation regardless of their artistic talents, as long as they learned what a dog, a baseball cap, and an ice cream are before they're asked to draw an image that combines all three. (s. Ill. 3)



Ill. 3: A drawing of a dog wearing a baseball hat while eating ice cream, by Alexander Naumann

A human, even if they're referencing an existing work, will almost always add their own personality and creativity to the creation, too, something an AI is incapable of providing.

The same is true for text-based content. In the case of the journalist with the summer reading list, two problems are obvious: First is the question on whether a list of book recommendations can be called that if the "author" has not read even one of the books on the

list. Second, and more to the point, is the fact that an AI cannot have an opinion. The AI does not “read” a book which it then recommends. The AI was prompted to generate a list of 15 books that fit a theme and because the AI cannot tell truth from lie, it invented titles that fit that list.

6. Conclusion

The question of the use of copyrighted material to train GenAI without the permission or compensation of the creators is of interest to me as I, too, am a creator who publishes their work online for free. I am acquainted with many artists in varying fields who likewise publish their works online, for free or as a paid service, and who have had their works copied into AI training machines or have had to remove their works to avoid that outcome.

A common thought is that what can be found on the internet for free is fair game in copyright law. While that may be true, it is in my opinion unethical to copy even free art, let alone art that isn’t freely accessible. This is especially true when the data is fed into a machine to train an AI with the goal to churn out the same or similar art at maximum speed and high quantities. This generated “art” will flood the market and put real artists out of business or stifle their drive to create.

Meta, Google and other companies have argued in court that their copies of these copyrighted works are not only necessary to effectively train the AI models but also legal under fair use.²⁶ Another argument brought up in court is that an AI being trained is comparable to a human reading a book or studying art.²⁷ However, an AI can process much larger quantities of information in a much shorter time than any person can. Likewise, a person, while learning from what they study, will always add their own personality to any new creation, whereas an AI can only remix the information it has been supplied.

Another more important aspect is that the fair use clause only applies to the transformation of a work, not the acquisition of the original text. Therefore, even if fair use can be applied to the AI-generated content, it does not absolve these companies from using copyrighted works without permission. Just like a human either buys a book or borrows it from a library – both situations guaranteeing that the author has been compensated – a tech company seeking to use literature not in the public domain to train their AI should license these works fairly.

Others argue that by forcing companies like Meta and Google to pay for licenses to train their AI on copyrighted works, smaller AI companies will not be able to survive and compete

²⁶ Kadrey v. Meta, 2023f, C.

²⁷ Google, 2023, p. 9

fairly.²⁸ That, however, is a straw man argument at best, as the network Fairly Trained has proven.

Ed Newton-Rex, a founder of Fairly Trained, disagreed strongly with Stability AI's policy to use copyrighted works without permission from or compensation to the copyright holders, so he resigned in November 2023.²⁹ He then went on to build a network of AI companies who seek to compensate creators fairly for the use of their work to train AI models.³⁰ Licensing models range from big upfront payments to use a creator's work for a limited time or indefinitely, to royalty schemes in which the creator is paid a set amount every time their work is recalled by the AI to generate output.³¹

If tech companies insist on taking works without creators' permissions, it will have dire consequences for all of society. The internet is becoming less free as creators everywhere begin to remove their works to avoid having it taken without their consent. The practices of these tech companies are having a direct impact on values such as free expression – artistic or otherwise – and privacy, as well as truth and transparency. While the EU requires that AI-generated content always be marked as such, the AIs themselves do not always distinguish between the reproduction of real facts, and their own generated content. To the AI, there is no real or fake; there is only statistical data and therefore a truth is the same as a lie.

Another consequential argument must be that if more and more original, human-made content disappears from the internet in an attempt to protect it from being stolen, soon all the GenAI models will be trained on the same content that these GenAI have produced themselves.³² As with the copy of a copy of a VHS-tape, the quality of the next copy will degrade with every reproduction. Soon, all that will be left is the so-called AI slop, which has already made its way into libraries.³³

Copyright holders deserve legal protection and a better solution than being required to actively opt out of having their works taken for AI training. In a world where GenAI are here to stay whether we like it or not, the rights of individual humans are more important than ever.

Characters: 16,910

28 Martens, 2024, p. 23.

29 Ed Newton-Rex @ednewtonrex, 2023 ; Kleinman, 2023

30 Fairly Trained, 2025

31 TED, Ed Newton-Rex, 2024, 9:45-10:08

32 Villalobos et al, 2024, pp. 6-9

33 Maiberg, 2025

Sources

- Aidun, Elias: Data Privacy in the Digital Age: A Comparative Analysis of U.S. and EU Regulations?. In: University of Cincinnati Law Review Blog, Vol. 93. Blog entry from 05.03.2025. Online at <https://uclawreview.org/2025/03/05/data-privacy-in-the-digital-age-a-comparative-analysis-of-u-s-and-eu-regulations/> [Accessed on 15.06.2025]
- Allyn, Bobby (2025): “The New York Times” Takes OpenAI to Court. ChatGPT’s Future Could Be on the Line. In: NPR: Business. News article from 14.01.2025. Online at <https://www.npr.org/2025/01/14/nx-s1-5258952/new-york-times-openai-microsoft> [Accessed 15.06.2025]
- Belanger, Ashley (2025): Meta Is Making Users Who Opted out of AI Training Opt out Again, Watchdog Says. EU users have less than two weeks to opt out of Meta's AI training. In: Ars Technica. Blog entry from 14.05.2025. Online at <https://arstechnica.com/tech-policy/2025/05/meta-is-making-users-who-opted-out-of-ai-training-opt-out-again-watchdog-says/> [Accessed 15.06.2025]
- Berner Verband: Berner Übereinkunft Zum Schutz von Werken der Literatur und Kunst. (1886 – 1979). Online at http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBI&jumpTo=bgbl273s1069.pdf [Accessed on 15.06.2025] and https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBI&jumpTo=bgbl285s0081b.pdf [Accessed on 15.06.2025]
- Blair, Elizabeth (2025): How an AI-Generated Summer Reading List Got Published in Major Newspapers. In: NPR: Books. News article from 20.05.2025. Online at <https://www.npr.org/2025/05/20/nx-s1-5405022/fake-summer-reading-list-ai> [Accessed on 15.06.2025]
- 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. Online at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0790> [Accessed on 15.06.2025]
- Ed Newton-Rex @ednewtonrex (2023): I’ve resigned from my role leading the Audio team at Stability AI [...]. Tweet from 15.11.2023. Online at <https://x.com/ednewtonrex/status/1724902327151452486> [Accessed on 15.06.2025]
- European Parliament (2025): EU AI Act: First Regulation on Artificial Intelligence. The use of artificial intelligence in the EU is regulated by the AI Act, the world’s first comprehensive AI law. Find out how it protects you. In: Topics: Digital – Artificial Intelligence. Blog entry from 08.06.2023. Last updated on 19.02.2025. Online at <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-act-different-rules-for-different-risk-levels-6> [Accessed on 15.06.2025]
- Fairly Trained (2025): About. Online at <https://www.fairlytrained.org/about> [Accessed 15.06.2025]

- Getty Images (2023): Getty Images Statement. In: Newsroom. News article from 17.01.2023. Online at <https://newsroom.gettyimages.com/en/getty-images/getty-images-statement> [Accessed on 15.06.2025]
- Getty Images (US), Inc. v. Stability AI, Inc. (2023a): Document 18. Exhibit A. Filed on 02.05.2023. Online at <https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.18.1.pdf> [Accessed on 15.06.2025] In: 1:23-cv-00135 (D. Del.). Court Case in 17:501 Copyright Infringement. Filed on 03.02.2023. Last known filing on 02.05.2025. Online at <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/> [Accessed on 15.06.2025]
- Getty Images (US), Inc. v. Stability AI, Inc. (2023b): Document 40. Filed on 02.07.2024. Online at <https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.40.1.pdf> [Accessed on 15.06.2025] In: 1:23-cv-00135 (D. Del.). Court Case in 17:501 Copyright Infringement. Filed on 03.02.2023. Last known filing on 02.05.2025. Online at <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/> [Accessed on 15.06.2025]
- Gilligan, Vince (2023): My name is Vince Gilligan, and I'm the creator of the television series "Breaking Bad." [...] Online at https://downloads.regulations.gov/COLC-2023-0006-8501/attachment_1.pdf [Accessed on 15.06.2025] Response to: U.S. Copyright Office Artificial Intelligence Study (2023-2025). Online at <https://www.copyright.gov/policy/artificial-intelligence/> [Accessed 15.06.2025]
- Google (2023): Artificial Intelligence and Copyright. Online at <https://www.documentcloud.org/documents/24117935-google/> [Accessed 15.06.2025]
- Kadrey v. Meta Platforms, Inc. (2023a): Document 449-4. Filed on 20.02.2025. Online at <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.449.4.pdf> [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]
- Kadrey v. Meta Platforms, Inc. (2023b): Document 417-6. Filed on 05.02.2025. Online at <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.417.6.pdf> [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]
- Kadrey v. Meta Platforms, Inc. (2023c): Document 391-24. Filed on 14.01.2025. Online at <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.391.24.pdf> [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]

- Kadrey v. Meta Platforms, Inc. (2023d): Document 562-43. Exhibit 147. Filed on 25.04.2025. Online at https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.562.43_1.pdf [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]
- Kadrey v. Meta Platforms, Inc. (2023e): Document 377. Filed on 08.01.2025. Online at <https://www.courtlistener.com/docket/67569326/377/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]
- Kadrey v. Meta Platforms, Inc. (2023f): Document 489. Filed on 24.03.2025. Online at <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.489.0.pdf> [Accessed on 15.06.2025] In: 3:23-cv-03417 (N.D. Cal.). Court Case in 17:504 Copyright Infringement. Filed on 07.07.2023. Last known filing on 14.05.2025. Online at <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/> [Accessed on 15.06.2025]
- Kleinman, Zoe (2023): AI Chief Quits over “exploitative” Copyright Row. In: BBC News: Technology. News article from 17.11.2023. Online at <https://www.bbc.com/news/technology-67446000> [Accessed on 15.06.2025]
- Koebler, Jason (2025): Chicago Sun-Times Prints AI-Generated Summer Reading List With Books That Don’t Exist. In: 404 Media. News article from 20.05.2025. Online at <https://www.404media.co/chicago-sun-times-prints-ai-generated-summer-reading-list-with-books-that-dont-exist/> [Accessed on 15.06.2025]
- Maiberg, Emanuel (2025): AI-Generated Slop Is Already In Your Public Library. In: 404 Media. News article from 04.02.2025. Online at <https://www.404media.co/email/7a1109dd-e6cd-46ac-be2f-a6d6461fc2ff> [Accessed 15.06.2025]
- Martens, Bertin (2024): Economic Arguments in Favour of Reducing Copyright Protection for Generative AI Inputs and Outputs. In: Bruegel, Iss. 09/2024, p. 23. Online at <https://www.bruegel.org/sites/default/files/private/2024-04/WP%2009%20040424%20Copyright%20final.pdf> [Accessed on 15.06.2025]
- Meta (2025): Meta AIs Terms of Service (EU). Last updated on 24.02.2025. Online at <https://www.facebook.com/legal/eu-ai-terms> [Accessed on 15.06.2025]
- Miao, Fengchun ; Holmes, Wayne (2023): Guidance for Generative AI in Education and Research. Published by UNESCO. Paris. Online at <https://doi.org/10.54675/EWZM9535> [Accessed on 15.06.2025]

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Online at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434> [Accessed on 15.06.2025]
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence. Online at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> [Accessed on 15.06.2025]
- Reisner, Alex (2025): The Unbelievable Scale of AI's Pirated-Books Problem. In: The Atlantic. News article from 20.03.2025. Online at <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/> [Accessed on 15.06.2025]
- Sipes, James L. (2006); Mining for Data. In: Landscape Architecture, Vol 96, No. 10, pp. 126–32. Online at <https://www.jstor.org/stable/44677035> [Accessed on 15.06.2025]
- Tan, Eli (2024): When the Terms of Service Change to Make Way for A.I. Training. In: The New York Times: Technology. News article from 26.06.2024. Online at <https://www.nytimes.com/2024/06/26/technology/terms-service-ai-training.html> [Accessed on 15.06.2025]
- TED ; Ed Newton-Rex (2024): How AI Models Steal Creative Work — and What to Do About It [TED Talk]. Recorded on 22.10.2024 at TEDAI San Francisco. Video published on 19.03.2025 on YouTube. Online at <https://www.youtube.com/watch?v=U9d0p96N1iw> [Accessed on 15.06.2025]
- UC Berkeley School of Information: What Is Machine Learning (ML)? In UC Berkeley School of Information: Blog. Blog entry from 26.06.2020. Last updated February 2022. Online at <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/> [Accessed on 15.06.2025]
- U.S. Copyright Office (2020): Copyright in Derivative Works and Compilations. In: Circular. No. 14. Online at <https://www.copyright.gov/circs/circ14.pdf> [Accessed on 15.06.2025]
- U.S. Copyright Office (2024): Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code. In: Circular. No. 92. Online at <https://www.copyright.gov/title17/title17.pdf> [Accessed on 15.06.2025]
- U.S. Copyright Office (2025a): Code of Federal Regulations. § 201.40 Exemptions to Prohibition against Circumvention. In: Federal Regulations : Patents, Trademarks, and Copyrights. Online at <https://www.copyright.gov/title37/201/37cfr201-40.html> [Accessed on 15.06.2025]
- U.S. Copyright Office (2025b): Copyright and Artificial Intelligence. Part 3: Generative AI Training. Pre-Publication Version. A Report of the register of copyrights. Online at <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> [Accessed on 15.06.2025]

- U.S. Copyright Office (2025): Fair Use Index. In: Home. Last updated February 2025. Online at <https://www.copyright.gov/fair-use/index.html> [Accessed on 15.06.2025]
- Verbraucherzentrale (2025): “Meta AI” bei Facebook, Instagram und WhatsApp – so widersprechen Sie. In: Wissen – Digitale Welt – Soziale Netzwerke. Blog entry from 27.05.2025. Last updated on 05.06.2025. Online at <https://www.verbraucherzentrale.de/wissen/digitale-welt/soziale-netzwerke/meta-ai-bei-facebook-instagram-und-whatsapp-so-widersprechen-sie-95646> [Accessed on 15.06.2025]
- Villalobos, Pablo et al (2022): Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data. Published on 26.10.2022. Last revised on 04.06.2024. Online at <https://doi.org/10.48550/arXiv.2211.04325> [Accessed on 15.06.2025]

Hiermit erkläre ich, dass ich die eingereichte Seminararbeit selbständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. KI-Anwendungen, die ich als Hilfsmittel verwendet habe, habe ich im Anhang meiner Arbeit dokumentiert. Daraus geht hervor, welche KI-Anwendung ich für die dort angegebenen Zwecke in den genannten Teilen der Arbeit eingesetzt habe.