

Dornis, Tim W. and Stober, Sebastian, Urheberrecht und Training generativer KI-Modelle - technologische und juristische Grundlagen

(September 4, 2024).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4946214

Abstract (EN)

Generative AI is transforming creative fields by rapidly producing texts, images, music, and videos. These AI creations often seem as impressive as human-made works but require extensive training on vast amounts of data, much of which is copyrighted. This dependency on copyrighted material has sparked legal debates, as AI training involves “copying” and “reproducing” these works, actions that could potentially infringe on copyrights. In defense, AI proponents in the United States invoke “fair use” under Section 107 of the Copyright Act, while in Europe, they cite Article 4(1) of the 2019 DSM Directive, which allows certain uses of copyrighted works for “text and data mining.”

This study challenges the prevailing European legal stance, presenting several arguments:

1. The exception for text and data mining should not apply to AI training because the technologies differ fundamentally - one processes semantic information, while the other extracts syntactic information.
2. There is no suitable copyright exception to justify the massive infringements occurring during AI training. Copyrighted works are copied during data collection, fully or partly replicated inside AI models, and can also be reproduced by end-users.
3. Even if AI training occurs outside Europe, developers cannot avoid European copyright laws. If works are replicated within a model, making it available in Europe could infringe the “right of making available” under Article 3 of the InfoSoc Directive. Thus, offering AI services to European users ultimately subjects developers to European copyright laws and European courts’ jurisdiction.

This study suggests to rethink copyright issues in the context of AI. Given the technical revolution and socio-economic disruptions generative AI brings, lawmakers should reconsider how to balance protecting human creativity and fostering AI innovation. The current lack of regulation neglects the technical realities and is thus not only legally unsound but also unjust.

Abstract (DE)

Generative KI verändert die kreative Welt. Sie produziert Texte, Bilder, Musik und Videos praktisch aus dem Nichts und in Sekundenschnelle. Diese KI-Kreationen wirken oft genauso beeindruckend wie von Menschenhand geschaffene Werke, erfordern aber ein umfangreiches Training auf der Grundlage riesiger Datenmengen, von denen viele urheberrechtlich geschützt sind. Diese Abhängigkeit von urheberrechtlich geschütztem Material und der damit verbundene Eingriff in Urheberrechte hat Debatten ausgelöst und in vielen Ländern zu gerichtlichen Auseinandersetzungen geführt. Zur Verteidigung berufen sich KI-Entwickler in den USA auf die „fair use defense“ nach section 107 des U.S. Copyright Act. In Europa wird vor allem Artikel 4 Absatz 1 der DSM-Richtlinie angeführt, der Nutzungen urheberrechtlich geschützter Werke für „Text und Data Mining“ erlaubt.

Die vorliegende Studie stellt die im Moment vorherrschende europäische Rechtsauffassung in Frage, insbesondere in folgenden Punkten:

1. Die Ausnahme für Text und Data Mining sollte nicht für das Training generativer KI-Modelle gelten, weil sich die beiden Technologien fundamental unterscheiden – die eine verarbeitet semantische Informationen, die andere extrahiert und verwertet umfangreich die syntaktischen und damit urheberrechtlich geschützten Informationen der Werke in den Trainingsdaten.
2. Es gibt keine praktisch relevante Schranke, welche die vielfachen, während des KI-Trainings stattfindenden Eingriffe in das Urheberrecht rechtfertigen könnte. Urheberrechtlich geschützte Werke werden bei der Datenerfassung kopiert, ganz oder teilweise in den KI-Modellen repräsentiert und können schließlich auch von den Endnutzern der Modelle vervielfältigt werden.
3. Auch wenn das Training generativer KI-Modelle außerhalb Europas stattfindet, können sich die Entwickler nicht der Geltung und Durchsetzung der europäischen Regeln des Urheberrechts entziehen. Da die zum Training eingesetzten Werke jedenfalls teilweise in den Modellen repräsentiert sind, kann die Bereitstellung von KI-Dienstleistungen an Nutzer in Europa in das „Recht der öffentlichen Zugänglichmachung“ in Art. 3 InfoSoc-Richtlinie eingreifen. Die Entwickler und Anbieter sind deshalb den europäischen Urheberrechtsgesetzen unterworfen und unterliegen der Zuständigkeit der europäischen Gerichte.

Diese Studie regt an, Urheberrechtsfragen im Zusammenhang mit dem Training generativer KI-Modelle einer noch gründlicheren Betrachtung zu unterziehen. Angesichts der technischen Revolution und der sozio-ökonomischen Umwälzungen, die generative KI mit sich bringt, muss der Gesetzgeber entscheiden, wie das Gleichgewicht zwischen dem Schutz menschlicher Kreativität und der Förderung von KI-Innovationen hergestellt werden kann. Die bestehenden Mängel der gesetzlichen Regelungen vernachlässigen die technischen Realitäten und sind daher nicht nur rechtlich zweifelhaft, sondern vor allem auch in der Sache ungerecht.

Urheberrecht und Training generativer KI-Modelle - technologische und juristische Grundlagen

171 Pages Posted:

[Tim W. Dornis](#)

Leibniz University Hannover; New York University School of Law

[Sebastian Stober](#)

Otto-von-Guericke University, Magdeburg

Date Written: September 4, 2024

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4946214