

Dornis, Tim W. and Stober, Sebastian, Urheberrecht und Training generativer KI-Modelle - technologische und juristische Grundlagen
(September 4, 2024). Available at SSRN:

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

Executive Summary

Inhalt des Urheberrechts

Im Rahmen des Trainings generativer KI-Modelle kommt es zu zahlreichen verschiedenen Handlungen der Vervielfältigung urheberrechtlich geschützter Werke im Sinne des § 16 Abs. 1 UrhG.

(1) Dies betrifft zunächst die Sammlung, Vorbereitung und Speicherung der beim Training verwerteten geschützten Werke.

(2) Darüber hinaus kommt es während des Trainingsprozesses generativer KI-Systeme – sowohl beim *Pre-Training* als auch beim *Fine-Tuning* – zu einer urheberrechtlich relevanten Vervielfältigung der zum Training verwerteten Werke „im Innern“ des Modells. Ein expliziter Speichermechanismus ist zwar nicht angelegt. Die Trainingsdaten werden in den aktuellen generativen Modellen – LLMs und (Latent) Diffusion Modellen – aber durchaus memorisiert.

(3) Schließlich kann es bei Einsatz generativer KI-Modelle, insbesondere durch die Nutzer von KI-Systemen (z.B. ChatGPT über die OpenAI-Webseite), zu Vervielfältigungen und Umgestaltungen der für das Training des zugrundeliegenden KI-Modells verwerteten Werke kommen.

(4) In der Zurverfügungstellung der in diese KI-Systeme implementierten generativen KI-Modelle zur Anwendung durch die Nutzer oder zum Download des Modells im Ganzen liegt schließlich eine öffentliche Zugänglichmachung (§§ 15 Abs. 2 Nr. 2, 19a UrhG) von Teilen der für das Training verwerteten und „im Innern“ des Modells vervielfältigten Werke.

Schranken des Urheberrechts

Der geltende Kanon urheberrechtlicher Schrankentatbestände erfasst die mit dem Training generativer KI-Modelle einhergehenden Eingriffe in das Urheberrecht lediglich in einigen wenigen, praktisch nicht relevanten Konstellationen. Vor allem findet die Schranke für Text und Data Mining (TDM) keine Anwendung.

(1) Die im Rahmen der Sammlung, Vorbereitung und Speicherung von Trainingsdaten stattfindenden Vervielfältigungshandlungen fallen nicht unter den

Schrankentatbestand für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG, Art. 5 Abs. 1 InfoSoc-Richtlinie).

(2) Auch die TDM-Schrankentatbestände finden keine Anwendung. Dies gilt grundsätzlich unbestritten im Hinblick auf die Schranke für das Text und Data Mining für Zwecke der wissenschaftlichen Forschung (§ 60d UrhG, Art. 3 DSM-Richtlinie).

(3) Darüber hinaus ist auch einer Anwendung der Schranke für das (kommerzielle) Text und Data Mining (§ 44b UrhG, Art. 4 DSM-Richtlinie) zu widersprechen. Der genaue Blick auf die Technologie generativer KI-Modelle erhellt, dass die Verwertung der urheberrechtlich geschützten Inhalte in den Trainingsdaten – anders als Text und Data Mining – nicht auf semantische Inhalte begrenzt ist. Die Betrachtung von Wortlaut, Systematik und Telos des Schrankentatbestandes verbietet daher eine Anwendung auf das Training generativer KI-Modelle.

(a) Dies kann zunächst durch eine vergleichende Untersuchung der technologischen Grundlagen des TDM und des Trainings generativer KI-Modelle, insbesondere der Unterschiede der dabei eingesetzten Methoden, gezeigt werden: Das Training generativer KI-Modelle begrenzt die Nutzung der Trainingsdaten nicht auf eine reine Auswertung der in den Werken enthaltenen semantischen Informationen. Es erfasst darüber hinaus auch und insbesondere die syntaktischen Informationen. Diese umfassende Verwertung führt zur Repräsentation der Trainingsdateninhalte im Vektorraum der Modelle und damit zu einer Vervielfältigung im Sinne des § 16 Abs. 1 UrhG. Das Training generativer KI-Modelle kann deshalb nicht unter den TDM-Schrankentatbestand gefasst werden.

(b) Eine historische Auslegung des TDM-Schrankentatbestandes bestätigt die technologisch-konzeptionellen Zusammenhänge: Der Gesetzgeber der DSM-Richtlinie hat die technologische Entwicklung kreativ-produktiver KI-Systeme und deren disruptive Auswirkungen nicht vorhergesehen. Das schließt es aus, den ausschließlich für die Auswertung semantischer Informationen konzipierten TDM-Schrankentatbestand auf umfassend syntax-verwertende generative KI-Modelle zu erstrecken. Auch für das Gesetzgebungsverfahren der KI-Verordnung ist kein spezifischer Regelungswille erkennbar, insbesondere keine Befassung des Gesetzgebers mit den technologischen Grundlagen sowie den Unterschieden zwischen TDM und dem Training generativer KI-Modelle.

(c) Auch kann gezeigt werden, dass das Training generativer KI-Modelle – selbst wenn man den TDM-Schrankentatbestand anwenden wollte – gegen den Dreistufentest des internationalen und europäischen Urheberrechts verstieße. Die umfassende Auswertung der syntaktischen Informationen urheberrechtlich geschützter Werke ist als Beeinträchtigung der den Rechteinhabern durch das Urheberrecht zugewiesenen „normalen Auswertung“ und damit als unheilbarer Verstoß gegen die Testvorgaben einzuordnen.

(d) Für den Zeitraum vor der Geltung der DSM-Richtlinie (vor dem 7. Juni 2021) ist ein Zustand der schrankenlos-rechtsverletzenden Verwertung urheberrechtlich geschützter Werke für das Training generativer KI-Modelle festzustellen.

(4) Im Hinblick auf Vervielfältigungen im Rahmen der öffentlichen Zugänglichmachung sowie des Einsatzes generativer KI-Modelle (insbesondere bei der Output-Erstellung) ist festzustellen, dass es für die meisten praktisch relevanten Szenarien an Schrankentatbeständen fehlt. Weder das Zitatrecht (§ 51 UrhG), noch die Schranken für unwesentliches Beiwerk (§ 57 UrhG), für Karikatur, Parodie und Pastiche (§ 51a UrhG) oder zum privaten und sonstigen Gebrauch (§ 53 UrhG) finden Anwendung.

Anwendbares Recht, internationale Zuständigkeit und KI-Verordnung

(1) Für die Rechtsanwendungsfrage und die internationale Zuständigkeit der Gerichte ist gegen die bislang einheitlich vertretene Perspektive der Unangreifbarkeit von KI-Trainingshandlungen im Ausland darauf hinzuweisen, dass es bei Zugänglichmachung von KI-Modellen zur Anwendung durch Nutzer in Deutschland (z.B. bei ChatGPT über die OpenAI-Webseite) aufgrund der Vervielfältigung der urheberrechtlich geschützten Trainingsdaten „im Innern“ der Modelle zu einer öffentlichen Zugänglichmachung im Sinne der §§ 15 Abs. 2 Nr. 2, 19a UrhG kommt. Aufgrund der Ausrichtung des Angebots entsprechender KI-Dienstleistungen auf Nutzer im Inland ist sowohl die Anwendbarkeit deutschen Rechts als auch die internationale Zuständigkeit deutscher Gerichte gegeben.

(2) Ergänzend ist darauf hinzuweisen, dass auch die KI-Verordnung eine Einhaltung europäischen Urheberrechts verlangt. Das Training generativer KI-Modelle ohne Einwilligung der Rechteinhaber ist demnach sowohl als Urheberrechtsverletzung als auch als Pflichtverstoß einzuordnen. Gegen derartige Verletzungen der Pflichten in der Verordnung kann unter Umständen auch privatrechtlich vorgegangen werden.

Anschlussfragen

Unter einer über die technischen und urheberrechtlichen Details hinausgehenden Perspektive sind vor allem drei in der Diskussion regelmäßig propagierte Narrative einer kritischen Prüfung zu unterziehen:

(1) Zunächst muss bezweifelt werden, dass sich die natürliche Kreativität des Menschen mittel- bis langfristig gegen die zunehmend anwachsenden Kapazitäten „künstlicher Kreativität“ behaupten können. Es ist vielmehr damit zu rechnen, dass menschliche Kreativität zunehmend verdrängt werden wird. Der Gesetzgeber kann sich deshalb nicht darauf beschränken, zunächst die weiteren Entwicklungen abzuwarten.

(2) Eine Steigerung der kreativen Produktion durch den Menschen wird aus dem Anwachsen „künstlicher“ Erzeugnisse, entgegen derzeit gängiger Prognosen, sehr wahrscheinlich nicht resultieren. Vielmehr ist damit zu rechnen, dass die Ergebnisse genuin menschlicher Kreativität in vielen Berufsgruppen und Branchen – insbesondere im Bereich der journalistischen Berichterstattung, der Unterhaltung und der Herstellung von Alltagsprodukten – in erheblichem Umfang durch generativen KI-Output ersetzt und verdrängt werden.

(3) Schließlich muss sich insbesondere der europäische Gesetzgeber die Frage stellen, ob er vor dem Hintergrund der auf anderen Gebieten kompromisslosen Sicherung regulatoriver Mindeststandards gerade für das Urheberrecht dem bereits begonnenen, globalen *race to the bottom* tatenlos zusehen möchte. Es geht dabei nicht um die Verhinderung von KI-Innovationen, sondern um faire Wettbewerbsbedingungen und einen angemessenen Ausgleich für die verwerteten Ressourcen.