

Qualitätssicherung mit Hilfe von KI?

Zur Zukunft des Peer Reviews bei wissenschaftlichen Publikationen

Zeitschriftenartikel:

Begutachtet

Begutachtung:

Prof. Dr. Ulrike Verch 

HAW Hamburg 

Erhalten: 02. April 2026

Akzeptiert: 21. April 2026

Publiziert: 08. Juli 2026

Lizenz:

© Lisa Eleonora Haase

Dieses Werk steht unter der Lizenz Creative Commons-Namensnennung 4.0 (CC-BY 4.0) International



Empfohlene Zitierung:

Haase, Lisa Eleonora (2026): Qualitätssicherung mit Hilfe von KI? Zur Zukunft des Peer Reviews bei wissenschaftlichen Publikationen, in: API Magazin, Nr. 7, Bd. 2. DOI:10.15460/apimagazin.2026.7.2.314.

Lisa Eleonora Haase^{1*} 

Studierende im 4. Semester des Masterstudiengangs Digitale Transformation der Informations- und Medienwirtschaft

¹Hochschule für Angewandte Wissenschaften Hamburg, Deutschland 

*Korrespondenz: redaktion-api@haw-hamburg.de

Zusammenfassung

Das wissenschaftliche Publizieren befindet sich im Zuge der digitalen Transformation in einem tiefgreifenden Wandel. Insbesondere der Peer-Review-Prozess als zentrales Instrument der Qualitätssicherung steht dabei unter Druck. Vor dem Hintergrund steigender Einreichungszahlen und begrenzter Kapazitäten der Reviewer*innen wird der Einsatz Künstlicher Intelligenz zunehmend als potenzielle Lösung diskutiert. Die vorliegende Arbeit untersucht, inwiefern KI den Peer-Review-Prozess verändert und welche Chancen und Herausforderungen sich daraus für die Qualität und Offenheit der Wissenschaft ergeben. Auf Basis aktueller empirischer Studien und normativer Leitlinien zeigt sich, dass KI insbesondere bei formalen und strukturellen Prüfungen Effizienzgewinne und Entlastungspotenziale bietet. Die Bewertung von Relevanz und wissenschaftlichem Erkenntnisgewinn hängt hingegen weiterhin stark von menschlicher Urteilskraft ab. Aktuell wirkt KI daher vor allem als Instrument zur Ergänzung bestehender Systeme. Inwieweit sie zu mehr Transparenz, Fairness und Integrität beiträgt, hängt maßgeblich von klaren Governance-Strukturen und der Verankerung menschlicher Verantwortung ab.

Schlagnote: Peer-Review, Künstliche Intelligenz, KI, Bias, Publikationsethik, Human-in-the-loop, HITL, Open Science

Abstract

In the wake of digital transformation, scientific publishing is undergoing profound change. The peer review process, a key instrument of quality assurance, is particularly under pressure. Against the backdrop of rising submission numbers and limited reviewer capacity, the use of artificial intelligence (AI) is increasingly being discussed as a potential solution. This paper examines how AI is changing the peer review process, and the opportunities and challenges this presents for scientific quality and openness. Drawing on current empirical studies and normative guidelines, it demonstrates that AI can deliver efficiency gains and provide much-needed relief, particularly in formal and structural reviews. However, the evaluation of novelty, relevance, and scientific knowledge gained continues to depend heavily on human judgement. AI therefore currently functions primarily as a tool for enhancing existing systems. Its contribution to greater transparency, fairness, and integrity largely depends on clear governance structures and the anchoring of human responsibility.

Keywords: Peer Review, Artificial Intelligence, Bias, Publication Ethics, Human-in-the-Loop, Open Science

1 Einleitung und Relevanz des Themas

Das wissenschaftliche Publizieren befindet sich in einem tiefgreifenden Transformationsprozess. Getrieben durch Digitalisierung, globale Vernetzung und steigende Publikationszahlen steht dabei auch der Peer-Review-Prozess als zentrales Instrument der wissenschaftlichen Qualitätssicherung zunehmend unter Druck. Gleichzeitig haben sich in den letzten Jahren leistungsfähige Systeme der künstlichen Intelligenz, insbesondere Large Language Models (LLMs), rasant entwickelt und finden bereits Anwendung in verschiedenen Bereichen der wissenschaftlichen Praxis. Künstliche Intelligenz (KI) wird als Instrument zur Effizienzsteigerung, Standardisierung und Entlastung diskutiert. Die Integration von KI wirft jedoch zugleich grundlegende Fragen nach Bewertungsmaßstäben, Transparenz, Verantwortlichkeit und wissenschaftlicher Integrität auf.

Daher untersucht die vorliegende Arbeit, inwiefern der Einsatz von KI den Peer-Review-Prozess im wissenschaftlichen Publizieren verändert und welche Chancen und Herausforderungen sich daraus für die Qualität und Offenheit der Wissenschaft ergeben. So ist es das Ziel, in dieser Arbeit die bestehenden empirischen Befunde, Standards und Anwendungsfelder für die Nutzung von KI im Peer-Review-Prozess systematisch zusammenzuführen und kritisch zu reflektieren.

2 Das wissenschaftliche Publizieren im digitalen Wandel

In den folgenden Unterkapiteln werden zunächst zentrale Begriffe geklärt und eine Auswahl aktueller Literatur zum Thema aufgezeigt.

Die Ursprünge des wissenschaftlichen Publizierens reichen bis in das 17. Jahrhundert zurück, als mit dem *Journal des sçavans* im Jahr 1665 eines der ersten wissenschaftlichen Journale gegründet wurde (Pagliaro 2020: 2). Somit haben wissenschaftliche Zeitschriften eine mehr als 350-jährige Tradition und haben sich historisch als das dominierende Format zur Verbreitung wissenschaftlicher Erkenntnisse etabliert (Ghasemi u. a. 2022: 1f.).

Jedoch erlebt dieses traditionsreiche Publikationssystem im Zuge der fortschreitenden Digitalisierung und der globalen Verbreitung des Internets einen grundlegenden Wandel. Der Anteil der Internetnutzenden belief sich 2024 weltweit laut Schätzungen auf rund 68 Prozent.¹ Diese hohe Nutzungsrate begünstigt die Entwicklung, dass die traditionelle physische Distribution weitestgehend durch Online-Publikationen ersetzt wurde (Pagliaro 2020: 1).

Doch trotz dieser Möglichkeit, wissenschaftliche Inhalte freier zugänglich zu machen, erschienen bis 2020 lediglich etwa ein Fünftel aller neu publizierten wissenschaftlichen Artikel im Open-Access-Format (Pagliaro 2020: 5). In Deutschland lässt sich in den letzten Jahren jedoch ein deutlich positiverer Trend beobachten: Im Zeitraum von 2022 bis 2026 wurden bereits rund 69 % aller Zeitschriftenartikel im Open Access veröffentlicht.²

Eng mit der Open-Access-Bewegung verbunden ist das übergreifende Rahmenkonzept der ‚Open Science‘. Open Science wird dabei als eine Sammlung von Prinzipien und Praktiken verstanden, welche darauf abzielen wissenschaftliche Forschung disziplinübergreifend frei zugänglich zu machen (UNESCO und Canadian National

1 Statista (2025): Internetnutzer - Anteil weltweit 2024, *Statista*, [online] <https://de.statista.com/statistik/daten/studie/805943/umfrage/anteil-der-internetnutzer-weltweit/> [Abrufdatum: 23.02.2026].

2 Open Access Monitor (o. J.): Open Access Status, *Open Access Monitor*, [online] <https://open-access-monitor.de/open-access> [Abrufdatum: 22.04.2026].

Commission for UNESCO 2022: 2f.). Das englische Open, also zu Deutsch die Offenheit, bezieht sich dabei nicht nur auf die finalen Forschungsergebnisse, sondern auch auf den offenen Zugang zu den erhobenen Daten, verwendete Methoden und den gesamten Produktionsprozess von wissenschaftlichem Wissen (UNESCO und Canadian National Commission for UNESCO 2022: 4).³ Somit soll Open Science die Transparenz, Nachnutzbarkeit und Kollaboration fördern (Ramachandran u. a. 2021: 1).

2.1 Der Peer-Review-Prozess im wissenschaftlichen Publizieren

Der Peer-Review-Prozess stellt einen zentralen Mechanismus der Qualitätssicherung im wissenschaftlichen Publizieren dar. Dabei ist das Ziel des Peer Reviews, eingereichte Manuskripte vor ihrer Veröffentlichung durch andere Wissenschaftler*innen beziehungsweise Expert*innen bewerten zu lassen, um so die wissenschaftliche Qualität, Validität und Relevanz der Forschung sicherzustellen (Drozd und Ladomery 2024: 1f.; Riding 2022: 1).

Doch das Peer Review ist nicht ausschließlich als Kontrollinstrument zu verstehen, sondern auch als ein formativer Prozess, welcher durch konstruktives Feedback zur inhaltlichen und formalen Verbesserung von wissenschaftlichen Arbeiten beitragen soll (Drozd und Ladomery 2024: 2f.). Trotz seiner zentralen Relevanz für die Wissenschaft ist das Peer-Review-System mit einer Reihe von Herausforderungen konfrontiert. So ist ein häufig genannter Kritikpunkt die hohe Zeitintensität des Verfahrens, die sich insbesondere vor dem Hintergrund steigender Einreichungszahlen weiter verschärft (Drozd und Ladomery 2024: 6). Gleichzeitig wird zunehmend ein Mangel an qualifizierten Reviewer*innen festgestellt, was zu einer Überlastung der verfügbaren Gutachter*innen und zu Verzögerungen im Publikationsprozess führen kann (Drozd und Ladomery 2024: 6f.).

Darüber hinaus wird auch die Subjektivität des Peer Reviews kritisch diskutiert. So können Bewertungen durch persönliche Präferenzen oder institutionelle Zugehörigkeiten der Reviewer*innen beeinflusst werden (Drozd und Ladomery 2024: 7f.). Ein weiterer Kritikpunkt ist die geringe Transparenz in diesem Verfahren, da Gutachten in der Regel nicht veröffentlicht werden, was die Entscheidungsprozesse für Außenstehende schwer nachvollziehbar macht (Drozd und Ladomery 2024: 8).

Diese strukturellen Herausforderungen bilden den Ausgangspunkt für aktuelle Diskussionen, in denen digitale und insbesondere KI-gestützte Ansätze zunehmend als mögliche Lösungsoptionen diskutiert werden. Vor diesem Hintergrund beschäftigt sich der folgende Abschnitt damit, wie die aktuelle Forschung den Einsatz von KI im Peer-Review-Prozess einordnet.

2.2 Aktuelle Studien und Veröffentlichungen zum Einsatz von KI im Peer Review

Die in Kapitel 2.1 skizzierten strukturellen Herausforderungen des Peer-Review-Prozesses bilden den Ausgangspunkt der Forschung zum Einsatz von KI. Dieser wird dabei nicht primär als technologische Innovation um ihrer selbst willen diskutiert, sondern als potenzieller Ansatz zur strukturellen Weiterentwicklung des bestehenden Systems (BaHammam 2025: 159f.; Kankanhalli 2024: 80).

Eine zentrale Veröffentlichung zur Rolle generativer KI im Begutachtungsprozess stellt das Editorial ‚Peer Review in the Age of Generative AI‘ von Kankanhalli

³ Open Science AG (2014): Mission Statement, *AG Open Science*, [online] <https://ag-openscience.de/mision-statement/> [Abrufdatum: 23.02.2026].

(2024) dar. Der Beitrag ist konzeptionell angelegt und untersucht, inwiefern Large Language Models einzelne Phasen des Peer-Review-Prozesses unterstützen oder teilweise automatisieren könnten (Kankanhalli 2024: 76f.). Die Autorin differenziert systematisch zwischen dem Pre-Peer-Review-Screening und der eigentlichen Begutachtungsphase (Kankanhalli 2024: 77). Während formale Prüfungen, wie beispielsweise die Überprüfung, ob das Manuskript den Formatvorgaben für Struktur, Literaturangaben und Metadaten der Publikationsplattform entspricht, ein hohes Automatisierungspotenzial aufweisen, gelten Bewertungen von Relevanz und Signifikanz der wissenschaftlichen Beiträge als nur begrenzt automatisierbar, da sie kontextabhängige Urteilsfähigkeit erfordern (Kankanhalli 2024: 79). Vor diesem Hintergrund plädiert Kankanhalli für das sogenannte ‚human-in-the-loop‘-Modell, bei dem die KI als unterstützendes Instrument fungiert, während die letztliche Bewertungsverantwortung beim Menschen verbleibt (Kankanhalli 2024: 80f.).

Auch BaHamam setzt sich in dem 2025 erschienen ‚Peer Review in the Artificial Intelligence Era: A Call for Developing Responsible Integration Guidelines‘ mit der Integration von KI in den Begutachtungsprozess auseinander. Der Beitrag ist konzeptionell-analytisch angelegt und argumentiert für klar definierte Integrationsrichtlinien. Empirische Untersuchungen deuten darauf hin, dass KI-generierte Reviews in bestimmten Dimensionen inhaltliche Überschneidungen mit menschlichen Gutachten aufweisen, insbesondere bei methodischen und technischen Aspekten (BaHamam 2025: 160). Gleichzeitig werden Grenzen bei der Bewertung von Neuheit und wissenschaftlicher Relevanz hervorgehoben (BaHamam 2025: 160f.). Auch BaHamam betont, dass die KI unterstützend eingesetzt werden sollte, während die finale Entscheidungsverantwortung bei den Reviewer*innen bleiben sollte (BaHamam 2025: 160ff.).

Eine umfassende und empirisch fundierte Analyse liefern Mann et al. in dem 2024 publizierten ‚AI and the Future of Academic Peer Review‘. Die Autor*innen stellen die normativen Zielsetzungen des Peer Reviews den in Kapitel 2.1 beschriebenen Herausforderungen des bestehenden Systems gegenüber (Mann u. a. 2025: 2ff.). Auf dieser Grundlage argumentieren sie, dass der Einsatz von LLMs nicht an einem idealisierten Peer Review gemessen werden dürfe, sondern am real existierenden Status quo (Mann u. a. 2025: 5). Gleichzeitig diskutieren die Autor*innen Risiken wie Halluzinationen, algorithmische Verzerrungen, Vertraulichkeitsfragen und die Gefahr eines Automation Bias (Mann u. a. 2025: 11ff.). Dieser Automation Bias beschreibt laut der Bundeszentrale für politische Bildung „das Phänomen, dass Menschen den Vorschlägen und Entscheidungen von Maschinen mehr vertrauen als ihren eigenen“.⁴ Abschließend plädieren sie für sorgfältig evaluierte, regulierte Pilotanwendungen, in denen KI als unterstützendes Instrument getestet wird (Mann u. a. 2025: 24f.).

Demgegenüber vertreten Wiechert, Wang und Chapman (2024) in ihrem Editorial ‚When Will Peer-Review Become A.I.-Review?‘, eine stärker normativ zugespitzte Position. Ausgehend von den strukturellen Herausforderungen medizinischer Fachzeitschriften erkennen die Autor*innen zwar an, dass KI bei formalen Prüfungen Effizienzgewinne ermöglichen kann (Wiechert u. a. 2024: 1463). Jedoch betonen sie auch, dass wissenschaftliche Begutachtung nicht allein auf der Analyse bestehenden Wissens beruhen würde, sondern auch die Einschätzung von Innovation und potenziellem Erkenntnisgewinn umfasse (Wiechert u. a. 2024). So ordnen die

4 Bundeszentrale für politische Bildung (2025): Automation Bias. Ausgewählte Beiträge von Instagram, bpb.de, [online] <https://www.bpb.de/lernen/digitale-bildung/werkstatt/569392/automation-bias/> [Abrufdatum: 22.04.2026].

Autor*innen eine vollständig KI-basierte Publikationskette als problematisch ein und mit Blick auf wissenschaftliche Glaubwürdigkeit als unvereinbar mit dem Selbstverständnis wissenschaftlicher Praxis (Wiechert u. a. 2024).

Zusammenfassend lässt sich festhalten, dass ein übergreifender Konsens darin besteht, dass KI-Systeme gegenwärtig als unterstützende Werkzeuge verstanden werden. Die Integration klarer Governance- und Transparenzregelungen sollte dabei nicht vernachlässigt werden und stellt eine zentrale Notwendigkeit dar (BaHammam 2025: 162; Mann u. a. 2025: 24). Für die vorliegende Arbeit ergibt sich daraus, dass der Einsatz von KI nicht nur deskriptiv zu betrachten ist. Vielmehr muss geklärt werden, welche Rolle KI im Peer-Review-Prozess einnehmen kann und wo die funktionalen, sowie normativen Grenzen in der Anwendung liegen.

3 Einfluss von Künstlicher Intelligenz auf den Peer-Review-Prozess

Vor diesem Hintergrund widmet sich das folgende Kapitel der systematischen Einordnung von KI im Peer-Review-Kontext.

3.1 Grundlagen und Begriffsbestimmungen

Es gibt bereits regelbasierte Automatisierungsinstrumente, die seit Jahren im redaktionellen Umfeld eingesetzt werden. Dazu zählen unter anderem Plagiatserkennungssysteme, statistische Prüfprogramme oder Struktur- und Formatierungsschecks, welche auf definierte Kriterien zurückgreifen, und die formalen Aspekte eines Manuskripts überprüfen (Daskaliuk u. a. 2025: 4; Kousha und Thelwall 2023: 6). Dahingegen ist generative KI prinzipiell in der Lage, argumentative Strukturen zu bewerten, methodische Schwächen zu identifizieren und Entscheidungsprognosen im Sinne einer ‚Accept/Reject-Empfehlung‘ abzugeben (Bharti u. a. 2023: 65ff.; Mann u. a. 2025: 17). Der zentrale Unterschied liegt somit weniger im Grad der Automatisierung als vielmehr in der Art der Tätigkeit. Während etablierte Systeme primär Formalien prüfen, kann generative KI die wissenschaftliche Urteilsbildung der Reviewer*innen beeinflussen.

In der wissenschaftlichen Literatur wird überwiegend zwischen einer ergänzenden und einer substitutiven Nutzung von KI unterschieden. Dabei wird unter einer ergänzenden Nutzung der Einsatz von KI als unterstützendes Instrument verstanden, wobei die finale Entscheidungsverantwortung bei den Reviewer*innen verbleibt (BaHammam 2025: 162; Daskaliuk u. a. 2025: 6; Kankanalli 2024: 81; Seghier 2025: 112). Eine substitutive Nutzung bezeichnet KI-Systeme, die autonom über die Annahme und Ablehnung wissenschaftlicher Beiträge entscheiden. Dieses Szenario wird, wie in Kapitel 2.2 besprochen, in der Literatur überwiegend kritisch beurteilt. Neben der funktionalen Abgrenzung spielt auch die normative Dimension eine zentrale Rolle. So wird in der Literatur betont, dass KI weder als Autorin noch als verantwortliche Akteurin im Peer-Review-Prozess auftreten kann, denn sie kann weder Rechenschaftspflicht noch institutionelle Verantwortlichkeit tragen (Leung u. a. 2023: 3f.). Ebenso wird auf die Bedeutung von Vertraulichkeit hingewiesen, insbesondere da das Einspeisen unveröffentlichter Manuskripte in externe KI-Systeme potenziell gegen zentrale Prinzipien des Peer Reviews verstoßen kann (BaHammam 2025: 111). So könnten dadurch die Autorenrechte in Teilen verletzt werden. Dadurch wird deutlich, dass die begriffliche Bestimmung von KI im Peer-

Review-Kontext nicht ausschließlich technisch, sondern untrennbar mit Fragen der Verantwortung, Integrität und wissenschaftlichen Ethik verbunden ist.

Dieser Diskurs bildet die Grundlage für die folgende Analyse konkreter Anwendungsfelder von KI im Peer-Review-Prozess.

3.2 Anwendungsfelder von KI im Peer-Review-Prozess

An welchen konkreten Stellen des Peer-Review-Prozesses wird KI bereits oder perspektivisch eingesetzt? Die Recherche zeigt, dass KI-Anwendungen entlang verschiedener Phasen des Prozesses verortet werden können und damit von der initialen Manuskripteinschätzung über die Auswahl geeigneter Gutachter*innen bis hin zur Analyse von Textqualität und Struktur eingesetzt werden (Kousha und Thelwall 2023: 6ff.).

3.2.1 Manuskripteinschätzung und Vorabprüfung

So stellt die automatisierte Vorabprüfung eingereichter Manuskripte ein zentrales Anwendungsfeld dar. Demnach können KI-gestützte Systeme formale und methodische Kriterien vorab prüfen, bevor eine externe Begutachtung eingeleitet wird, und können damit den redaktionellen Prozess beschleunigen (Bauchner und Rivara 2024: 1f.).

Bereits etablierte Anwendungen umfassen die Plagiatserkennung, etwa durch Systeme wie ‚iThenticate‘, die eingereichten Texte mit umfangreichen Datenbanken vergleichen.⁵ Gleichzeitig existieren Programme zur automatisierten statistischen Prüfung, wie ‚StatCheck‘ oder ‚StatReviewer‘ (Kousha und Thelwall 2023: 6).

Mit dem Einsatz generativer KI erweitert sich dieses Anwendungsfeld um inhaltlich-analytische Funktionen. So können LLMs Manuskripte zusammenfassen, argumentative Strukturen sichtbar machen und potenzielle methodische Schwächen identifizieren (BaHamam 2025: 160). Empirische Untersuchungen zeigen, dass KI-generierte Reviews in Teilen inhaltliche Überschneidungen mit menschlichen Gutachten aufweisen, insbesondere bei der Identifikation technischer oder struktureller Aspekte (Mann u. a. 2025: 24). Zudem zeigen Systeme wie ‚PEERRec‘, dass auf Basis von Paper- und Reviewtexten Empfehlungsscores und Entscheidungsprognosen modelliert werden können, welche die Reviewer*innen bei ihrer Entscheidungsfindung unterstützen können (Bharti u. a. 2023: 69).

3.2.2 Auswahl von Reviewer*innen

Ein weiteres (potenzielles) Anwendungsfeld liegt in der Identifikation geeigneter Gutachter*innen. Die Auswahl passender Reviewer*innen ist ein ressourcenintensiver Prozess, da fachliche Expertise und potenzielle Interessenkonflikte berücksichtigt werden müssen (BaHamam 2025: 159f.; Mann u. a. 2025: 5f.). KI-gestützte Systeme können hier unterstützen, indem sie Manuskriptinhalte mit bibliometrischen Profilen und Publikationsdaten potenzieller Reviewer*innen abgleichen.

So nutzen Systeme wie der ‚Clarivate Reviewer Locator‘ oder ‚Elseviers Reviewer Finder‘ textbasierte Matching-Verfahren zur Identifikation geeigneter Expert*innen (Kousha und Thelwall 2023: 7). Auch KI-gestützte Systeme wie ‚AIRA‘ von ‚Frontiers‘ dienen der Strukturierung redaktioneller Workflows und der Unterstützung bei der Reviewer*innen-Auswahl.

⁵ Turnitin (o. J.): iThenticate. Publish with confidence using iThenticate, [online] <https://www.turnitin.com/products/ithenticate/> [Abrufdatum: 24.02.2026].

3.2.3 Analyse von Textqualität und Struktur

Ein drittes Anwendungsfeld betrifft die Analyse sprachlicher und struktureller Aspekte wissenschaftlicher Texte. KI-Systeme können Grammatik, Stil, Kohärenz und terminologische Konsistenz prüfen und dadurch sowohl Autor*innen als auch Gutachter*innen unterstützen (Daskaliuk u. a. 2025: 5; Seghier 2025: 108). Dies kann auch insbesondere im internationalen Kontext zu einer Verbesserung der Verständlichkeit eingereichter Arbeiten beitragen (Seghier 2025: 106).

4 Herausforderungen und Risiken des KI-gestützten Peer Reviews

Die Literatur weist auf eine Reihe substanzieller Herausforderungen und Risiken hin, die mit dem Einsatz künstlicher Intelligenz im Peer-Review-Prozess verbunden sind. So geht es um Verzerrungen durch Algorithmen (Bias), Transparenz, Verantwortlichkeit und neue Formen der Manipulation.

4.1 Bias und Diskriminierung

So betrifft ein zentrales Risiko die Möglichkeit, dass KI-Systeme bestehende Verzerrungen nicht reduzieren, sondern reproduzieren oder sogar verstärken. Demnach können Trainingsdaten von KIs, die beispielsweise überproportional aus bestimmten Institutionen oder Sprachräumen stammen, diese Muster in die Bewertungsprozesse einfließen lassen (Seghier 2025: 108).

Vor diesem Hintergrund argumentieren Befürworter der KI, dass algorithmische Systeme ebenso anfällig für Verzerrungen sein können wie menschliche Gutachter*innen (Kankanhalli 2024: 80; Mann u. a. 2025: 5). Zwar wird in der Literatur die Hoffnung formuliert, dass KI textzentrierter und weniger personenbezogen operieren könnte (Bauchner und Rivara 2024: 1f.), zugleich wird jedoch betont, dass algorithmische Fairness keine automatische Eigenschaft technischer Systeme ist, sondern kontinuierlicher Überprüfung und Anpassung bedarf (Seghier 2025: 113).

4.2 Transparenz und Nachvollziehbarkeit

Ein weiteres Spannungsfeld betrifft die Transparenz algorithmischer Entscheidungsprozesse. LLMs operieren auf Basis hochkomplexer neuronaler Strukturen, deren interne Gewichtungen kaum nachvollziehbar sind. Diese sogenannte Black-Box-Problematik wird in der Literatur wiederholt als Herausforderung für wissenschaftliche Bewertungsprozesse beschrieben (Leung u. a. 2023: 1f.; Seghier 2025: 104f.). Wenn nicht klar ersichtlich ist, auf welcher Grundlage eine KI bestimmte Kritikpunkte generiert oder eine Entscheidungsprognose abgibt, stellt sich die Frage nach der Legitimität dieser Bewertungen.

Darüber hinaus betrifft die Intransparenz auch Fragen der Verantwortlichkeit. Während menschliche Gutachter*innen für ihre Einschätzungen einstehen, kann ein KI-System weder Rechenschaft ablegen noch ethische Verantwortung übernehmen (Leung u. a. 2023: 3). Aus diesem Grund betonen zahlreiche Beiträge, dass die letztliche Entscheidungsverantwortung zwingend bei den Reviewer*innen verbleiben muss (Kankanhalli 2024: 80f.; Seghier 2025: 112).

4.3 Prompt Injections

Eine weitere Verwundbarkeit KI-gestützter Begutachtungssysteme sind sogenannte Prompt-Injection-Angriffe. Der Artikel „Prompt Injection Attacks on LLM-Assisted

Peer Review' zeigt, dass Large Language Models, die zur Generierung von Reviewberichten eingesetzt werden, anfällig für versteckte Instruktionen sind, die direkt im eingereichten Manuskript eingebettet werden können (Collu u. a. 2025: 2). Dabei können Autor*innen einen vermeintlich ‚unsichtbaren‘ Text in PDF-Dokumente integrieren, etwa in Form von weißer Schrift auf weißem Untergrund oder anderen Formatierungen. Dieser ist für menschliche Reviewer*innen nicht erkennbar, wird jedoch vom Sprachmodell als reguläre Eingabe verarbeitet (Collu u. a. 2025: 5). Damit besteht die Möglichkeit mithilfe von KI generierte Reviews gezielt zu manipulieren (Collu u. a. 2025: 28).

4.4 Auswirkungen auf wissenschaftliche Integrität und Vertrauen

Letztendlich kumulieren die genannten Risiken in einer übergeordneten Frage: Welche Auswirkungen hat der Einsatz von KI auf die wissenschaftliche Integrität und das Vertrauen in das Publikationssystem? Das Peer Review ist nicht nur ein technischer Qualitätssicherungsmechanismus, sondern auch ein soziales Vertrauensarrangement zwischen Autor*innen, Gutachter*innen, Herausgeber*innen und der wissenschaftlichen Gemeinschaft. Wenn dieser Prozess zunehmend von algorithmischen Systemen geprägt wird, könnte sich auch die Wahrnehmung von Verantwortlichkeit und Legitimität ändern. Seghier (2025) warnt in diesem Zusammenhang vor einem Szenario, in dem wissenschaftliche Kommunikation zu einer „awkward conversation between an AI writing a paper and an AI evaluating that paper“ (Seghier 2025: 104) degenerieren könnte. Eine solche Entwicklung würde die Rolle menschlicher Urteilkraft erodieren und das Vertrauen in die Authentizität wissenschaftlicher Begutachtung untergraben.

Insgesamt verdeutlicht die Literatur, dass der Einsatz von KI im Peer Review nicht nur eine technische, sondern eine institutionelle und ethische Transformation darstellt. Die Herausforderung besteht darin, Effizienz- und Qualitätssteigerungen zu realisieren, ohne dabei die normativen Grundlagen wissenschaftlicher Begutachtung zu gefährden. Zu diesen Grundlagen können auch die bestehenden Standards gezählt werden, die im nächsten Abschnitt besprochen werden sollen.

5 Standards und Richtlinien

Die Core Practices des Committee on Publication Ethics (COPE) bilden einen zentralen normativen Referenzrahmen für die Sicherung wissenschaftlicher Integrität im Publikationswesen. Sie wurden 2017 als Ersatz des bisherigen „Code of Conduct“, eingeführt, um die Erwartungen an professionelle Standards im wissenschaftlichen Publizieren klarer und zugleich flexibler zu formulieren.⁶ Die Core Practices gelten für teilnehmende Herausgeber*innen, Reviewer*innen, Journale, Verlage und Institutionen gleichermaßen. Sie sind nicht als rechtlich bindendes Regelwerk, sondern als professioneller Orientierungsrahmen konzipiert. Ihr Ziel besteht darin, Transparenz, Integrität und eine verantwortungsvolle Publikationspraxis zu fördern. In diesem Sinne bieten die Core Practices eine institutionelle Struktur, in die auch der Einsatz von KI im Peer-Review-Prozess normativ eingeordnet werden kann.

Für den KI-Kontext sind insbesondere die COPE-Core-Practices zu Transparenz, Peer-Review-Prozessen und Rechenschaftspflicht relevant. COPE fordert klar

⁶ Committee on Publication Ethics (2017): Why the Code of Conduct for Journal Editors was replaced with the Core Practices, o. S., [online] <https://publicationethics.org/home/api/Dokumente/Magazin/2026-Sommer/articles/2026-06-haase-ki/media/827/download?attachment> [Abrufdatum: 26.04.2026].

geregelte und transparent beschriebene Begutachtungsverfahren sowie definierte Prozesse im Umgang mit Interessenkonflikten und Fehlverhalten. Im Kontext der KI-gestützten Begutachtung gewinnen diese Anforderungen an Bedeutung, da Fragen der Verantwortungszuschreibung und der Offenlegung algorithmischer Unterstützung in den Vordergrund treten. Die Core Practices legen damit nahe, dass eine Integration von KI in den Peer-Review-Prozess nur im Rahmen klarer Governance-Strukturen erfolgen kann.

Die praktische Relevanz dieser Governance-Fragen zeigt sich exemplarisch im *COPE Case 26-03 (2026)*. Herausgeber*innen berichten darin von einem deutlichen Anstieg mutmaßlich KI-generierter Peer Reviews. Die betreffenden Gutachten erschienen zunächst detailliert, enthielten jedoch zahlreiche inhaltlich unzutreffende oder irrelevante Kritikpunkte. KI-Detektionstools wiesen bei ihnen teilweise einen Anteil KI-generierten Textes von über 80 % aus. Dies verstieß gegen die jeweiligen Standards des Journals, die lediglich eine begrenzte sprachliche Überarbeitung mit Offenlegung zuließ und das Hochladen unveröffentlichter Manuskripte in generative Systeme untersagte. In seiner Stellungnahme betont COPE, dass das Peer Review primär der Einholung fachlicher Expertise dient und KI diese gegenwärtig nicht zuverlässig ersetzen kann. Entsprechend empfiehlt COPE klare Richtlinien zur zulässigen KI-Nutzung, verbindliche Offenlegungspflichten bei der Nutzung von KI sowie Maßnahmen bei Verstößen gegen diese Regularien.⁷

Auch die Deutsche Forschungsgemeinschaft (DFG) hat Leitlinien für den Einsatz von KI in Begutachtungsverfahren formuliert, die insbesondere Vertraulichkeit, Transparenz, Qualitätssicherung und Verantwortung in den Fokus rücken. Demnach sei der Einsatz von KI-Systemen ausschließlich in unterstützender Funktion zulässig, während die inhaltliche Bewertung weiterhin allein den Gutachter*innen obliegen sollte. So betreffen die zentralen Anforderungen den Schutz sensibler Daten, die Offenlegung des KI-Einsatzes sowie die verpflichtende kritische Prüfung KI-generierter Inhalte, um Fehler und Verzerrungen zu vermeiden. Dabei betont die DFG zudem, dass die wissenschaftliche und ethische Verantwortung nicht auf KI-Systeme übertragen werden dürfe, sondern uneingeschränkt bei den Gutachter*innen verbleibe, wodurch der Einsatz von KI klar in ein Rahmenwerk menschlicher Kontrolle und institutioneller Regulierung eingebettet werden soll.⁸

6 Diskussion und Beantwortung der Forschungsfrage

Die vorhergehende Analyse hat gezeigt, dass KI-Systeme bereits heute in verschiedenen Phasen des Begutachtungsprozesses eingesetzt werden. Gleichzeitig wurde jedoch auch deutlich, dass zentrale Elemente wissenschaftlicher Bewertung weiterhin nur eingeschränkt algorithmisch erfassbar sind (Kankanhalli 2024: 80f.). Dieser Diskurs wirft Fragen zu dem Verhältnis von Qualität, Effizienz, Offenheit und Kontrolle auf, die in den folgenden Unterkapitel 6.1 und 6.2 diskutiert werden sollen. Anschließend soll die in Kapitel 1 gestellte Forschungsfrage beantwortet werden.

7 Committee on Publication Ethics (2026): Editors suspect that reviewers are using artificial intelligence, *COPE: Committee on Publication Ethics*, [online] <https://publicationethics.org/guidance/case/editors-suspect-reviewers-are-using-artificial-intelligence> [Abrufdatum: 23.02.2026].

8 Deutsche Forschungsgesellschaft (2026): Leitlinie zur Nutzung von Künstlicher Intelligenz in der Begutachtung, DFG-Vordruck 4.04, [online] <https://www.dfg.de/resource/blob/387664/4-04-de.pdf> [Abrufdatum: 17.04.2026].

6.1 Qualität versus Effizienz

Eines der am häufigsten genannten Argumente für den Einsatz von KI ist die Aussicht auf Effizienzgewinne. Angesichts der in Kapitel 2.1 besprochenen steigenden Einreichungszahlen scheint dies zunächst eine pragmatische Antwort für das überlastete System zu sein. Die vorhergegangene Analyse zeigt jedoch gleichzeitig, dass eine Effizienzsteigerung nicht automatisch eine Qualitätsverbesserung bedeuten muss. Es besteht daher die Gefahr, dass sich die Bewertungsmaßstäbe von einer kontextgebundenen Urteilsbildung hin zu stärker formal messbaren Kriterien verschieben. Eine solche Verschiebung könnte langfristig zu einer Normierung wissenschaftlicher Beiträge führen und experimentelle oder theoretisch unkonventionelle Arbeiten benachteiligen. Effizienzgewinne dürfen daher nicht isoliert betrachtet werden. Sie müssen vielmehr kritisch reflektiert werden.

6.2 Offenheit versus Kontrolle

Es gibt auch widersprüchliche Entwicklungen in Bezug auf die Offenheit und Kontrolle von wissenschaftlichen Publikationen. So könnten die KI-gestützten Analysen Open Science unterstützen, indem Bewertungsmuster systematisch erfasst und implizite Kriterien sichtbar gemacht werden (Kousha und Thelwall 2023: 10). Theoretisch eröffnet dies die Möglichkeit einer stärker datenbasierten Reflexion von Peer-Review-Praktiken.

Doch gleichzeitig kann der Einsatz von KI auch neue Formen der Intransparenz erzeugen. So machen die in Kapitel 4 diskutierten Problematiken deutlich, dass KI-Systeme nicht nur unterstützende Werkzeuge sind, sondern auch neue Angriffsflächen im Publikationsprozess bieten. Damit wird deutlich, dass Offenheit im Sinne von Open Science nur dann wirksam ist, wenn sie durch institutionelle Kontroll- und Verantwortungsstrukturen unterstützt wird. Transparenz allein genügt nicht, sie muss mit klaren Verantwortlichkeits- und Prüfmechanismen einhergehen.

6.3 Beantwortung der Forschungsfrage

Vor dem Hintergrund der vorhergegangenen Analyse und Diskussion lässt sich die zu Beginn dieser Arbeit gestellte Forschungsfrage ‚Inwiefern verändert der Einsatz von KI den Peer-Review-Prozess im wissenschaftlichen Publizieren? Und welche Chancen und Herausforderungen ergeben sich daraus für die Qualität und Offenheit der Wissenschaft?, differenziert beantworten. So verändert der Einsatz von KI den Peer-Review-Prozess vor allem strukturell, indem er bestimmte Aufgaben standardisiert, beschleunigt und datenbasiert unterstützt. Diese Veränderungen können zur Effizienzsteigerung und formalen Qualitätssicherung beitragen. Jedoch bergen sie gleichzeitig das Risiko die gesetzten Bewertungsmaßstäbe zu verschieben, neue Manipulationsmöglichkeiten zu eröffnen und Verantwortungsstrukturen zu verkomplizieren.

KI ist somit weder als rein technisches Hilfsmittel noch als autonome Bewertungsinstanz zu verstehen, sondern als Element in einem komplexen sozialen und institutionellen Gefüge. Die entscheidende Frage lautet daher nicht, ob KI eingesetzt werden soll, denn die Integration in den Prozess ist bereits angelaufen, sondern unter welchen normativen Bedingungen ihre Nutzung vertretbar ist. Eine unreflektierte Technologisierung des Peer Reviews könnte die Qualität und Glaubwürdigkeit der Wissenschaft untergraben. Eine kontrollierte, transparent regulierte und klar begrenzte Integration kann hingegen unterstützende Funktionen übernehmen, ohne die zentrale Rolle menschlicher Urteilskraft zu ersetzen.

7 Fazit und Ausblick

Abschließend zeigt sich, dass der Einsatz von KI im Rahmen des Peer Reviews neben dem technischen Fortschritt auch eine institutionelle Weichenstellung bedeutet. So birgt KI das Potenzial, Effizienz, Standardisierung und Transparenz zu fördern sowie bestehende Überlastungsprobleme zu lösen. Gleichzeitig besteht die Gefahr einer Verschiebung der Bewertungsmaßstäbe hin zu stärker formalisierten Kriterien. Ob KI langfristig zur Stärkung von Qualität und Offenheit beiträgt, hängt daher maßgeblich von klaren Governance-Strukturen, verbindlichen Offenlegungsregeln und einer konsequenten Verankerung menschlicher Verantwortung im Peer-Review-Prozess ab.

Zukünftige Forschung könnte insbesondere die disziplinspezifischen Effekte KI-gestützter Begutachtungsverfahren empirisch untersuchen, die Auswirkungen auf die Innovationsfähigkeit und Diversität analysieren, Maßnahmen zum Schutz vor Manipulationen entwickeln und bestehende institutionelle Standards weiterentwickeln. Die Transformation des Peer Reviews durch KI ist somit kein abgeschlossener Wandel, sondern ein fortlaufender Balanceakt zwischen technologischem Fortschritt und wissenschaftlicher Verantwortung.

Literatur

- BaHamam, Ahmed (2025): Peer Review in the Artificial Intelligence Era: A Call for Developing Responsible Integration Guidelines, in: *Nature and Science of Sleep*, Jg. 17, S. 159–164, DOI: [10.2147/NSS.S513872](https://doi.org/10.2147/NSS.S513872).
- Bauchner, Howard; Rivara, Frederick P. (2024): Use of Artificial Intelligence and the Future of Peer Review, in: *Health Affairs Scholar*, Jg. 2, Nr. 5, S. 1–3, DOI: [10.1093/haschl/qxae058](https://doi.org/10.1093/haschl/qxae058).
- Bharti, Prabhat Kumar; Ghosal, Tirthankar; Agarwal, Mayank; u. a. (2023): PEERRec: An AI-based Approach to Automatically Generate Recommendations and Predict Decisions in Peer Review, in: *International Journal on Digital Libraries*, Jg. 25, Nr. 1, S. 55–72, DOI: [10.1007/s00799-023-00375-0](https://doi.org/10.1007/s00799-023-00375-0).
- Collu, Matteo Gioele; Salviati, Umberto; Confalonieri, Roberto; u. a. (2025): Publish to Perish: Prompt Injection Attacks on LLM-Assisted Peer Review, arXiv, DOI: [10.48550/arXiv.2508.20863](https://doi.org/10.48550/arXiv.2508.20863).
- Doskaliuk, Bohdana; Zimba, Olena; Yessirkepov, Marlen; u. a. (2025): Artificial Intelligence in Peer Review: Enhancing Efficiency While Preserving Integrity, in: *Journal of Korean Medical Science*, Jg. 40, Nr. 7, S. 1–9, DOI: [10.3346/jkms.2025.40.e92](https://doi.org/10.3346/jkms.2025.40.e92).
- Drozd, John A.; Lodomery, Michael R. (2024): The Peer Review Process: Past, Present, and Future, in: *British Journal of Biomedical Science*, Jg. 81, S. 1–10, DOI: [10.3389/bjbs.2024.12054](https://doi.org/10.3389/bjbs.2024.12054).
- Ghasemi, Asghar; Mirmiran, Parvin; Kashfi, Khosrow; u. a. (2022): Scientific Publishing in Biomedicine: A Brief History of Scientific Journals, in: *International Journal of Endocrinology and Metabolism*, Jg. 21, Nr. 1, S. 1–10, DOI: [10.5812/ijem-131812](https://doi.org/10.5812/ijem-131812).
- Kankanhalli, Atreyi (2024): Peer Review in the Age of Generative AI, in: *Journal of the Association for Information Systems*, Jg. 25, Nr. 1, S. 76–84, DOI: [10.17705/1jais.00865](https://doi.org/10.17705/1jais.00865).

- Kousha, Kayvan; Thelwall, Mike (2023): Artificial Intelligence to Support Publishing and Peer Review: A Summary and Review, in: *Learned Publishing*, Jg. 37, Nr. 1, S. 4–12, DOI: [10.1002/leap.1570](https://doi.org/10.1002/leap.1570).
- Leung, Tiffany I.; De Azevedo Cardoso, Taiane; Mavragani, Amaryllis; u. a. (2023): Best Practices for Using AI Tools as an Author, Peer Reviewer, or Editor, in: *Journal of Medical Internet Research*, Jg. 25, S. 1–8, DOI: [10.2196/51584](https://doi.org/10.2196/51584).
- Mann, Sebastian Porsdam; Aboy, Mateo; Seah, Joel Jiehao; u. a. (2025): AI and the Future of Academic Peer Review, arXiv, DOI: [10.48550/arXiv.2509.14189](https://doi.org/10.48550/arXiv.2509.14189).
- Pagliaro, Mario (2020): Publishing Scientific Articles in the Digital Era, in: *Open Science Journal*, Jg. 5, Nr. 3, S. 1–12, DOI: [10.23954/osj.v5i3.2617](https://doi.org/10.23954/osj.v5i3.2617).
- Ramachandran, Rahul; Bugbee, Kaylin; Murphy, Kevin (2021): From Open Data to Open Science, in: *Earth and Space Science*, Jg. 8, Nr. 5, S. 1–17, DOI: [10.1029/2020EA001562](https://doi.org/10.1029/2020EA001562).
- Riding, James B. (2022): An Evaluation of the Process of Peer Review, in: *Palynology*, Jg. 47, Nr. 1, S. 1–8, DOI: [10.1080/01916122.2022.2151052](https://doi.org/10.1080/01916122.2022.2151052).
- Seghier, Mohamed L. (2025): AI-powered Peer Review Needs Human Supervision, in: *Journal of Information, Communication and Ethics in Society*, Jg. 23, Nr. 1, S. 104–116, DOI: [10.1080/01916122.2022.2151052](https://doi.org/10.1080/01916122.2022.2151052).
- UNESCO; Canadian National Commission for UNESCO (2022): *An Introduction to the UNESCO Recommendation on Open Science*, (technical report), DOI: [10.54677/XOIR1696](https://doi.org/10.54677/XOIR1696).
- Wiechert, Karsten; Wang, Jeffrey C.; Chapman, Jens R. (2024): When Will Peer-Review Become A.I.-Review?, in: *Global Spine Journal*, Jg. 14, Nr. 5, S. 1463, DOI: [10.1177/21925682241242892](https://doi.org/10.1177/21925682241242892).