Alles klar! Ich erstelle einen ausführlichen Forschungsbericht auf Deutsch zum Thema KI-unterstützte Literaturrecherche in der Forschung unter Verwendung ausschließlich aktueller wissenschaftlicher Literatur. Der Bericht wird systematisch gegliedert sein und eine vollständige Literaturliste enthalten. Ich melde mich, sobald der Bericht bereit ist.

* **KI-unterstützte Literaturrecherche in der Forschung**

**Einleitung**

Die systematische **Literaturrecherche** ist ein grundlegender Schritt im Forschungsprozess, doch angesichts der **Informationsflut** in vielen Fachgebieten wird sie zunehmend zur Herausforderung. Jährlich erscheinen unzählige neue Publikationen, was es Forschenden erschwert, den Überblick zu behalten und relevante Arbeiten effizient zu identifizieren ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=%5B1%20,generated%20by%20skilled%20human%20researchers)). Studien zeigen, dass die Geschwindigkeit wissenschaftlicher Veröffentlichungen kontinuierlich steigt ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=%5B1%20,generated%20by%20skilled%20human%20researchers)). In diesem Kontext wächst das Interesse an **KI-gestützten Methoden**, welche die Recherche erleichtern und beschleunigen könnten. Insbesondere seit der breiten Verfügbarkeit von **generativer KI** wie *ChatGPT* (veröffentlicht von OpenAI Ende 2022) wird intensiv darüber diskutiert, wie solche *Large Language Models (LLMs)* die wissenschaftliche Literaturrecherche unterstützen können. Erste Anwendungen und Untersuchungen deuten darauf hin, dass KI-Tools in der Tat helfen können, beispielsweise indem sie Literatur schneller durchforsten und zusammenfassen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Forschende%20von%20zeitaufw%C3%A4ndigen%20Routinet%C3%A4tigkeiten%20entlastet)).

Der vorliegende Bericht bietet einen umfassenden Überblick über das Thema *KI-unterstützte Literaturrecherche in der Forschung*. Zunächst werden Hintergrund und grundlegende Technologien (generative KI, LLMs, ChatGPT) erläutert. Anschließend werden zentrale **Anwendungsbereiche** aufgezeigt sowie gängige **Werkzeuge und Plattformen** vorgestellt. Ein Schwerpunkt liegt auf den **Vorteilen** dieser neuen Methoden, aber auch auf den **Herausforderungen** und Limitierungen, die mit ihrem Einsatz einhergehen. Abschließend werfen wir einen Blick auf **Zukunftsperspektiven** und Entwicklungsrichtungen. Alle Aussagen stützen sich auf aktuelle wissenschaftliche Literatur der letzten Jahre, um einen fundierten und aktuellen Einblick in diese Thematik zu ermöglichen.

**Hintergrund**

Die Unterstützung der Literaturrecherche durch KI ist kein völlig neues Konzept – bereits seit einigen Jahren existieren spezialisierte Suchmaschinen und Algorithmen, die über die klassische Stichwortsuche hinausgehen. So nutzt z. B. **Semantic Scholar** (ein 2015 gestartetes Projekt des Allen Institute for AI) KI-Methoden, um relevante Publikationen zu finden und sogar **automatische Zusammenfassungen** bereitzustellen (sogenannte *TL;DR*-Kurzfassungen) ([TLDR Feature - Semantic Scholar](https://www.semanticscholar.org/product/tldr#:~:text=The%20TLDR%20feature%20from%20Semantic,on%20the%20search%20results%20page)). Dennoch hat erst der jüngste Durchbruch bei *Large Language Models* eine breite Diskussion in Gang gesetzt. Seit der Einführung von **ChatGPT** und ähnlichen Modellen stellt sich in der wissenschaftlichen Gemeinschaft verstärkt die Frage, inwieweit solche generativen KI-Modelle für Recherchezwecke nutzbar sind ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Bing%20and%20Google%E2%80%99s%20Gemini%20to,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)). Erste Erfahrungen waren zwiespältig: Einerseits können LLMs erstaunlich flüssig Texte generieren und Fragen beantworten, andererseits wurde früh deutlich, dass eine *direkte* Literatursuche über Chatbots (noch) nicht zuverlässig funktioniert ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Large%20Language%20Models%20are%20usable,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)). Beispielsweise zeigte sich, dass ChatGPT zwar auf Aufforderung Literaturangaben liefern kann, diese aber teils fehlerhaft oder frei erfunden (*halluziniert*) sind ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=match%20at%20L713%20entirely%20fabricated,2%2C15)). Ein prominentes Beispiel ist **Meta’s Galactica**, ein KI-Modell, das 2022 speziell für wissenschaftliche Inhalte entwickelt wurde – es musste nach wenigen Tagen vom Netz genommen werden, weil es *plausibel klingenden Unsinn* und falsche Quellenangaben produzierte ([Meta Abruptly Pulls Galactica 'Science' Model After Launch](https://www.cdotrends.com/story/17674/meta-abruptly-pulls-galactica-%E2%80%98science%E2%80%99-model-after-launch#:~:text=Meta%20Abruptly%20Pulls%20Galactica%20%27Science%27,of%20generating%20alarmingly%20plausible%20nonsense)). Diese Ereignisse machten deutlich, dass zwar enormes Potenzial besteht, aber auch Risiken, wenn KI ungeprüft für die Literaturrecherche eingesetzt wird.

Trotz solcher Anfangsschwierigkeiten hat die rasante Entwicklung der KI dazu geführt, dass immer mehr **wissenschaftliche Studien** die Rolle von KI in der Literaturrecherche untersuchen. In den letzten drei Jahren erschien eine Reihe von Arbeiten, die den Einsatz von *generativer KI* und *LLMs* in verschiedenen Phasen des Recherche- und Review-Prozesses evaluieren. Eine Scoping-Review aus dem Jahr 2025 identifizierte 37 Studien seit 2021, in denen LLMs zur Unterstützung von systematischen Übersichtsarbeiten (*Systematic Reviews*) eingesetzt wurden ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Our%20database%20search%20yielded%208054,fifth%20as%20nonpromising%20%28n%C2%A0%3D%C2%A08%2C%2022)). Diese Studien decken nahezu alle Schritte des Review-Prozesses ab – von der Literatursuche über die Studienauswahl bis zur Extraktion von Daten – und zeigen damit, dass KI-Unterstützung in der Breite der Recherche zunehmend an Relevanz gewinnt ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Our%20database%20search%20yielded%208054,fifth%20as%20nonpromising%20%28n%C2%A0%3D%C2%A08%2C%2022)). Insgesamt zeichnen die bisherigen Arbeiten ein Bild, in dem KI zwar bereits **vielversprechende Ergebnisse** liefert, aber noch keine ausgereifte Lösung darstellt ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Conclusion)) ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=studies%20%2838,important%20in%20creating%20systematic%20reviews)). Dieses Spannungsfeld zwischen Potenzial und Unreife wird im Folgenden näher beleuchtet.

**Technologien: Generative KI, LLMs und ChatGPT**

Moderne KI-Tools für die Literaturrecherche basieren vor allem auf **generativer KI** und großen Sprachmodellen. Generative KI bezeichnet Verfahren des *maschinellen Lernens*, die neue Inhalte (Text, Bilder usw.) auf Grundlage von gelernten Mustern erzeugen können ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Im%20Zentrum%20der%20KI,basierend%20auf%20zuvor%20gelernten%20Mustern)). Ein *Large Language Model (LLM)* wie **GPT-3** oder **GPT-4** ist ein typisches Beispiel: Das Modell wird auf riesigen Textmengen trainiert und lernt dabei, statistische Zusammenhänge zwischen Wörtern und Sätzen zu erkennen. Dadurch ist es in der Lage, auf Eingaben in natürlicher Sprache mit kohärenten Texten zu antworten. **ChatGPT** von OpenAI ist eine benutzerfreundliche Schnittstelle zu solchen LLMs, die im Dialogformat Antworten liefert ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Artificial%20intelligence%20,1)). ChatGPT und vergleichbare Modelle (z. B. **Bing Chat** mit GPT-4 oder Googles **Bard**/Gemini) zählen zur *generativen KI*, da sie auf Basis ihres Trainingswissens neu formulierte Antworten generieren.

Wichtig für die Recherche ist, dass LLMs nicht wie klassische Suchmaschinen arbeiten, die Dokumente indizieren und bei einer Anfrage passende Ergebnisse **auffinden**. Stattdessen **generieren** LLMs zunächst einmal Texte auf Basis statistischer Muster. Ohne weitere Steuerung kann dies dazu führen, dass zwar sprachlich plausible, aber faktisch falsche Aussagen entstehen. Um LLMs für die Literaturrecherche nutzbar zu machen, werden verschiedene Strategien eingesetzt, etwa: (1) **Prompt Engineering** – die Eingaben werden so gestaltet, dass das Modell gezielt nach Fakten sucht und seine Antworten mit Quellen belegt; (2) **Retrieval-Augmentation** – das Modell greift während der Generierung auf externe Datenbanken oder Dokumentenkorpora zu, um aktuelle und verifizierte Informationen einzubinden ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=for%20relevant%20prior%20work,hallucinations%20can%20be%20significantly%20reduced)) ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=results,the%20proposed%20retrieval%20and%20generation)). Letzteres ähnelt der Funktionsweise eines Suchassistenten: Das LLM formuliert z. B. selbst Suchanfragen, ruft relevante Texte ab und baut sie in die Antwort ein. OpenAI und Google arbeiten an spezialisierten Forschungs-KI-Systemen (z. B. *OpenAI’s “Deep Research”* und *Google’s “Gemini Deep Research”*), die genau diesen Ansatz verfolgen ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Students%20can%20enter%20a%20query%2C,model%20we%E2%80%99ve%20released%2C%E2%80%9D%20she%20says)). Diese sollen längere **Literaturanalysen** autonom durchführen können – Berichten zufolge können solche Systeme eine halbe Stunde lang selbstständig Fachartikel durchsuchen und dem Nutzer einen ausführlichen, vollständig **zitierbaren Bericht** mit Text, Abbildungen und Quellen liefern ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Students%20can%20enter%20a%20query%2C,any%20other%20model%20we%E2%80%99ve%20released%2C%E2%80%9D)). Diese neuen Technologien versprechen, die bisherigen Grenzen von LLMs – etwa der geschlossene Trainingsstand oder das Halluzinieren von Referenzen – zu überwinden, indem sie die Stärke der generativen Antwort mit echter **Informationssuche** kombinieren.

Zusammengefasst stellen generative KI und LLMs die technische Grundlage für viele aktuelle KI-Tools in der Literaturrecherche. Ihre Fähigkeit, natürliche Sprache zu verstehen und zu erzeugen, ermöglicht neuartige Formen der Interaktion: Anstatt nur Schlüsselwörter einzugeben, können Forschende komplexe Fragen stellen und erklärende Antworten erhalten. Gleichzeitig erfordert der effektive Einsatz dieser Modelle ein Verständnis ihrer Funktionsweise, um ihre Ausgaben richtig einordnen und steuern zu können. Die nächsten Abschnitte zeigen, wie diese Technologien konkret angewendet werden und welche **Werkzeuge** Forschenden bereits zur Verfügung stehen.

**Anwendungsbereiche KI-gestützter Literaturrecherche**

KI-Unterstützung kommt in verschiedenen **Phasen** der Literaturrecherche und -auswertung zum Einsatz. Im Folgenden werden einige zentrale Anwendungsbereiche beschrieben:

* **Literatursuche und Screening:** Ein naheliegender Einsatzbereich ist die eigentliche Suche nach relevanter Literatur in digitalen Datenbanken. KI-Tools können hier helfen, indem sie Suchanfragen intelligenter verarbeiten (z. B. Synonyme und verwandte Begriffe erkennen) und eine große Menge an Treffern vorselektieren. In der Medizin und Gesundheitsforschung wird erprobt, LLMs für das **Screening** von Titeln und Abstracts in systematischen Übersichtsarbeiten einzusetzen ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Our%20database%20search%20yielded%208054,fifth%20as%20nonpromising%20%28n%C2%A0%3D%C2%A08%2C%2022)). So können z. B. KI-Modelle Tausende von Abstracts vorsortieren und diejenigen identifizieren, die mit hoher Wahrscheinlichkeit relevant sind, wodurch die Anzahl der von Menschen gründlich zu lesenden Texte drastisch reduziert wird. Eine aktuelle Studie in *PNAS (2025)* untersuchte 18 verschiedene LLMs bei der Auswahl von Studien für drei systematische Reviews. Die besten Modelle erreichten eine korrekte Klassifikation von über **88 %** der Abstracts im Vergleich zur menschlichen Auswahl, was eine erhebliche Arbeitsersparnis bedeutet ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=titles%20and%20abstracts%20as%20either,in%20SI%20Appendix%20%2C%20Figs) ) ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=,In%20conclusion) ). Allerdings variierten die Ergebnisse je nach Thema und Suchstrategie, und insbesondere bei sehr spezifischen Fragestellungen (mit wenigen relevanten Studien) ließ die Genauigkeit noch zu wünschen übrig ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=,In%20conclusion) ) ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=match%20at%20L427%20titles%20and,in%20SI%20Appendix%20%2C%20Figs) ). Nichtsdestotrotz zeigt dieser Anwendungsfall, dass KI die *Screening*-Phase deutlich beschleunigen kann – eine wichtige Entwicklung, da gerade in systematischen Reviews das Aussortieren irrelevanter Treffer extrem zeitaufwändig ist.
* **Literaturübersichten und -synthesen:** Neben der reinen Suche kann KI auch bei der **Synthese** von Literatur helfen. Das Verfassen von *Literature Reviews* (Literaturüberblicken) gehört zu den Kernaufgaben in vielen Forschungsarbeiten und erfordert, vorhandene Studien zusammenzufassen und einzuordnen. Hier experimentieren Forschende mit LLMs, um etwa automatisiert den **Forschungsstand zusammenzufassen** oder Entwürfe von Literaturüberblicken verfassen zu lassen ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=Literature%20reviews%20are%20an%20essential,papers%20by%20querying%20an%20external)) ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=show%20promising%20potential%20for%20writing,com)). Ein Ansatz ist, einem KI-Modell eine bereits gefundene Menge von relevanten Abstracts oder Papern zu geben und es anzuweisen, daraus eine strukturierte Zusammenfassung zu erstellen. Agarwal et al. (2024) zeigten, dass GPT-4 in der Lage ist, einen groben ersten Entwurf eines Literature Reviews zu erzeugen, wenn man den Prozess in Teilschritte zerlegt ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=which%20we%20will%20make%20available,hallucinations%20can%20be%20substantially%20reduced)). Zunächst extrahiert das Modell dabei aus einem Start-Papier (*Query*-Papier) wichtige Schlüsselwörter, dann ruft es auf Basis dieser Begriffe und semantischer Ähnlichkeit weitere relevante Publikationen ab (über externe Datenbanken), und schließlich verfasst es einen zusammenhängenden Text, der diese Quellen einbezieht ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=review%20based%20on%20the%20retrieved,based%20literature%20review)) ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=show%20promising%20potential%20for%20writing,com)). Dieser mehrstufige Workflow führte zu deutlich besseren Ergebnissen als eine einfache direkte Generierung: Die Kombination aus *schlüsselwortbasierter* und *embeddings-basierter* Suche steigerte die Trefferqualität (Präzision und Recall) um 10–30 % ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=show%20promising%20potential%20for%20writing,com)). Zudem konnte durch planvolles Vorgehen die Anzahl halluzinierter Referenzen um ca. 18–26 % reduziert werden ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=find%20that%20combining%20keyword,com)). Dies verdeutlicht, dass KI-Assistenten beim Schreiben von Literaturübersichten nützlich sein können – allerdings eher als Werkzeuge zur Entlastung (etwa beim Entwurf oder der Struktur), während die inhaltliche Feinabstimmung und Bewertung weiterhin von Menschen übernommen werden muss. Tatsächlich kam die Studie zu dem Schluss, dass trotz aller Fortschritte die vollständige Automatisierung von Literature Reviews noch nicht erreicht ist („*Are we there yet?*“ – die Antwort war **nein**, noch nicht ganz) ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=results,the%20proposed%20retrieval%20and%20generation)).
* **Wissensvernetzung und Exploration:** Ein weiterer Anwendungsbereich ist die **Entdeckung neuer Zusammenhänge** in der Literatur. KI-gestützte Tools können Publikationen inhaltlich miteinander verknüpfen und so Forschenden helfen, thematische Netzwerke zu erkennen. Beispielsweise gibt es **Mapping-Tools**, die anhand von Zitations- und Inhaltsmustern visuelle Netzwerke erstellen: Arbeiten mit ähnlichen Themen werden gruppiert, häufig zitierte Schlüsselstudien hervorgehoben, etc. KI kann diese Verknüpfungen noch erweitern, indem auch inhaltliche Ähnlichkeiten (z. B. mittels Embeddings) einbezogen werden ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=,basierend%20auf%20vorhandenen%20Quellen%20identifizieren)). Ein Forschender könnte so ein zentrales Paper eingeben und das Tool schlägt andere Veröffentlichungen vor, die *konzeptionell* verwandt sind, selbst wenn keine direkten Zitate vorliegen. Dies erleichtert das Auffinden von **interdisziplinären Verbindungen** oder von relevanter Literatur, die über klassische Schlagwortsuche leicht übersehen würde. Zudem können KI-Assistenten bei der **Ideenfindung** helfen – etwa indem sie auf Basis einiger Stichworte oder eines groben Forschungsvorhabens relevante Unterthemen und Literaturhinweise generieren. Solche assistiven Funktionen wurden von Bibliotheken kategorisiert: man unterscheidet z.B. KI-*Assistenten* für Themenfindung, *Analyse*-Tools für Textzusammenfassungen, *Such*-Tools für die Literaturrecherche an sich und *Mapping*-Tools für das Aufzeigen von Vernetzungen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Die%20Auswahl%20des%20richtigen%20KI,je%20nach%20Einsatzzweck%20unterschiedlich%20kategorisieren)) ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=,basierend%20auf%20vorhandenen%20Quellen%20identifizieren)). In all diesen Bereichen gibt es bereits Prototypen und erste Produkte, die teils beachtliche Resultate liefern.
* **Datenextraktion und Evidenzsynthese:** Über die Literaturrecherche hinaus setzt KI auch im nächsten Schritt an – dem **Extrahieren von Informationen** aus gefundenen Studien. Beispielsweise experimentieren einige Arbeiten damit, LLMs Tabellen aus Artikeln oder statistische Kennzahlen entnehmen zu lassen, um Meta-Analysen zu erleichtern. Hierbei handelt es sich jedoch um sehr anspruchsvolle Anwendungen, da die Modelle sehr präzise sein müssen. In der erwähnten Scoping-Review waren ca. 30 % der identifizierten Studien dem Bereich *Datenextraktion* zuzuordnen ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Our%20database%20search%20yielded%208054,fifth%20as%20nonpromising%20%28n%C2%A0%3D%C2%A08%2C%2022)). Ein Beispiel ist die automatische Erkennung von Studienparametern (Population, Intervention, Outcomes) in klinischen Studienberichten mithilfe von NLP. Erste Ergebnisse sind vielversprechend, aber dieser Bereich steht noch am Anfang, da die Anforderungen an Genauigkeit und Kontextverständnis hier besonders hoch sind.

Zusammenfassend durchdringt KI den Prozess der Literaturrecherche von der **ersten Suche** bis zur **Zusammenfassung und Analyse** gefundener Quellen. Die größten sofortigen Effekte werden in der Beschleunigung routinemäßiger Aufgaben gesehen: Große Trefferlisten können schneller gefiltert, bekannte Informationen automatisiert zusammengefasst und Wissensgebiete zügig kartiert werden ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Forschende%20von%20zeitaufw%C3%A4ndigen%20Routinet%C3%A4tigkeiten%20entlastet)) ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=KI%20revolutioniert%20die%20Literaturrecherche%2C%20indem,Menschliche)). Dies verschafft Forschenden mehr Zeit für wertschöpfende Tätigkeiten wie Interpretation und kritische Diskussion der Literatur. Gleichzeitig bleiben für jeden dieser Anwendungsbereiche sorgfältige Validierung und oft manuelle Nachkontrolle nötig – ein Thema, das im Abschnitt *Vorteile und Herausforderungen* detaillierter diskutiert wird.

**Werkzeuge und Plattformen**

In den vergangenen Jahren sind zahlreiche **Tools** und **Plattformen** entstanden, die KI für die Literaturrecherche nutzbar machen. Im Folgenden eine Auswahl wichtiger Werkzeuge sowie deren Kernfunktionen:

* **Semantic Scholar**: Eine kostenlose akademische Suchmaschine, die KI nutzt, um die Suche nach Fachliteratur zu verbessern. Semantic Scholar liefert nicht nur klassische Suchergebnisse, sondern auch KI-generierte **Kurzfassungen (TLDR)** für viele Paper ([TLDR Feature - Semantic Scholar](https://www.semanticscholar.org/product/tldr#:~:text=The%20TLDR%20feature%20from%20Semantic,on%20the%20search%20results%20page)). Diese Ein-Satz-Zusammenfassungen helfen, rasch einzuschätzen, ob ein Treffer relevant ist. Zudem bietet Semantic Scholar Filterfunktionen (z. B. nach Themen, Autoren) und einen semantischen Suchmodus, der verwandte Arbeiten auch bei unterschiedlichen Begrifflichkeiten findet. Das Projekt ist seit 2015 online und wird kontinuierlich um KI-Funktionen erweitert. Semantic Scholar steht exemplarisch für *AI-Powered Search Tools*, die das Auffinden von Literatur durch intelligente Algorithmen unterstützen.
* **Elicit**: *Elicit* ist ein KI-gestützter Rechercheassistent, entwickelt vom Forschungsprojekt Ought. Die Plattform durchsucht eine große Paper-Datenbank (über 125 Millionen Artikel aus der Semantic Scholar Corpus) und kombiniert **semantische Suche** mit **automatischer Zusammenfassung** ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=1,%E2%80%9D)) ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=3,whether%20the%20paper%20is%20relevant)). Nutzer können in natürlicher Sprache eine Forschungsfrage eingeben (z. B. *„Was sind die Hauptfaktoren für X laut aktueller Forschung?“*). Elicit sucht dann nach relevanten Publikationen – auch wenn diese nicht exakt die gleichen Keywords enthalten – indem es inhaltliche Ähnlichkeit nutzt ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=1,%E2%80%9D)). Für die Top-Ergebnisse generiert Elicit kurze **abstract-spezifische Zusammenfassungen**, die direkt die gestellte Frage adressieren ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=3,whether%20the%20paper%20is%20relevant)). Dies erspart es, jedes Abstract selbst lesen zu müssen, und gibt einen fokussierten Überblick. Zudem ermöglicht Elicit es, Ergebnisse nach verschiedenen Kriterien zu sortieren und zu filtern, z. B. nach Studiendesign (randomisierte Studien, Reviews etc.) oder nach Zitationszahl ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=4,with%20the%20largest%20sample%20size)) ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=5,that%20cited%20a%20specific%20paper)). Das Tool zielt darauf ab, sowohl *Breite* (durch semantische Suche auch über Begriffsgrenzen hinweg) als auch *Tiefe* (durch gezielte Detailabfrage von Papers) zu bieten. Elicit ist ein prominentes Beispiel dafür, wie LLMs (*GPT-3*-basierte Module) mit traditionellen Datenbanken kombiniert werden, um eine **Literaturrecherche in Dialogform** zu ermöglichen. Wichtig zu beachten: Elicit selbst betont, dass es Mechanismen zur Reduktion von Halluzinationen eingebaut hat (z. B. stützt es sich stark auf echte Paper-Inhalte) ([Elicit Help Center](https://support.elicit.com/en/categories/146369#:~:text=we%E2%80%99ll%20share%20some%20of%20the,and%20remain%20private%20to%20your)), dennoch sollten die angebotenen Zusammenfassungen und extrapolierten Antworten immer kritisch geprüft werden.
* **Scite**: *Scite* ist ein weiteres KI-gestütztes Werkzeug, allerdings mit einem etwas anderen Fokus. Es analysiert Zitationen und versucht qualitativ auszuwerten, *wie* eine Arbeit in anderen Papers zitiert wird. Scite klassifiziert Zitate danach, ob eine Arbeit unterstützt, widersprochen oder einfach nur erwähnt wird. Kürzlich hat Scite auch einen Chatbot-Interface („Scite Assistant“) eingeführt, das Fragen mit Bezug auf die Literatur beantwortet und dabei Belege aus Papers liefert – ähnlich einer fachbezogenen ChatGPT-Version, aber mit realen Referenzen. Ein Nature-Artikel nennt Scite – neben Elicit – einen der neuen „Forschungsassistenten“, die Experimente vorschlagen oder Hypothesen testen können ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=AI%20can%20still%20function%20as,experiments%20to%20test%20the%20theory)) ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=and%20resources%20such%20as%20Scite,experiments%20to%20test%20the%20theory)). In der Praxis wird Scite vor allem genutzt, um Evidenz aus der Literatur schnell zu beurteilen: Gibt es kontroverse Befunde zu einer bestimmten These? Welche Paper stützen eine Aussage? Solche Informationen kann Scite (teil-)automatisiert liefern, was insbesondere in medizinischen Feldern hilfreich ist, um den **Konsens- oder Dissensgrad** zu einem Forschungsstand abzuschätzen.
* **Connected Papers / Research Rabbit**: Diese Tools dienen primär dem **Visualisieren von Literatur-Netzwerken**. Connected Papers generiert ausgehend von einem Ausgangs-Paper eine Netzwerkgraphik von thematisch verwandten Arbeiten. Obwohl das zugrundeliegende Verfahren nicht im Detail offenliegt, handelt es sich um eine Kombination aus Zitationsanalysen und semantischen Ähnlichkeiten – mithin auch eine Form von KI-gestützter Mustererkennung. *Research Rabbit* verfolgt ein ähnliches Ziel und bietet zusätzlich die Möglichkeit, sog. Playlists von Papern anzulegen, aus denen es kontinuierlich Empfehlungen für neue relevante Arbeiten erstellt (eine Art *personalisiertes Empfehlungssystem* für Literatur). Beide Plattformen helfen Forschenden, nichts Wichtiges zu übersehen und neue Publikationen zu entdecken, die man mit klassischen Suchen vielleicht nicht gefunden hätte. Diese **Mapping-Werkzeuge** entsprechen dem in der Theorie beschriebenen Kategorie "*Mapping Tools*" für Literaturvernetzung ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=,basierend%20auf%20vorhandenen%20Quellen%20identifizieren)). Sie sind besonders nützlich in frühen Phasen der Recherche, wenn man sich in ein neues Thema einarbeitet und zunächst das Forschungsfeld erkundet.
* **ChatGPT, Bing Chat und ähnliche KI-Chatbots**: Schließlich darf man auch die generischen KI-Plattformen nicht außer Acht lassen, denn viele Forschende nutzen sie ganz pragmatisch für Recherchezwecke. Zwar ist **ChatGPT** kein spezialisiertes Literaturrecherche-Tool, aber durch geschicktes *Prompting* kann man es dennoch einsetzen, um z. B. Zusammenfassungen von bekannten Konzepten zu erhalten oder sich wichtige Publikationen zu einem Thema *nennen* zu lassen. Allerdings muss man hier sehr vorsichtig sein: Da ChatGPT (ohne Plug-ins oder Browsing) kein direktes Wissen über aktuelle Paper hat, sondern nur sein Trainingswissen (Stand 2021 für GPT-4 in der Basisversion) nutzt, kommt es häufig vor, dass es zwar echte Journals und plausible Titel nennt, die Referenzen aber **fingiert** sind ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=match%20at%20L713%20entirely%20fabricated,2%2C15)). Für eine ernsthafte Recherche ist das direkte Vertrauen auf ChatGPT-Ausgaben daher nicht empfehlenswert ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Large%20Language%20Models%20are%20usable,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)). Anders sieht es mit neueren Varianten aus: **Bing Chat** (Microsofts Suchchatbot auf Basis von GPT-4) greift live auf das Web zu und zeigt zu seinen Antworten Quellen an. Damit kann man bereits deutlich besser arbeiten, da jeder Faktenaussage ein Link zu einer echten Webseite (etwa einem Journal-Artikel oder einer Datenbank) beigefügt ist. Ähnliches gilt für **Google Bard** mit den neuesten Updates. Diese allgemeinen KI-Chatbots entwickeln sich also schrittweise in Richtung Forschungsassistenten. Ein Experiment von Quaasdorf (2024) verglich ChatGPT, Bing (GPT-4) und einen Vorläufer von Google Gemini bei einfachen Recherchefragen. Das Fazit war, dass **keines** der Modelle eine verlässliche vollständige Literaturliste liefern konnte – Bing schnitt noch am besten ab, weil es echte Websuche einbezog, während ChatGPT und Co. zwar schnell Antworten gaben, aber Fakten und Quellen nicht konsequent korrekt waren ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Bing%20and%20Google%E2%80%99s%20Gemini%20to,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)) ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Large%20Language%20Models%20are%20usable,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)). Dennoch werden diese Tools zunehmend in den Forschungskontext integriert, vor allem für informelle Vorab-Recherchen oder zum Erhalten eines groben Überblicks. Zukünftig könnten Weiterentwicklungen wie *OpenAI Deep Research* speziell für akademische Zwecke optimierte Chatbot-Systeme darstellen ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=%E2%80%9Cadvanced%20enormously%E2%80%9D%20in%20an%20area,companies%20are%20launching%20similar%20products)) ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Students%20can%20enter%20a%20query%2C,model%20we%E2%80%99ve%20released%2C%E2%80%9D%20she%20says)).

Neben den genannten gibt es noch weitere nennenswerte Werkzeuge, z.B. **Consensus** (ein KI-Suchdienst, der evidenzbasierte Antworten aus Papers generieren will) oder **SciSpace** (eine Plattform, die Paper durchsucht und in einem Chat Fragen zu PDF-Inhalten beantwortet). Auch klassische Literaturdatenbanken wie **PubMed** experimentieren mit KI-Features – etwa der *PubMed GPT* Pilot, der kurze abstrakt-ähnliche Antworten generiert, oder automatische Klassifizierungen nach Studienqualität. Viele dieser Tools sind derzeit in Entwicklung oder Beta-Stadium. Es ist typisch, dass sie **kombinierte Funktionen** anbieten: Suche, Empfehlung, Zusammenfassung, Zitatanalyse – die Grenzen zwischen den Kategorien verwischen. Forschende haben heute die Möglichkeit, ganze *Pipelines* aus verschiedenen KI-Werkzeugen aufzubauen: So berichtet ein Doktorand in *Nature*, dass er eine Reihe von KI-Plattformen hintereinander schaltet, die ihn tagesaktuell über neue Veröffentlichungen informieren, komplexe Themen vereinfachen, beim Schreiben helfen und sogar seine Quellen organisieren ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=He%20quickly%20came%20around%2C%20however%2C,demands%20of%20classes%20and%20research)) ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Here%2C%20Nature%20explores%20how%20academics,parts%20of%20the%20research%20process)). Dieser ganzheitliche Ansatz zeigt, dass KI zunehmend in alle Facetten des wissenschaftlichen Arbeitens vordringt, mit der Literaturrecherche als einem der ersten und naheliegendsten Einsatzfelder.

**Vorteile und Herausforderungen**

Der Einsatz von KI in der Literaturrecherche bietet **erhebliche Vorteile**, bringt aber auch **Herausforderungen** mit sich. Im Folgenden werden die wichtigsten Pro- und Contra-Punkte – gestützt auf aktuelle Studien – dargestellt.

**Vorteile / Chancen:**

* **Geschwindigkeit und Effizienz**: Einer der größten Pluspunkte ist die enorme Geschwindigkeit, mit der KI-Tools arbeiten. Ein LLM wie GPT-4 kann in Sekundenbruchteilen auf eine Anfrage reagieren und (unter Zuhilfenahme von Datenbanken) innerhalb weniger Minuten eine Liste potenziell relevanter Artikel generieren ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Results%3A%20GPT,of%20knowledge%2C%20and%20contextual%20understanding)). Ein Vergleich von ChatGPT-4 mit menschlichen Forschenden ergab, dass das KI-Modell innerhalb von Sekunden umfangreiche Listen von Faktoren und Referenzen liefern konnte, während die menschliche Literaturrecherche sich über **mehrere Monate** erstreckte ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=match%20at%20L582%20The%20AI,Had%20the%20review%20been)). Natürlich war die menschliche Aufarbeitung gründlicher, doch für einen **ersten Überblick** liegt hier ein klarer Vorteil der KI: Sie kann in kurzer Zeit das Terrain abstecken und somit den Einstieg in ein neues Thema erleichtern. Durch diese Zeitersparnis werden Forschende von Routineaufgaben entlastet und können sich auf höherwertige Tätigkeiten konzentrieren ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Forschende%20von%20zeitaufw%C3%A4ndigen%20Routinet%C3%A4tigkeiten%20entlastet)). Das Prinzip des *„cognitive offloading“* – also das Auslagern kognitiver Routinearbeit an eine Maschine – ermöglicht es, die menschliche Aufmerksamkeit gezielter für Analyse und kreatives Denken einzusetzen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,und%20analytische%20Aufgaben%20konzentrieren%20%E2%80%93)).
* **Breite Abdeckung & Vernetzung**: KI-Systeme, insbesondere solche mit Zugriff auf große Literaturdatenbanken, können eine **breitere Streuung** an Ergebnissen liefern als manuell erstellte Suchanfragen. Wo ein Forschender vielleicht 2–3 Schlüsselbegriffe nutzt, kann ein KI-Tool dank semantischer Suche auch relevante Arbeiten finden, die andere Begriffe verwenden ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=1,%E2%80%9D)) ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=even%20if%20they%20don%E2%80%99t%20use,%E2%80%9D)). Dadurch sinkt das Risiko, wichtige Publikationen zu übersehen, die nur aufgrund von Terminologie nicht auf den ersten Blick erkennbar waren. Zudem sind KI-Tools gut darin, **thematische Verbindungen** aufzuzeigen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Informationen%2C%20sondern%20verbessert%20auch%20die,dem%20Prinzip%20des%20%E2%80%9Ecognitive%20offloading%E2%80%9C)). Sie können z.B. automatisch erkennen, wenn zwei Forschungsstränge inhaltlich überlappen, und den Nutzer auf solche Zusammenhänge hinweisen. Diese Vernetzung verbessert nicht nur die Rechercheergebnisse, sondern fördert auch interdisziplinäre Sichtweisen. KI kann etwa Vorschläge machen wie: "Paper X aus Gebiet A könnte relevant für Ihre Fragestellung in Gebiet B sein, obwohl das auf den ersten Blick nicht offensichtlich ist." Insgesamt erhöht sich dadurch die **Qualität der Recherche**, da ein umfassenderes Bild des Forschungsfeldes entsteht ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Forschende%20von%20zeitaufw%C3%A4ndigen%20Routinet%C3%A4tigkeiten%20entlastet)).
* **Automatische Zusammenfassungen und Inhaltsanalyse**: Ein weiterer Vorteil ist die Fähigkeit moderner KI, Texte inhaltlich zu analysieren und zusammenzufassen. KI-gestützte Tools können für gefundene Artikel **automatisch Abstracts oder Kernaussagen generieren** ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=3,whether%20the%20paper%20is%20relevant)). Dies hilft enorm beim Screening: Anstatt jedes Paper vollständig lesen zu müssen, bekommt man eine KI-Zusammenfassung, die speziell auf die gestellte Frage zugeschnitten sein kann. Beispielsweise kann ein Tool wie Elicit zu jedem Suchtreffer einen Satz liefern, der erklärt, warum dieses Paper für die Anfrage relevant ist ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=3,whether%20the%20paper%20is%20relevant)). Auch detailliertere Zusammenfassungen oder Extraktionen (z. B. Herausfiltern von Ergebnissen oder Methoden) sind möglich. In Zukunft könnten solche Funktionen die Grundlage für **teilautomatisierte systematische Reviews** bilden, in denen KI Vorarbeit leistet, die der Mensch dann validiert und verfeinert. Schon jetzt berichten Wissenschaftler, dass durch KI-Hilfen die Orientierung in komplexer Literatur deutlich erleichtert wird ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Durch%20den%20Einsatz%20dieser)). Komplizierte oder sehr lange Texte lassen sich von KI erklären oder in Stichpunkte zerlegen, was gerade für Nachwuchsforschende eine wertvolle Lernhilfe sein kann. Letztlich führen diese Möglichkeiten dazu, dass der **Forschungsprozess beschleunigt** und effizienter gestaltet wird ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Informationen%2C%20sondern%20verbessert%20auch%20die,dem%20Prinzip%20des%20%E2%80%9Ecognitive%20offloading%E2%80%9C)), ohne (im Idealfall) dass die inhaltliche Gründlichkeit leidet.
* **Neue Einsichten und Hypothesen**: KI kann auch **kreative Impulse** geben, indem sie Muster erkennt, die dem Menschen vielleicht entgehen. Beispielsweise könnte eine KI bei der Analyse von 100 Abstracts feststellen, dass ein bestimmtes Methodenmuster immer mit einem bestimmten Ergebnis korreliert – ein Zusammenhang, der verstreut in der Literatur existiert, aber noch nicht ausdrücklich thematisiert wurde. Solche Entdeckungen können zu neuen Forschungsfragen anregen. Einige Forscher sehen hierin eine der spannendsten Chancen von KI in der Wissenschaft: nicht nur das Automatisieren, was wir ohnehin tun, sondern das Aufdecken *verborgener Strukturen*. Zwar steckt dieser Bereich noch in den Kinderschuhen, aber erste Fälle wurden dokumentiert, etwa KI-Vorschläge für experimentelle Designs oder die Identifikation von Wissenslücken, auf die Menschen nicht sofort gestoßen wären ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=AI%20can%20still%20function%20as,experiments%20to%20test%20the%20theory)) ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=and%20resources%20such%20as%20Scite,experiments%20to%20test%20the%20theory)). Das Konzept der *Augmented Intelligence*, also der gezielten Verstärkung menschlicher Intelligenz durch KI, zielt genau darauf ab – die KI bringt Vorschläge und Verbindungen ein, der Mensch bewertet und nutzt diese kreativ weiter ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=basierend%20auf%20dem%20Prinzip%20des,%E2%80%9Ecognitive%20offloading%E2%80%9C)).

Zusammengefasst zeigen sich erhebliche Vorteile durch KI-Unterstützung: eine schnellere, umfassendere und manchmal auch intelligentere Literaturrecherche. Viele Routinearbeiten können beschleunigt werden, und in manchen Fällen eröffnet KI ganz neue Perspektiven.

**Herausforderungen / Risiken:**

* **Qualität und Verlässlichkeit der Ergebnisse**: Die perhaps größte Herausforderung ist die Frage der **Zuverlässigkeit**. Aktuelle generative Modelle neigen dazu, sogenannte *Halluzinationen* zu produzieren – also Antworten, die zwar selbstbewusst präsentiert werden, aber nicht der Realität entsprechen. In der Literaturrecherche ist dies besonders heikel, etwa wenn ein KI-Assistent Quellenangaben liefert, die es in Wirklichkeit gar nicht gibt ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=match%20at%20L713%20entirely%20fabricated,2%2C15)). Mehrere Untersuchungen haben festgestellt, dass ChatGPT und ähnliche LLMs Referenzen erfinden können, wenn sie dazu aufgefordert werden, Literatur zu nennen ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=match%20at%20L713%20entirely%20fabricated,2%2C15)). Ein Beispiel: Das Modell wird gefragt nach wichtigen Artikeln zu einem Thema – es nennt Titel, Autor und Journal, die sehr glaubwürdig klingen, doch eine Überprüfung zeigt, dass keine solchen Artikel existieren. Solche Fehler können gravierende Folgen haben, wenn sie unbemerkt bleiben. Auch abseits von Referenzen ist die **Inhaltsgenauigkeit** ein Problem: GPT-4 etwa konnte in einem Test zwar eine beeindruckend umfangreiche Liste von Faktoren zu einem medizinischen Thema liefern, enthielt aber auch **irrelevante oder falsche Informationen** darunter ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Results%3A%20GPT,of%20knowledge%2C%20and%20contextual%20understanding)). Die KI hat nicht das tiefe Kontextverständnis, um Fakten zu gewichten oder Wichtiges von Unwichtigem immer sicher zu trennen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=KI%20revolutioniert%20die%20Literaturrecherche%2C%20indem,Menschliche)). Daher besteht die Gefahr, dass eine KI-Suche zwar schnell viele Ergebnisse liefert, die *Qualität* dieser Ergebnisse aber gemischt ist – es können falsche Treffer oder Fehlschlüsse dabei sein. Ohne menschliche Überprüfung besteht also ein hohes Risiko, sich auf fehlerhafte Grundlagen zu stützen. Einhellig warnen die Studien: **Menschliche Expertise bleibt unverzichtbar**, um KI-Ergebnisse einzuordnen und zu validieren ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Verbindungen%20aufzeigt.%20KI,Menschliche%20Expertise%20bleibt%20daher%20unverzichtbar)).
* **Mangelndes Kontextverständnis und fehlende Urteilsfähigkeit**: KI-Systeme, so leistungsfähig sie in Mustererkennung sind, fehlt es an echtem **Verständnis**. Sie können Sprache verarbeiten, aber sie *begreifen* nicht in menschlichem Sinne. Das zeigt sich z.B. darin, dass eine KI keine zuverlässige **qualitative Bewertung** einer Studie durchführen kann ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=Verbindungen%20aufzeigt.%20KI,Menschliche%20Expertise%20bleibt%20daher%20unverzichtbar)). Ob ein Paper methodisch solide ist, ob die Schlussfolgerungen der Autoren berechtigt sind oder ob widersprüchliche Befunde vorliegen – all das sind Bewertungen, die Expertenwissen und Kontextkenntnis erfordern. LLMs können solche Bewertungen höchstens nachahmen, indem sie gelernte Textbausteine reproduzieren. Somit stößt KI an Grenzen, wenn es um die **tiefergehende Analyse** geht: Sie kann etwa fünf Artikel zum gleichen Thema zusammenfassen, aber nicht mit letzter Sicherheit entscheiden, welche Studie die robustesten Evidenzen liefert. Ebenso können feine **Nuancen** in Texten (Sarkasmus, vorsichtige Formulierungen, Einschränkungen) von KI übersehen oder falsch interpretiert werden. Hier lauert die Gefahr, dass KI-generierte Zusammenfassungen *übervereinfachen* oder wichtigen Kontext auslassen. Ein praktisches Beispiel: In einer medizinischen Übersichtsarbeit mag die KI übersehen, dass alle identifizierten Studien z.B. nur an Männern durchgeführt wurden – etwas, das ein menschlicher Reviewer als wichtigen Hinweis (Limitation) aufnehmen würde. Solche kontextuellen Aspekte kann KI (noch) nicht zuverlässig berücksichtigen ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=liefern%2C%20was%20den%20Forschungsprozess%20erheblich,Menschliche)). Damit einher geht die fehlende **Urteilsfähigkeit** im ethischen oder analytischen Sinn: Eine KI hat keine Intuition, was eine "bedeutende Entdeckung" ist oder welche Frage man besser nicht mit statistischen Mitteln allein beantwortet. Die *Entscheidungshoheit* muss daher beim Forschenden bleiben.
* **Bias und Datengrundlage**: KI-Modelle sind so gut wie die Daten, auf denen sie trainiert wurden. Wenn es in diesen Daten systematische **Verzerrungen (Bias)** gibt, reproduziert die KI diese unter Umständen. In der wissenschaftlichen Literatur gibt es beispielsweise sprachliche oder regionale Bias (mehr englischsprachige Publikationen, mehr Studien aus bestimmten Ländern oder von bestimmten Autoren in den Trainingsdaten). Ein KI-Tool könnte dadurch z.B. die Tendenz haben, westliche Quellen zu bevorzugen oder Mainstream-Meinungen zu verstärken, während abweichende oder neuere Ansätze untergehen. Auch das Problem des **publication bias** in der Forschung (dass positive Ergebnisse häufiger publiziert werden als negative) könnte von einer KI unreflektiert übernommen werden und dann etwa vorschlagen: "Die Mehrheit der Literatur sagt X wirkt, also wirkt X." – ohne zu erkennen, dass vielleicht viele Gegenstudien nie publiziert wurden. Zwar sind sich Entwickler dieser Probleme bewusst und versuchen, Modelle entsprechend zu kalibrieren, jedoch bleibt es eine Herausforderung, KI-Ausgaben auf **Unvoreingenommenheit** und Vollständigkeit zu trimmen. Nutzer von KI in der Literaturrecherche müssen sich also der möglichen blinden Flecken bewusst sein und im Zweifelsfall eigene Suchstrategien ergänzend einsetzen, um ein ausgewogenes Bild zu erhalten.
* **Transparenz und Nachvollziehbarkeit**: Ein praktisches Problem bei komplexen KI-Systemen ist die oft fehlende **Transparenz**. Wenn ein KI-Tool 20 Paper vorschlägt, ist nicht immer klar, *warum* genau diese ausgewählt wurden. Klassische Datenbanken erlauben zumindest eine Reproduzierbarkeit: Gleiche Suchanfrage, gleiche Filter -> gleiches Resultat. Bei lernenden Systemen können aber kleine Änderungen im Prompt andere Ergebnisse liefern ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=references,We)), und die internen Entscheidungswege sind nicht offen gelegt. Dies kann das Vertrauen der Nutzenden beeinträchtigen und macht auch wissenschaftliche Vergleiche schwierig. Wenn zwei Forschende mit dem gleichen KI-Assistenten arbeiten und unterschiedliche Outputs bekommen, weil das Modell stochastisch oder mit Undokumentierten externen Daten operiert, stellt sich die Frage der *Reliabilität*. Erste Arbeiten fordern daher Frameworks für eine bessere **Nachvollziehbarkeit** der KI-Ausgaben in Reviews ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=generation%20results,the%20proposed%20retrieval%20and%20generation)). Bis solche Standards etabliert sind, muss man KI-Ergebnisse immer mit einer gewissen Skepsis behandeln und idealerweise mehrfach verifizieren. Auch sollte man im Forschungsbericht offenlegen, wie KI zum Einsatz kam (z. B. welche Version, welche Parameter), um die Transparenz zu erhöhen – einige Journals verlangen inzwischen eine Angabe, ob und wie AI-Assistenz genutzt wurde.
* **Ethik und akademische Integrität**: Nicht zuletzt wirft die KI-Nutzung Fragen der **wissenschaftlichen Integrität** auf. Das Spektrum reicht von Urheberrechtsfragen (dürfen KI-Zusammenfassungen aus geschützten Papers erstellt werden?) über Datenschutz (bei Nutzung eigener Literaturverwaltung in KI-Tools) bis hin zu Bedenken, ob das Verwenden von KI bei Literaturarbeit als *Ghostwriting* anzusehen ist. Universitäten versuchen derzeit Leitlinien für den Umgang mit Tools wie ChatGPT zu formulieren. Einig ist man sich, dass KI ein Hilfsmittel sein kann, aber die **Verantwortung** für Korrektheit und Originalität der Arbeit beim Forschenden bleibt. Insbesondere das blinde Vertrauen auf KI-generierte Inhalte ohne Prüfung würde gegen die Sorgfaltspflicht verstoßen. Zudem sollte transparent bleiben, was eigenständige geistige Leistung ist und was (teil-)automatisiert erfolgte. Diese Meta-Aspekte sind zwar nicht technisch, aber für die Praxis relevant: Der Nutzen von KI darf nicht dazu führen, dass etablierte Qualitätskriterien aufgeweicht werden. Im Gegenteil, viele argumentieren, man müsse **kritischer denn je** sein, wenn KI im Spiel ist ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=liefern%2C%20was%20den%20Forschungsprozess%20erheblich,Menschliche)), um sicherzustellen, dass wissenschaftliche Standards gehalten werden.

Zusammenfassend lässt sich sagen, dass KI-Tools für die Literaturrecherche zweifellos mächtig sind und bereits jetzt greifbare Vorteile bringen. Gleichzeitig sind die Herausforderungen real: ohne sorgfältige Validierung durch Menschen kann der Schuss nach hinten losgehen. Die aktuelle Literatur betont einhellig ein hybrides Modell: KI als **Assistenz**, der Mensch als **Kontrolleur und Interpret** ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Conclusions%3A%20The%20study%20suggests%20that,and%20contextually%20rich%20scholarly%20outputs)). So formulieren es Mostafapour et al. (2024) in ihrer Vergleichsstudie treffend: GPT-4 kann mit guter Anleitung ein wertvolles Werkzeug für **vorläufige** Literaturüberblicke sein, aber es ersetzt nicht die Expertise von Forschenden – vielmehr müsse man KI **kombinieren** mit menschlicher Qualitätssicherung, um wirklich verlässliche Ergebnisse zu erzielen ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Conclusions%3A%20The%20study%20suggests%20that,with%20human%20expertise%20for%20more)). Anders ausgedrückt: die Stärken der KI (Schnelligkeit, Breite) sollten genutzt, ihre Schwächen (fehlende Tiefe, mögliche Fehler) durch menschliches Eingreifen ausgeglichen werden ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Results%3A%20GPT,of%20knowledge%2C%20and%20contextual%20understanding)) ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Conclusions%3A%20The%20study%20suggests%20that,and%20contextually%20rich%20scholarly%20outputs)).

**Zukunftsperspektiven**

Angesichts der schnellen Fortschritte im Bereich KI ist zu erwarten, dass die nächsten Jahre bedeutende Weiterentwicklungen für die literaturbezogene Forschungspraxis bringen. Abschließend seien einige *Zukunftsperspektiven* skizziert:

* **Verfeinerte LLMs und spezialisierte Modelle**: Große Anbieter wie OpenAI, Google und Meta arbeiten bereits an der nächsten Generation von Sprachmodellen, die explizit für wissenschaftliche Anwendungen optimiert sind. *GPT-4* ist bereits deutlich leistungsfähiger als seine Vorgänger in Bezug auf Faktentreue, und zukünftige Modelle (GPT-5 und darüber hinaus, oder Googles *Gemini*) könnten noch größere Kontextfenster haben, aktuelleres Wissen und bessere Zitationsfähigkeiten. Ein spannender Ansatz sind **domänenspezifische LLMs** – etwa Modelle, die ausschließlich mit wissenschaftlichen Publikationen trainiert werden (ähnlich *SciBERT* für wissenschaftliche Texte). Solche Modelle könnten ein tieferes Verständnis von Fachjargon und Methoden mitbringen. Wir dürfen z.B. eine Art "ScholarGPT" erwarten, das bei Eingabe einer Forschungsfrage automatisch nicht nur antwortet, sondern direkt auf die *relevante Fachliteratur* verweist und die Antwort mit DOI-verknüpften Quellen untermauert. Die Ergebnisse von Tools wie *OpenAI Deep Research* deuten an, wohin die Reise geht: LLMs, die eigenständig recherchieren und dabei **verlässliche Zitate** liefern ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Students%20can%20enter%20a%20query%2C,model%20we%E2%80%99ve%20released%2C%E2%80%9D%20she%20says)). Dies könnte die Qualität der KI-Ausgaben dramatisch erhöhen und das Vertrauen der Nutzer stärken.
* **Integration in wissenschaftliche Arbeitsumgebungen**: KI-Funktionen werden voraussichtlich nahtlos in die üblichen Plattformen der Wissenschaft integriert. Man kann sich vorstellen, dass **Datenbanken und Bibliothekskataloge** eingebaute KI-Assistenten haben: Etwa einen Button "Frage an den Literatur-Assistenten", der eine natürliche Frage entgegennimmt und daraus eine ausgefeilte Suchstrategie plus Zusammenfassung generiert. Erste Ansätze sieht man bereits – z.B. testet Elsevier KI-Erweiterungen für Scopus, und Tools wie *Dimensions* bieten AI-Analysen an. Auch in **Literaturverwaltungstools** (Zotero, EndNote etc.) könnten KI-Module helfen, automatisch neue relevante Publikationen vorzuschlagen oder beim Einsortieren von PDFs Schlüsselinformationen zu extrahieren (wie Studiendesign, Stichprobengröße, etc., vgl. Elicit-Features ([Why Elicit Is Different from Other Research Tools](https://support.elicit.com/en/articles/552705#:~:text=4,with%20the%20largest%20sample%20size))). Langfristig könnten KI-Assistenten zu einer Art *persönlichem Research-Co-Piloten* avancieren, der ständig im Hintergrund mitläuft: Er beobachtet, welche Papers man liest oder zitiert, und schlägt proaktiv weitere Literatur vor, warnt vor möglichen Oversights („Du hast Studie X noch nicht berücksichtigt, die könnte wichtig sein“), oder fasst auf Wunsch einen gelesenen Stapel Literatur in einem einheitlichen Bericht zusammen.
* **Besseres Handling von Halluzinationen und Transparenz**: Die Forschung der letzten Jahre hat viel Aufklärungsarbeit geleistet, wo aktuelle KI-Tools versagen (Halluzinationen, Intransparenz). Daraus ergeben sich konkrete Entwicklungsziele. Künftige Systeme werden vermutlich Mechanismen enthalten, die *jede Aussage mit Evidenz verknüpfen* – z.B. indem im generierten Text jeder faktische Satz eine Referenznummer trägt, die auf die Quelle verweist (einen solchen Ansatz verfolgt z.B. das Projekt *Galactica* in seinem Ideal, bevor es scheiterte). Ebenso wird daran gearbeitet, **Erklärbarkeit** zu verbessern: KI könnte z.B. begründen, warum sie ein Paper vorschlägt („Dieses Paper wurde von 5 Ihrer bereits ausgewählten Quellen zitiert“ oder „Das Abstract enthält folgende Schlüsselbegriffe, die zu Ihrer Frage passen…“). Durch solche Meta-Informationen könnte die *Nachvollziehbarkeit* steigen. Auch *Benutzerschulungen* im Prompting werden wohl selbstverständlicher: Da Prompt-Engineering eine große Rolle spielt ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=generation%20results,the%20proposed%20retrieval%20and%20generation)), könnten Tools sog. *Prompt-Vorlagen* oder Assistenten zur Formulierung mitliefern, um optimale Ergebnisse zu erzielen. Insgesamt ist zu erwarten, dass die Interaktion mit KI transparenter und kontrollierbarer wird, damit Nutzende sich auf die Resultate eher verlassen können.
* **Veränderte Forschungskultur und Arbeitsabläufe**: Sollten KI-Assistenten tatsächlich so leistungsfähig werden, dass sie große Teile der Literaturrecherche autonom erledigen, wird dies die wissenschaftliche Arbeitskultur verändern. Schon jetzt sprechen einige von einer *„Revolution für die Forschung“*, da viele Studierende und Wissenschaftler KI beinahe täglich nutzen ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Shafi%20now%20says%20that%20the,hard%20to%20imagine%20wanting%20to%E2%80%9D)). Zukünftig könnte es zum Standard werden, dass ein erster Literaturüberblick immer zunächst mit KI-Hilfe erstellt wird – ähnlich wie heute niemand mehr von Hand in gedruckten Indexbänden sucht, sondern Datenbanken nutzt. Der **Skillset** von Forschenden wird sich anpassen müssen: Kompetenz im Umgang mit KI-Tools (Stichwort *KI-Literacy*) wird wichtiger. Einige Bibliotheken bieten bereits Schulungen dazu an, wie man KI bei der Recherche effektiv und ethisch korrekt einsetzt. Gleichzeitig werden Peer-Review-Prozesse und Publikationsrichtlinien nachziehen, um ggf. KI-unterstützte Arbeiten einordnen zu können. Im besten Fall führt die Entlastung von Routinetätigkeiten dazu, dass mehr Zeit für Hypothesenbildung, Experimente und gründliche inhaltliche Diskussion bleibt – Bereiche also, in denen menschliche Kreativität und kritisches Denken gefragt sind. KI könnte somit die **Qualität der Forschung** insgesamt heben, wenn sie richtig eingesetzt wird, indem sie repetitives „Literaturgraben“ minimiert und Raum für *Denkarbeit* schafft.
* **Ethik & Regulierung**: Ein Blick in die Zukunft muss auch ethische und regulatorische Aspekte berücksichtigen. Es ist denkbar, dass Richtlinien entwickelt werden, wann und wie KI in wissenschaftlichen Reviews eingesetzt werden darf. Vielleicht entstehen Standards für **KI-gesteuerte Reviews**, die genau definieren, welche Schritte von KI übernommen werden dürfen und wo menschliche Validierung Pflicht ist. Ebenso könnte es Zertifizierungen für KI-Software geben, die deren Zuverlässigkeit bestätigen (ähnlich einem Gütesiegel). Auf der anderen Seite könnten Missbrauchsfälle (etwa Plagiate durch KI oder massenhaft fehlerhafte KI-Übersichtsarbeiten) zu strengeren Regeln führen. Die kommenden Jahre werden hier ein Austarieren bringen müssen zwischen Innovation und Integrität.

Insgesamt blicken viele Expertinnen und Experten **optimistisch** auf die Entwicklung: Die bisherigen Studien zeigen trotz aller Probleme eine Tendenz, dass die **Treffsicherheit und Nützlichkeit** von KI in der Literaturrecherche kontinuierlich zunimmt ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Conclusion)) ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=studies%20%2838,are%20becoming%20increasingly%20important%20in)). So lautet das Fazit einer umfassenden Übersichtsarbeit: "*Obwohl LLMs noch nicht vollständig einsatzbereit sind, unterstreicht der rasche Zuwachs an Forschungsarbeiten in diesem Bereich ihre wachsende Relevanz*" ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Conclusion)). Es ist daher plausibel, dass in naher Zukunft KI-Tools aus dem Forschungsalltag nicht mehr wegzudenken sind – als leistungsfähige Assistenten, die Wissenschaftlerinnen und Wissenschaftler bei der Bewältigung der Informationsflut unterstützen, ohne jedoch die wissenschaftliche Urteilsbildung zu ersetzen. Die **Zukunft der Literaturrecherche** dürfte in einer engen Kollaboration von Mensch und Maschine liegen: KI liefert Entwürfe, Vorschläge und Automation, der Mensch steuert Expertise, Bewertung und Kreativität bei. In diesem Sinne stehen wir am Beginn einer neuen Ära, in der literaturorientierte Forschung deutlich effizienter werden kann, solange wir die Technik mit Augenmaß und Verantwortungsbewusstsein einsetzen.

**Literaturverzeichnis**

* **Agarwal, S.**, *et al.* (2024). *LLMs for Literature Review: Are we there yet?* arXiv Preprint arXiv:2412.15249 ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=Literature%20reviews%20are%20an%20essential,papers%20by%20querying%20an%20external)) ([LLMs for Literature Review: Are we there yet?](https://arxiv.org/html/2412.15249v1#:~:text=which%20we%20will%20make%20available,hallucinations%20can%20be%20substantially%20reduced)).
* **Delgado-Chaves, F. M.**, *et al.* (2025). *Transforming literature screening: The emerging role of large language models in systematic reviews*. **Proceedings of the National Academy of Sciences**, 122(2), e2411962122 ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=,In%20conclusion) ) ( [Transforming literature screening: The emerging role of large language models in systematic reviews - PMC](https://pmc.ncbi.nlm.nih.gov/articles/pmid/39761403/#:~:text=titles%20and%20abstracts%20as%20either,in%20SI%20Appendix%20%2C%20Figs) ).
* **Heidt, A.** (2025). *AI for research: the ultimate guide to choosing the right tool*. **Nature** (Career Feature), 07 April 2025 ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=Students%20can%20enter%20a%20query%2C,model%20we%E2%80%99ve%20released%2C%E2%80%9D%20she%20says)) ([AI for research: the ultimate guide to choosing the right tool](https://www.nature.com/articles/d41586-025-01069-0#:~:text=AI%20can%20still%20function%20as,experiments%20to%20test%20the%20theory)).
* **Lieberum, J.-L.**, *et al.* (2025). *Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review*. **Journal of Clinical Epidemiology**, 151 (in press) ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=Conclusion)) ([Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review - Journal of Clinical Epidemiology](https://www.jclinepi.com/article/S0895-4356%2825%2900079-4/fulltext#:~:text=studies%20%2838,important%20in%20creating%20systematic%20reviews)).
* **Mostafapour, M.**, *et al.* (2024). *Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study*. **JMIR AI**, 3(1), e56537 ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Results%3A%20GPT,of%20knowledge%2C%20and%20contextual%20understanding)) ([JMIR AI - Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study](https://ai.jmir.org/2024/1/e56537#:~:text=Conclusions%3A%20The%20study%20suggests%20that,and%20contextually%20rich%20scholarly%20outputs)).
* **Quaasdorf, F.** (2024). *Künstliche Intelligenz in der Literaturrecherche*. **API Magazin**, 5(2), 1–14 ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Bing%20and%20Google%E2%80%99s%20Gemini%20to,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)) ([200204\_SUB\_OJS\_Artikel](https://journals.sub.uni-hamburg.de/hup3/apimagazin/article/download/198/229#:~:text=Large%20Language%20Models%20are%20usable,Intelligence%2C%20ChatGPT%2C%20Chatbot%2C%20Literature%20Research)).
* **Universität Augsburg Bibliothek** (2023). *KI in der Literaturrecherche* (Webseite) ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=F%C3%BCr%20die%20wissenschaftliche%20Literaturrecherche%20bietet,Forschende%20von%20zeitaufw%C3%A4ndigen%20Routinet%C3%A4tigkeiten%20entlastet)) ([KI in der Literaturrecherche](https://www.uni-augsburg.de/de/organisation/bibliothek/kurse-beratung/ki-in-der-literaturrecherche/#:~:text=KI%20revolutioniert%20die%20Literaturrecherche%2C%20indem,Menschliche)).