



Literaturrecherche in Zeiten der KI

von Heike Baller

Literaturrecherche gehört zu jeder wissenschaftlichen Arbeit, sei es eine Haus- oder Abschlussarbeit, und überhaupt zu jeder Art wissenschaftlichen Arbeitens. Auch in der Forschung begleitet sie Forschende ihr ganzes Berufsleben lang. Der Umfang kann variieren, die Methode „Literaturrecherche“ nicht.

Nun ist seit November 2022 Künstliche Intelligenz (KI, oder englisch AI) in aller Munde. Der Launch von ChatGPT hat einen Hype zum Thema ausgelöst. Den meisten ist schon klar, dass ChatGPT nicht zur Literaturrecherche taugt – aber in den letzten Monaten sind neue KI-Tools aufgekommen, die Unterstützung bei der Literaturrecherche versprechen. Was ist da dran?

Bevor in dieser Reihe einzelne Tools im Vordergrund stehen, soll es um ein paar Gedanken rund um die KI-gestützte Form der Literaturrecherche gehen.

Welche Quellen nutzen KI-gestützte Recherche-Tools?

Wenn Sie seriös und umfassend recherchieren wollen, müssen Sie sicher sein, möglichst breit aufgestellte Quellen zu nutzen. Dazu gehören:

- „Sammelstellen“ wichtiger Journals – von Google Scholar über OPAC, Datenbanken bis Web of Science, alles Tools, bei denen keine KI eingesetzt wird. Seit einiger Zeit gilt es nun auch, geeignete KI-gestützte Tools zu nutzen.
- Je nach Fachgebiet sollte der Erscheinungszeitraum der von diesem Tool erfassten Literatur länger sein als die letzten zehn Jahre.
- Quellen mit Literatur, die nicht nur im Open Access verfügbar ist.

Dieser letzte Punkt ist ein Thema, das je nach Fachgebiet unterschiedlich zu gewichten ist – es ist jedoch klar, dass die Forschung in vielen Bereichen in Zeitschriften oder anderen Publikationen weitergetrieben wird, die nicht im Open Access zur Verfügung stehen.

Die Quellenlage eines Recherche-Tools zu kennen, sei es ein OPAC, eine Datenbank oder eines der neuen KI-gestützten Hilfsmittel, ist also sehr wichtig. Verfügt ein Tool über eine „About“- oder FAQ-Seite, ist das oft ein guter Anlaufpunkt, um diese Informationen zu bekommen.

Es gibt KI-gestützte Recherche-Tools, bei denen nicht klar ist, welche Quellen ihnen zugrundeliegen. Es dann mit Trial and Error herauszufinden, ist mühsam und nicht wirklich erfolgversprechend. Ein paar Tests können zeigen, ob da nur dubiose Repositorien kleiner Universitäten hinterlegt sind oder verlässliche Quellen, aber das ist dann keine gesicherte Angabe zur Quellengrundlage.

Verschiedene Tools nutzen

Wenn die Quellenlage unklar ist, sollten Sie auf jeden Fall weitere Tools hinzunehmen, am besten solche, bei denen Sie sehen können, woher sie ihre Informationen beziehen.

Ein zweiter Grund, mehrere Tools zu nutzen, auch wenn Sie die Quellenlage kennen: Sie kann sehr unterschiedlich sein – ein OPAC z.B. kann nur den Bestand einer Bibliothek verzeichnen, ein anderer umfasst mehrere Bibliotheken. Eine Datenbank kann einem bestimmten Thema dienen oder – gerade bei den neuen Tools mit KI kommt das häufiger vor – die Quellen stammen vorwiegend aus dem englischsprachigen Raum. Diese Informationen werden oft auf „About“- oder FAQ-Seiten bereitgestellt. Klar ist, dass Tools, die solche Informationen nicht bereitstellen, schwierig sind – sie sind schwer einzuschätzen und deshalb ist auch ihre Nutzung Abwägungssache.

Die Empfehlung, mehrere Recherchequellen zu verwenden, gilt natürlich auch für die Angebote ohne KI-Unterstützung – denken Sie nur an den OPAC Ihrer Hochschule oder Angebote wie Google Scholar oder das DBIS zur Suche nach Datenbanken. Sie werden sich nie nur auf eines davon beschränken.

Was unterscheidet KI-gestützte Literaturrecherche von der ohne KI?

Die Unterschiede zwischen KI-gestützten Tools und solchen ohne KI lässt sich am Beispiel von Semantic Scholar ganz gut nachvollziehen. Dieses Tool ist schon seit 2015 am Start und bietet KI-gestützten Zusatznutzen.

KI-Elemente bei Semantic Scholar

Folgende Funktionen bei Semantic Scholar beruhen auf dem Einsatz von KI:

1. Summary/Zusammenfassung: Semantic Scholar fasst die wichtigsten Erkenntnisse eines Papers zusammen. In der Trefferliste oberhalb des Abstracts erscheint dann: TLDR – hinter dem Kürzel verbirgt sich der Satz „Too long, didn’t read“. Semantic Scholar erstellt hier eine Kurzzusammenfassung zum jeweiligen Paper. Laut Semantic Scholar wird dabei die Kernaussage hervorgehoben. Diese Zusammenfassung generiert eine KI anhand des Titels, des Abstracts und ggf. der Keywords.

2. Zuordnung zu Fachgebieten: Bis zu drei Fachgebiete können einem Artikel auf Grundlage von Titel und Abstract zugeordnet werden. Auch das geschieht automatisiert.

3. Citations: Wenn Sie einen Titel in der Langanzeige aufrufen, bekommen Sie nicht nur das Abstract und alle bibliographischen Angaben angezeigt, sondern auch:

- Tabellen und Bilder
- References, also die Literaturliste
- Citations – also Titel, die den aufgerufenen nutzen
- Related Titles/verwandte Titel

Das „Verständnis“ der KI von bestimmten Begriffen hilft dabei, Tabellen und Bilder aus den Artikeln zu extrahieren und einzeln darzustellen, die Literaturliste auszuwerten (References), weitere Titel zu finden, die entweder den gerade aufgerufenen zitieren (Citations) oder thematisch verwandt sind (Related Titles).

Vergleicht man diese Art der Recherche mit der „auf Papier“ und „am Regal“, entspricht der Hinweis auf die verwendete Literatur dem Schneeballsystem, bei dem man die Referenzen des vorliegenden Artikels darauf prüft, ob sie zum eigenen Vorhaben passen. Die Nutzung der verwandten Begriffe passt zum sogenannten „berry picking“: also zu schauen, was neben dem Titel noch so steht, wenn die Regale der Bibliothek systematisch sortiert sind. Auch Schlagwörter bei Online-Katalog-Einträgen kann man so nutzen: Ruft man bei der Angabe eines Titels in einem OPAC die verlinkten Schlagwörter auf, ist man auf diesen einzelnen Begriff beschränkt. „Berry picking“ ist also noch unsystematischer als das Schneeballsystem, bei dem eine Beziehung zwischen dem Artikel, den Sie lesen, und den Referenzen besteht. Bücher, die

in demselben Regal stehen oder Titel, die Schlagwörter miteinander teilen, sind im Vergleich dazu eher lose miteinander verknüpft.

Einschränkungen bei Semantic Scholar

Menschen, die Beiträge aus anderen als englischsprachigen Quellen benötigen, müssen auf jeden Fall weitere Quellen nutzen, denn Semantic Scholar hat seinen Schwerpunkt ganz klar im angelsächsischen Raum. Sehen kann man das in der Liste der Partner*innen – die Verlage und Repositorien sind großenteils in den USA.

Semantic Scholar beschränkt sich auf Titel, die frei zugänglich sind – lizenzierte Journals können Sie hier nicht finden: „Semantic Scholar is free to use and unlike similar search engines (i.e. Google Scholar) does not search for material that is behind a paywall.“^{[5]⁴¹}

Textgenerierende KI und Literaturrecherche?

Dass ChatGPT und zuverlässige Recherche mit belastbaren Angaben nicht zusammenpassen, ist inzwischen vielen klar. Vor allem, wenn es um die kostenlose Version gibt, deren Weltwissen im September 2021 endet und die keinen Zugriff aufs WWW hat. Auch GPT 4, das bei ChatGPT+ (dem Bezahl-Modell von Open.ai) und z.B. beim Chatbot von Bing eingesetzt wird, ist keine verlässliche Quelle – die KI „halluziniert“ sich im Zweifel Quellen zusammen, wenn man danach fragt.

Eine andere textgenerierende KI ist Perplexity.ai, das vom „Virtuellen Kompetenzzentrum - Schreiben lehren und lernen mit KI“ in der Liste von KI-Tools zur Recherche (VKKIWA) mit genannt wird. Grund genug, diesen Aspekt einmal zu testen:

Auch Perplexity.ai beruht auf GPT von Open.ai – doch bringt es zu seinen Texten immer Links. Ist das also eine verlässliche Rechercheoption, die zuverlässige Ergebnisse bringt? Ein kleiner Test mit einem ausführlichen Prompt:

„Nenne mir fünf aktuelle wissenschaftliche Artikel rund um das Thema Brustrekonstruktion nach einer Brustamputation aufgrund einer Krebserkrankung, in denen die Argumente der Betroffenen für die Rekonstruktion untersucht wurden.“

Schon das braucht mehr Zeit als das Eintippen der relevanten Begriffe in eine oder mehrere Suchzeilen bei PubMed oder LIVIVO. Zumindest für Menschen, die regelmäßig mit Boole'schen Operatoren, Befehlen und Filtern bei der Literaturrecherche arbeiten.

Wenn Sie textgenerierende KI nutzen, wie z.B. Perplexity.ai, sollten Sie die Texteingaben, die sogenannten Prompts, trainieren. Sie müssen detailliert und verständlich sein und möglichst viele Aspekte abdecken. Ja, Sie können im „Dialog“ mit der KI Informationen nachbessern – so wie Sie bei der Trefferliste einer Datenquelle Filter setzen –, aber bis so ein Prompt sitzt, das braucht Übung. Sie sollten möglichst ausführlich die verschiedenen Aspekte Ihrer Forschungsfrage einbauen. Im „Dialog“ mit der KI können Sie schreiben, was fehlt, welche Informationen nicht zutreffen und wo Sie Ergänzung wünschen. Das ist ein Prozess in natürlicher Sprache. Der Umstand, in vollständigen Sätzen zu schreiben, kann die Recherche auch vereinfachen – v.a., wenn es um eher einfache Fragen geht.

Die von Perplexity.ai aufgeführten Ergebnisse meiner Recherche im Einzelnen:

- Deutsches Ärzteblatt,
- die Site apotheke.de,
- die Site einer Klinik in der Schweiz,
- die Wiener Klinische Wochenschrift,
- die Leitlinie zum Thema von 2015.

Die gefundenen Artikel waren aus den Jahren 2020, 2021 und 2022 und damit einigermaßen aktuell, für ein medizinisches Thema allerdings ungewöhnlich „alt“, wie die Kontrollrecherche bei PubMed noch zeigen wird.

Beim Wiedereintippen des Prompts ein paar Minuten später brachte Perplexity.ai dieselben Treffer. Das ist eine Verbesserung gegenüber ChatGPT, das jedes Mal neue

1 Quelle: https://en.wikipedia.org/wiki/Semantic_Scholar#Technology (letzter Zugriff am 05.09.2023).

Ergebnisse produziert und die Reproduzierbarkeit von Recherche so unmöglich macht.

Jeder Treffer hat bei Perplexity.ai einen kurzen Satz zum Inhalt – die KI fasst den Text möglichst kurz zusammen. Aber: Die Treffer sind keine wissenschaftlich relevanten Quellen. Selbst der Artikel aus der Zeitschrift „Wiener Klinische Wochenschrift“ nicht, denn es handelt sich dabei nicht um eine Fachzeitschrift zum Thema.

Eine kurze Kontrollrecherche bei PubMed mit dem auf (englische) Stichwörter reduzierten Prompt: „breast reconstruction breast cancer surgery arguments for reconstruction“ erbrachte mindestens einen passenden Treffer für das Jahr 2023 – also aktueller als die Treffer von Perplexity.ai. Für den Zeitraum von fünf Jahren erbrachte PubMed sieben Treffer, von denen mindestens drei vollumfänglich der Frage entsprachen. Die Treffer von Perplexity.ai waren dagegen Überblicksdarstellungen, in denen das Thema teils nur in einem Nebensatz behandelt wurde, teils gar nicht.

Perplexity.ai bringt Links – und oft passen sie sowohl zur Aussage als auch zum Thema. Allerdings nicht immer. Am ehesten noch bei einer Internetrecherche zu allgemeinen Themen.

Die Suche nach dem Prompt: „Wie beeinflussen Klimawandel oder Ozeanerwärmung die Korallenbleiche? Antwort bitte auf Deutsch“ brachte eine ausführliche Antwort. Oberhalb des Textes sind die verwendeten Links angegeben, die mit Fußnoten im Text zitiert werden. Der erste Abschnitt bringt inhaltlich passende Aussagen: Die Ziffern der Fußnoten machen deutlich, dass vor allem die erste Quelle – ein Beitrag von ARDalpha – zur Erstellung herangezogen wurde.

Perplexity.ai hat analog zu Suchmaschinen keinen Zugriff auf Datenbanken. Deshalb ist eine Recherche mit dem Tool nur innerhalb des WWW möglich. Der ursprüngliche Test mit dem Tool fand im Juli 2023 statt – bei dem Versuch, ihn im August 2023 zu wiederholen, erschien ein anderer Text. Textgenerierende KI geben keine konsis-

Wie beeinflussen Klimawandel oder Ozeanerwärmung die Korallenbleiche? Antwort bitte auf Deutsch

⚡ Quick Search

Klimawandel: Warum das Korallensterben uns alle...
α* ardalpha · 1

Klimakrise und Meer – WWF Deutschland
wwf · 2

Great Barrier Reef: Wie der Klimawandel das...
youtube · 3

Forschungsthema: Ozeanerwärmung – ESKP
eskp · 4

☰ Answer

Der Klimawandel und die Ozeanerwärmung haben einen großen Einfluss auf die Korallenbleiche ¹ ². Die Erwärmung der Ozeane, die auf den Klimawandel zurückzuführen ist, ist die größte Gefahr für alle Korallen ¹. Tropische Korallen benötigen eine Wassertemperatur zwischen 20 und 29 Grad, um überleben zu können ¹. Wenn die Wassertemperatur zu hoch wird, sterben die einzelligen Algen, mit denen Korallen in Symbiose leben, ab und die Korallen bleichen aus ¹. Die Korallenbleiche ist ein wichtiger Indikator für den Gesundheitszustand eines Riffs ¹. Die Erwärmung des Ozeans führt auch dazu, dass Wasserschichten sich schlechter durchmischen, was die Korallen zusätzlich belastet ². Weltweit sind bereits ein Großteil der Korallenriffe von der Erwärmung und

tenten Aussagen ab, die bei ein und derselben Frage in zeitlichem Abstand immer das gleiche Ergebnis bringen. Damit sind die Ergebnisse nicht reproduzierbar und für wissenschaftliche Belange nicht geeignet.

Fazit

Der Einsatz von KI-gestützten Tools in der Literaturrecherche ist eine Option, bei der Sie viel abwägen müssen. Wie bei anderen Recherche-Tools gilt auch für KI-gestützte Hilfsmittel, dass verschiedene Quellen die Gefahr verringern, nicht alle relevanten Titel zu erfassen. Sie benötigen also Hintergrundwissen zu den Tools: zur Quellenlage und ggf. zu thematischen Schwerpunkten. Um Ihre Recherche ausreichend breit aufzustellen, sollten Sie sich klarmachen, welche Quellen Sie benötigen, und überprüfen, ob die Hilfsmittel die Ihrem Vorhaben entsprechende Grundlage bieten, also zum Thema passen und einen passenden Publikationszeitraum abdecken. Meist ist das nur in Kombination verschiedener Tools zu erreichen. Das gilt für KI-gestützte Angebote wie für die bisher üblichen Angebote, wie OPAC und Datenbanken.

Im weiteren Verlauf der Reihe stelle ich Ihnen verschiedene Tools vor, die explizit für Literaturrecherche konzipiert wurden oder Literaturrecherche als wichtiges Element berücksichtigen. Wie sich der Einsatz von KI dabei bemerkbar macht, welche Vorteile daraus entstehen und welche Risiken eventuell zu berücksichtigen sind, werden neben einer Anleitung für die Tools meine Themen sein.



© Susanne Fern

Die Autorin

Heike Baller arbeitet seit 1995 als freiberufliche Rechercheurin. Ihr in diesen Jahren erworbenes Wissen rund um Internet- und Literaturrecherche gibt sie in Seminaren und Workshops weiter und bietet individuelle Unterstützung in Form von 1:1-Recherche-Coachings. Ihr Blog bei Profi-Wissen ist eine Fundgrube voller Recherche-Tipps. <https://www.profi-wissen.de/recherche-blog/>