



Literaturrecherche in Zeiten der KI

Consensus und Elicit

von Heike Baller

Welchen Nutzen bringen KI-gestützte Tools bei der Literaturrecherche tatsächlich? Inzwischen steht eine ganze Reihe unterschiedlicher Werkzeuge zur Verfügung. Das Element „KI“ bezieht sich dabei vor allem auf den Einsatz textgenerierender KI. Damit ist eine Kommunikation in natürlicher Sprache möglich, die sich von der Literaturrecherche mit Kombinationen aus Suchbegriffen unterscheidet. Das klingt sehr attraktiv, denn die Suche nach passenden Wörtern für die Recherche kann aufwändig sein. Der Anteil dieser KI an den Recherche-Tools hat jedoch auch Nachteile. An zwei Tools möchte ich das erläutern:

- Consensus – ein frei nutzbares Tool, für das Sie einen kostenlosen Account benötigen
- Elicit – ein Tool, das nach einer bestimmten Nutzungsmenge nur kostenpflichtig weiter nutzbar ist

Consensus

Bei Consensus handelt es sich um ein Tool einer privaten Firma, das mit Hilfe von KI die Literaturrecherche unterstützen soll. Die Besonderheiten bestehen darin, dass Sie am besten mit einer Frage in natürlicher Sprache beginnen und zu den Ergebnissen verschiedene Arten von Zusammenfassungen gehören.

Sie benötigen einen kostenlosen Account, den Sie mit Ihrer Mail-Adresse anlegen können. Sie erhalten pro Monat 20 sogenannte „AI-Credits“. Diese erlauben die Nutzung ergänzender Optionen, z.B. faktentreuere Zusammenfassungen mit dem Modell GPT4. Für die, die eine unbegrenzte Anzahl dieser Credits benötigen, gibt es die Bezahlversion von Consensus. Die allererste Frage funktioniert noch ohne Account – aber wenn Sie damit weiter arbeiten wollen, müssen Sie sich anmelden.

Consensus arbeitet mit den bibliografischen Daten von Semantic Scholar, einem Tool zur Literaturrecherche, das ich im letzten Heft vorgestellt habe. Semantic Scholar verwendet in erster Linie Daten aus englischsprachigen Quellen und hat den Schwerpunkt auf frei verfügbaren Artikeln. Mit anderen Worten: Lizenzierte wissenschaftliche Literatur oder Monografien und Sammelwerke kommen hier nicht vor. Das ist eine Einschränkung der durchsuchbaren Literatur schon bei Semantic Scholar. Bei Consensus kommt noch hinzu, dass die KI-gestützte Auswertung nicht alle Titel aus Semantic Scholar umfasst. Die Basis der von dem Tool zur Verfügung gestellten Literatur ist also noch schmaler als die von Semantic Scholar selbst.

Die Oberflächensprache von Consensus ist Englisch. Tatsächlich sind die Ergebnisse nach einer englischsprachigen Frage besser als mit deutschsprachigem Input.

Wie funktioniert Consensus in der Praxis?

Wenn Sie mit Consensus arbeiten, geben Sie in die Suchzeile Ihre Frage ein. Sie erhalten dann eine Liste von Titeln, deren größtes Element das „Summary“ ist – eine Zusammenfassung in einem Satz. Dieser Satz steht unter dem Titel des Artikels und ist farbig hinterlegt. Die Zusammenfassung soll Ihnen die Einordnung des Artikels in Ihrer Recherche erleichtern.

Die Qualität dieser Summaries ist sehr unterschiedlich. Ich habe darunter auch Sätze mit völlig unwahren Aussagen gefunden. Sie sollen auf Abstract, Titel und Keywords des betreffenden Artikels beruhen. Da es sich um Texte einer textgenerierenden KI handelt, wird für jede dieser Ein-Satz-Zusammenfassungen der nächste Wortbestandteil nach Wahrscheinlichkeit berechnet. Das ist das Prinzip textgenerierender KI. Für einen Beitrag über die Religionspolitik von Heinrich VIII z.B. hieß es im Summary, Thomas Cromwell sei Mitglied des Klerus gewesen, was nicht stimmt. Das Abstract des Artikels bietet keinerlei Anlass für diese Annahme. Hier die beiden Textausschnitte (Recherche am 11.3.2024):

Der Titel: „THE MAKING OF RELIGIOUS POLICY, 1533–1546: HENRY VIII AND THE SEARCH FOR THE MIDDLE WAY“

Das Summary: „Thomas Cromwall was a prominent member of the clergy during the reign of Henry VIII, playing a significant role in the making of religious policy.“

Bei Aufruf der Seite mit allen Angaben zum Artikel gab es dann folgende Informationen:

Zusammenfassung auf der Seite des Artikels:

„Key Takeaway: Henry VIII was the dominant force in religious policy during his reign, and his policy, characterized as a search for the middle way, is best characterized as a consistent pursuit.“

Abstract: Too often religious policy in Henry VIII's reign after the break with Rome has been seen as fluctuating and inconsistent as he was influenced first by one group of ministers and courtiers and then by another. Here it is argued by contrast that Henry VIII was very much the dominant force in the making of religious policy and that his policy, which he pursued skilfully and consistently, is best characterized as a search for the middle way.“

Weder aus Abstract noch aus dem Titel geht irgendeine Information zu Thomas Cromwell hervor. Die Aussage, er gehörte dem Klerus an, ist eindeutig falsch – Thomas Cromwell war Jurist.

Andere Summaries zu derselben Rechercheanfrage stimmen zwar mit der Wirklichkeit überein, ergeben sich allerdings ebenfalls nicht aus dem Abstract oder Titel der entsprechenden Artikel. Bei der Einschätzung solcher Zusammenfassungen ist also eine Kenntnis des Themas unerlässlich.

Die weiteren Angaben zu den einzelnen Artikeln stehen unterhalb dieser Zusammenfassung und sind sehr knapp: Zeitschriftentitel, Autor*innennamen, wie oft der Text zitiert wurde und das Jahr. Erst bei Aufruf der Seite zum Titel gibt es die vollständigen Angaben. Die Darstellung ist nicht sehr übersichtlich, v.a., weil sie grafisch gegen andere Elemente wie „Summary“ oder „Key Takeaway“ abfällt.

Auffällig war bei den Recherchen, mit denen ich das Tool getestet habe, dass die zuoberst genannten Treffer oft schon älter waren – teils aus den 80er oder 90er Jahren oder den frühen 2000er Jahren. Bei einem Technik-Thema von hoher Aktualität – es ging um Batterie-Forschung – kam erst an achter Stelle ein Artikel von 2021. In der Gesamtliste waren aktuelle Titel vertreten, teilweise von 2024, die auch genau zur Frage passten, doch zum Vorschein kamen sie erst mit Hilfe des Filters für den Publikationszeitpunkt.

Die AI-Credits und was sie können

Pro Monat gibt es 20 AI-Credits – damit können Sie weitere Funktionen wie z.B. „Synthesize“ nutzen. Den Button dazu finden Sie oberhalb der Treffer. Es gibt dort auch einen Button für den Co-Piloten. Wenn Sie „Synthesize“ anklicken, erhalten Sie oberhalb der „Trefferliste“ ein Summary verschiedener Artikel. Zurzeit – März 2024 – handelt es sich noch um ein Angebot im Beta-Stadium; es ist also in der Entwicklung. Es wird angegeben, wie viele Artikel für das Summary analysiert wurden und dann folgen einige Sätze, die den Inhalt wiedergeben. Oder wiedergeben sollen. Im Recherchebeispiel vom

März 2024 wurde wieder Thomas Cromwell als Mitglied des Klerus und der späteren anglikanischen Kirche bezeichnet. Diese Aussagen treffen nicht zu.

Auch der Co-Pilot ist ein Angebot im Beta-Status. Hier gibt es einen längeren Text, der Erkenntnisse bündelt und mit Fußnoten auf seine Quellen verweist. Dieser Text war in meiner Beispiel-Recherche inhaltlich von besserer Qualität, wies aber immer noch Fehlzuordnungen zu Thomas Cromwell auf. Die zusammengerechneten Texte der KI-Tools kommen in dieser Form – mit Aufzählungspunkten und Fazit – sehr überzeugend daher und es bedarf fundierten Wissens, um die unzutreffenden Aussagen auszusondern.

Die dritte Funktion, die mit AI-Credits genutzt werden kann, ist der „Consensus Meter“. Dafür müssen Sie in die Suchzeile eine Forschungsfrage eintippen, die mit „Ja“ oder „Nein“ beantwortet werden kann. Dann bekommen Sie – bei der Nutzung von Synthesize – eine kleine Grafik angezeigt, aus der hervorgeht, wie viele der gefundenen Treffer die Antwort mit Ja, Nein oder Vielleicht beantwortet haben.

Copilot beta
8 papers analyzed

Introduction

The relationship between Henry VIII and Thomas Cromwell is a subject of historical interest, often analyzed through the lens of political maneuvering, religious reform, and personal dynamics within the Tudor court.

Key Insights

- Thomas Cromwell played a pivotal role in the execution of Anne Boleyn, leveraging legal systems to satisfy his ambitions and the king's desires, with Henry VIII's approval ¹.
- Cromwell was instrumental in the English Reformation, complementing Henry VIII with his attention to detail and imaginative power, without holding supreme power himself ².
- Despite his ultimate execution for treason and other charges, Cromwell is recognized as one of England's greatest statesmen, having significantly influenced the Reformation and government ³.
- Contrary to some portrayals, Cromwell rose to power not as a Machiavellian figure or revolutionary, but as a highly capable administrator serving the Crown ⁴.
- Cromwell's legacy has been debated, with some historians viewing him as a key architect of modern English governance, while others see him as a mere instrument of Henry VIII's policies ⁵.

Als letztes gibt es noch den „Study Snapshot“ zu den einzelnen Titeln. Damit wird aufgezeigt, welches Design einer Studie zugrunde liegt, wie groß die Datengrundlage ist, welche Bevölkerungsgruppe interviewt wurde und wie die Ergebnisse lauten.

Fazit zu Consensus

Kurz auf den Punkt gebracht: Die KI-Unterstützung bei Consensus verspricht mehr, als sie halten kann.

- Die Datenbasis von Semantic Scholar ist bereits eingeschränkt, da v.a. frei verfügbare Artikel aus englischsprachigen Quellen verwendet werden. Consensus kann nicht alle diese Titel mit der eigenen KI-Technik erschließen, so dass die Basis noch schmaler wird.
- Textgenerative KI produziert Texte nach dem Prinzip der statistischen Wahrscheinlichkeit – bei Chatbots wie ChatGPT sind die dadurch entstehenden Fehler als „Halluzination“ bekannt. Dieselbe Fehleranfälligkeit zeigen die Summaries von Consensus.
- Die Reihenfolge der Treffer ist intransparent – weder stehen die jüngsten noch die am häufigsten zitierten Titel oben. Erst mit Filtereinstellungen lassen sich die Titel nach Erscheinungszeitpunkt durchsuchen.

Elicit

Hinter Elicit steht eine gemeinnützige Organisation mit Namen Ought. Das Projekt ist älter als der Launch von ChatGPT, der so viele weitere KI-gestützte Tools nach sich gezogen hat. Elicit weist viele Parallelen zu Consensus auf: die Datenbasis über Semantic Scholar – frei verfügbare, in erster Linie englischsprachige Literatur. Das Finetuning der eigenen KI kann nicht alle Datensätze von Semantic Scholar erfassen – die Grundlage für die Recherche wird noch einmal schmaler. Da ist einiges im Bewegung: Semantic Scholar bekommt immer mehr Artikel in seinen Bestand – Elicit (und auch Consensus) können zunehmend mehr Artikel mit den eigenen Tools bearbeiten. So kann es passieren, dass eine Suchanfrage nach ein paar Wochen andere Ergebnisse zeitigt. Nach welchen Kriterien sich diese Form der „Aktualisierung“ abspielt, ist nicht klar. Es ist nicht gesagt, dass auf diese Weise aktuellere Treffer mit Hilfe der KI zusammengefasst werden.

Auch für Elicit benötigen Sie einen Account, den Sie kostenlos mit E-Mail-Adresse anlegen können. Sie starten mit 5.000 Credits. Sind diese aufgebraucht, geht es nur noch mit einem Bezahlmodell weiter. 5.000 Credits klingt nach einer Menge, wenn man es mit den 20 AI-Credits von Consensus vergleicht – doch das täuscht. Eine Suche nach Artikeln „kostet“ 100 bis 120 Credits. Elicit hat neben der Literaturrecherche noch weitere Tools im Angebot. Diese sind deutlich „teurer“. Wie bei Consensus ist auch hier die Oberflächensprache und die am besten bediente Anfragesprache Englisch.

Mit Elicit Literatur suchen

Auch bei Elicit startet Ihre Literaturrecherche mit einer Forschungsfrage in natürlicher Sprache. Als Ergebnis bekommen Sie eine zusammenfassende Antwort auf die Frage mit Verweis auf einzelne Artikel. In der kostenlosen Version von Elicit werden hier vier Artikel analysiert, in der kostenpflichtigen acht.

Darauf folgt dann die Liste der gefundenen Titel. Sie sind in Spalten angeordnet. Ganz links gibt es die Informationen zum Titel. Die zweite Spalte „Abstract Summary“ bietet dann eine Ein-Satz-Zusammenfassung, die mit Hilfe textgenerierender KI erstellt wird.

Sie können nun weitere Spalten hinzufügen – dabei handelt es sich um Informationen, die im Falle von Studien interessant sind: „Methodology“, „Outcome Measured“ oder „Invention“. Die Einstellung für die Reihenfolge der Titel ist „Relevanz“ – Sie können wechseln zu „most cited“ und „least cited“, sich die Titel alphabetisch anzeigen lassen oder nach Erscheinungszeitpunkt sortiert. Die Filtermöglichkeiten umfassen

- Schieberegler für einen Publikationszeitraum
- Auswahl des Studien-Typs
- Ein- oder Ausschluss bestimmter Begriffe im Abstract

Die Ein-Satz-Zusammenfassungen – Abstract Summary genannt – unterliegen wie in Consensus den Beschränkungen textgenerativer KI. Auch hier gibt es Zusammenfassungen, die fehlerhafte Informationen verbreiten. Wenn Ihnen ein solcher Satz also suspekt vorkommt, können Sie die Seite für den Artikel durch Anklicken des

Titels aufrufen und das Abstract einsehen, um es mit dem Summary abzugleichen.

Auch bei Elicit hatte ich eine Frage zu Thomas Cromwell gestellt. Ein Titel, der recht weit oben, also als relevant, gerankt wurde, hatte nicht Thomas Cromwell, sondern Thomas Cranmer als Thema. Da beide Heinrich VIII dienten, wenn auch in unterschiedlicher Funktion, kommt Thomas Cromwell sicher auch in diesem Werk vor. Doch weder das Abstract-Summary noch das Abstract selbst gaben darüber Aufschluss – nach einem ersten Eindruck handelte es sich um einen Treffer, der nicht zum Thema passte, das ausdrücklich Thomas Cromwell nannte. Auch ein Werk zu den Kreuzzügen, das recht weit oben landete, hatte weder nach Titel noch Abstract etwas mit Thomas Cromwell zu tun. Hier machte sich die Grundlage der KI bemerkbar, bei der es um statistische Wahrscheinlichkeit, also um Häufigkeit von Wortverbindungen geht. „Thomas“ ist demnach häufiger mit „Cranmer“ verbunden als mit Cromwell. In der Liste der „Treffer“ befanden sich noch einige solcher Fehler – in einem Titel ging es um Henry VII... Genaues Lesen und Faktenkenntnis sind für die Einschätzung der Ergebnisse nötig.

Ein Vorteil: Wenn es sich um völlig frei verfügbare Titel handelt, können Sie auf der Seite, die sich nach Klicken auf den Titel öffnet, den Text direkt einsehen. Der Nachteil hierbei: Der Text ist unformatiert, die Lesefreundlichkeit also nicht so gegeben. Mit der Tastenkombination Strg + f (cmd + f bei Mac) können Sie aber zentrale Be-

griffe im Text aufspüren und sehen, ob sich eine gründliche Lektüre lohnt.

Weitere Angebote von Elicit

Elicit bietet neben der Literaturrecherche zwei weitere Optionen an:

- Daten aus PDFs ziehen
- Eine Liste von Begriffen rund um ein Thema (*list of concepts*) erstellen

Die Möglichkeit, mit einem PDF zu „reden“, also direkt Fragen zum Inhalt zu stellen und Antworten darauf zu erhalten, gibt es bei verschiedenen Tools – Elicit bietet sie auch. Dazu muss man das entsprechende PDF hochladen. Bitte beachten Sie dabei die Urheberrechte! Sie müssen über die Rechte am Paper verfügen oder es muss gemeinfrei sein. Sie erhalten dann im Idealfall die in der Studie erarbeiteten Ergebnisse. Bei einem Versuch, ein solches PDF auslesen zu lassen, bekam ich allerdings nur die bibliografischen Daten. Das PDF enthielt durchaus Tabellen und andere auslesbare Daten.

Eine *list of concepts* oder Begriffsliste kann dabei helfen, ein Thema von verschiedenen Blickpunkten aus zu betrachten. Dazu gibt Elicit eine Liste von Begriffen aus, die in Zusammenhang mit dem ursprünglichen Begriff vorkommen. Bei meinem Test war es „responsible leadership“, ein Begriff aus dem Bereich Unternehmensführung. Als Ergebnisse bekam ich sowohl Synonyme angezeigt als auch Unter- und Oberbegriffe dazu. Diese Ergebnisse

The screenshot shows the Elicit interface with a light blue background. At the top, there are buttons for 'Synthesize' and 'Copilot'. On the right, there are buttons for 'Filter', 'Save search', and 'Share'. The main content is divided into two panels. The left panel is titled 'Summary beta' and indicates '8 papers analyzed'. It contains a text block: 'These studies suggest that Thomas Cromwell was Henry VIII's chief minister and secretary, playing a crucial role in the English Reformation and the king's governance, but his methods and alliances eventually led to his execution for multiple charges including treason.' The right panel is titled 'Consensus Meter beta' and contains a message: 'Not a 'Yes or No' question. To use the Consensus Meter, try asking a simple Yes/No question.' Below this message is a search bar with the text 'Was Thomas Cromwell a key figure in the Englis...' and a magnifying glass icon.

gab es wieder in einer Tabelle – links den Begriff mit einer kurzen Definition, rechts Zitate aus Artikeln, die den Kontext näher beleuchten mit Link zu dem zitierten Beitrag. Das Ranking dieser Liste ist nicht transparent und auch nicht filter- oder sortierbar.

Fazit zu Elicit

Auch bei Elicit verspricht die KI mehr, als sie halten kann. Ähnlich wie bei Consensus liegt nur eine schmale Datenbasis zugrunde – Semantic Scholar mit den Elicit-eigenen Einschränkungen, da die KI nicht alle Titel aus der Datenbasis bearbeiten kann.

Textgenerative KI erstellt Texte nach der Wahrscheinlichkeit, mit der der nächste Wortbestandteil folgt – die Faktentreue bleibt dabei oft auf der Strecke. Es kommen Zusammenfassungen vor, die mit den Informationen aus Titel und Abstract nicht zusammenpassen. Bei Elicit kamen zudem noch Treffer vor, die überhaupt nicht zum Thema passten.

Um Nutzen aus der Liste ziehen zu können, muss der Prompt möglichst aussagekräftig, ausführlich und konkret sein. Danach ist dann Arbeit mit den Filtermöglichkeiten nötig – dabei kann die Option, Keywords im Abstract ein- oder auszuschließen, tatsächlich hilfreich sein.

Die beiden anderen Angebote von Elicit sind ebenfalls nicht einfach so zu nutzen: Wenn Sie Daten aus einem PDF ziehen wollen, muss die Rechtslage klar sein. Um eine Liste von Begriffen rund um ein Thema zu erstellen, ist die *list of concepts* ganz „nett“, kostet Sie aber bis zu 2.500 Credits – also die Hälfte des Guthabens. Da die Reihenfolge der Liste nicht begründet ist und Sie sie auch nicht bearbeiten können, hält sich der Nutzen in Grenzen.

Fazit zu beiden Tools

Die beiden KI-gestützten Tools zur Unterstützung der Literaturrecherche – Consensus und Elicit – versprechen zum jetzigen Stand mehr, als sie halten können. Der Vorteil, mit Fragen in natürlicher Sprache agieren zu können, wird von der Notwendigkeit aufgewogen, diese Fragen möglichst präzise und konkret zu stellen, damit keine oder nur wenige fehlerhaften Treffer in der Liste auftau-

chen. Das Ranking der Treffer ist nicht nachvollziehbar, die Auswahl der „besten“ Titel für die Zusammenfassung aus mehreren Titeln ebenso wenig. Um Nutzen aus der Recherche ziehen zu können, gibt es noch die Möglichkeit, die Liste mit Hilfe von Filtern und Sortieroptionen zu verändern und zu durchsuchen. Dieser Vorgang unterscheidet sich nicht von Datenbank- und Katalogrecherche. Der Vorteil, dass Ein-Satz-Zusammenfassungen bei der Einordnung einzelner Titel helfen sollen, wird durch fehlerhafte KI-Texte zunichte gemacht. Um diese Sätze einordnen zu können, bedarf es bereits gründlicher Kenntnisse rund ums Thema.

KI-gestützte Tools befinden sich immer noch in der Entwicklung – und sie werden es bleiben. Daher sollten wir sie kritisch im Auge behalten, um mitzubekommen, wann sie nicht nur zur Erstellung von Texten hilfreich sind, sondern auch bei einer Suche in natürlicher Sprache die Arbeit wirklich erleichtern.



© Susanne Fern

Die Autorin

Heike Baller arbeitet seit 1995 als freiberufliche Rechercherin. 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/>