

u^b

«If it's not free, you're the product»

Datafizierte Wissenschaftskommunikation als Herausforderung für Bibliotheken

Kathi Woitas, M.A.

Bibliosuisse-Kongress, 02.11.2023

DOI: [10.5281/zenodo.10068283](https://doi.org/10.5281/zenodo.10068283)

u^b

Datafication (2013)

...is not the same as digitization, which takes analog content [...] and converts it into digital information, a sequence of ones and zeros that computers can read.

*...is a far broader activity: **taking all aspects of life and turning them into data** [...]*

*Once we **datafy things**, we can **transform their purpose** and turn the information into **new forms of value**.*

Big-Data-Technologien: verteilte, redundante, beliebig skalierbare Datensysteme

Data Science: ML, Deep Learning („AI“), NLP, **generative AI** (LLM, VLM)

Digital Scholarship Services

Neue Dienstleistungen von wissenschaftlichen Bibliotheken für die datenbasierte Forschung

Kathi Woitas, Universitätsbibliothek Bern

27.10.2021 Bibliosuisse Kongress

Large-scale language models

A robot wrote this entire article. Are you scared yet, human?

GPT-3

We asked GPT-3, OpenAI's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

- For more about GPT-3 and how this essay was written and edited, please read our editor's note below

„A Robot Wrote This Entire Article. Are You Scared yet, Human?“ *The Guardian*, 8. September 2020.

<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>

GPT-3

- Zusammenfassungen
- Übersetzungen
- Dialog-Generierung
- Semantische Suche
- Vervollständigung von Programm-Code
- ...?

u^b Wissenschaftskommunikation

Aufgaben

- Bekanntmachung, Qualitätsprüfung und Dokumentation wiss. Erkenntnisse
- Zuschreibung von Urheberschaft und Reputation

→ Schreiben, Reviewen, Editieren, Rezipieren = **Praktiken** (und Normen)

→ institutioneller + infrastruktureller Rahmen = **Business**

Wie wirkt sich die Datafizierung auf diese Ebenen aus?

Was heisst das für Bibliotheken?

u^b Praktiken: Schreiben mit AI

u^b

Schreiben mit AI...

Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals

Guillaume Cabanac, Cyril Labbé, Alexander Magazinov

Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing

J. Elliott Casal^a  , Matt Kessler^b

Scientific sleuths spot dishonest ChatGPT use in papers

Manuscripts that don't disclose AI assistance are slipping past peer reviewers.

Gemma Conroy

Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened

Martin Májovský¹ ; Martin Černý¹ ; Matěj Kasal² ;
Martin Komarc^{3,4} ; David Netuka¹ 

Casal, J. Elliott, and Matt Kessler. „Can linguists distinguish between ChatGPT/AI and human writing? A study of research ethics and academic publishing“. *Research Methods in Applied Linguistics* 2, Nr. 3 (1. Dezember 2023): 100068.

<https://doi.org/10.1016/j.rmal.2023.100068>.

Májovský, Martin, Martin Černý, Matěj Kasal, Martin Komarc, and David Netuka. „Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened“. *Journal of Medical Internet Research* 25, Nr. 1 (31. Mai 2023): e46924. <https://doi.org/10.2196/46924>.

Cabanac, Guillaume, Cyril Labbé, and Alexander Magazinov. „Tortured phrases: a dubious writing style emerging in science. evidence of critical issues affecting established journals“. arXiv, 12. Juli 2021. <http://arxiv.org/abs/2107.06751>.

Conroy, Gemma. „Scientific Sleuths Spot Dishonest ChatGPT Use in Papers“. *Nature*, 8. September 2023. <https://doi.org/10.1038/d41586-023-02477-w>.

Introducing Microsoft 365 Copilot – your copilot for work

Mar 16, 2023 | Jared Spataro - CVP Modern Work & Business

Researcher Discovery



Affiliations



Scopus AI **Beta**

Learn with AI-generated overviews based on documents since 2018

[How it works](#)

What would you like to learn more about?



Scientific experimentation with generative AI

Gary Charness, Brian Jabarian, John List / 16 Oct 2023

Please be aware

While Scopus AI leverages trusted Scopus content, it is possible to produce misleading or inaccurate results. We recommend exercising your expertise and discretion before incorporating its outputs into official documents like dissertations or manuscripts. It is essential to

ChatGPT use shows that the grant-application system is broken

The fact that artificial intelligence can do much of the work makes a mockery of the process. It's time to make it easier for scientists to ask for research funding.

[Juan Manuel Parrilla](#)

<https://www.scopus.com/search/form.uri?zone=TopNavBar&origin=searchbasic&display=basic#scopus-ai> (Lizenz nötig)

Gary Charness, Brian Jabarian, und John List. „Scientific Experimentation with Generative AI“. *CEPR* (blog), 16. Oktober 2023.

<https://cepr.org/voxeu/columns/scientific-experimentation-generative-ai>.

Parrilla, Juan Manuel. „ChatGPT Use Shows That the Grant-Application System Is Broken“. *Nature*, 13. Oktober 2023. <https://doi.org/10.1038/d41586-023-03238-5>.

Spataro, Jared. „Introducing Microsoft 365 Copilot – Your Copilot for Work“. The Official Microsoft Blog, 16. März 2023. <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/>.

“Scholarly AI Taxonomy”

1. **Extract:** Identify and isolate specific entities or data points within the content.
2. **Validate:** Verify the accuracy and reliability of the information.
3. **Generate:** Produce new content or ideas, such as text or images.
4. **Analyse:** Examine patterns, relationships, or trends within the information.
5. **Reformat:** Modify and adjust information to fit specific formats or presentation styles.
6. **Discover:** Search for and locate relevant information or connections.
7. **Translate:** Convert information from one language or form to another.

Risiko: Verstöße gegen Wiss. Integrität

- **Vorspiegeln angeblicher Tatsachen** (engl. fabrication): Behauptung [...] von nicht existierenden Daten, Grundlagen oder Ergebnissen, auch das falsche oder irreführende Zitieren
- **Fälschung**: unlautere, vorsätzliche oder **grob fahrlässige** Manipulation [...] von Daten oder Ergebnissen und deren Darstellung
- **Plagiat**: eigene Leistung nicht hinreichend unterscheidbar von fremder oder früherer eigener Leistung
- Keine oder **unvollständige Angabe** von Daten und **Datenquellen**

u^b

Neue normative Aushandlungen

Reger Diskurs über Umgang mit LLM, Chancen + Risiken von AI-Schreibassistenz!

- Positionierungen von Editoren, z.B. aus [Humanities/Bioethics](#)
- Policies von Verlagen, z.B. [Sage](#)
- Empfehlungen von Forschungsförderern, z.B. [DFG](#)

→ vollumfängliche **Verantwortung des Autors**

→ **Transparenz/Offenlegung** aller Hilfsmittel (wie detailliert?)

→ klassisches Zitat bei wortwörtlichen Übernahmen?

u^b Praktiken: Reviewen + Editieren

u^b Gutachter und Herausgeber

Gutachter: Durchführung des Peer Review

- **Bewertung** des Manuskripts, Hinweise zur Verbesserung
- **Sicherung** von Qualität, Originalität, ethischen Standards
- **100 Mio Arbeitsstunden p.a.** global; **Wert: 1.5 Mrd. USD** nur in USA

Herausgeber: Überwachung des Peer Review

- **Auswahl** der Gutachter, Auswertung der Reviews
- **Entscheid** über Annahme oder Ablehnung



☆ System Recommendations

Elsevier is committed to furthering inclusion and diversity together with you as our editors, and strives to avoid creating or reinforcing unfair bias. In the whole list before sending invitations, taking gender, career stage, and global representation into consideration.

🔍 Keyword Search

Filter on h-index Filter on expertise Filter on connections Filter on review history

AIRA reads every research manuscript we receive and makes up to 20 checks a second. These checks cover, among other things, language quality, the integrity of figures and images, plagiarism, and conflicts of interest. The results give editors and reviewers another perspective as they decide whether to put a research paper through our rigorous and transparent peer review.

👥 Selected Candidates 0

📄 Manuscript Details

☰ Session Preferences

Reviewed for this journal 3 h-index 22

Similar works 2

Reviewed for this journal 20 h-index 26

Total invitations declined

Total invitations uninvited

Publication history

Years active 12

Total publications 242

Showing 3 most recent published works

(2023) [...](#)

(2023) [...](#)

Elsevier Journal Article Publishing Support Center. „Video Guide: Find reviewers using Scopus“. Zugegriffen 28. Oktober 2023. https://service.elsevier.com/app/answers/detail/a_id/33906/supporthub/publishing/

Frontiers Science Communications. „Science for All with Compatible AI“, 8. Juni 2023. <https://blog.frontiersin.org/2023/06/08/science-for-all-with-compatible-ai/>.

u^b

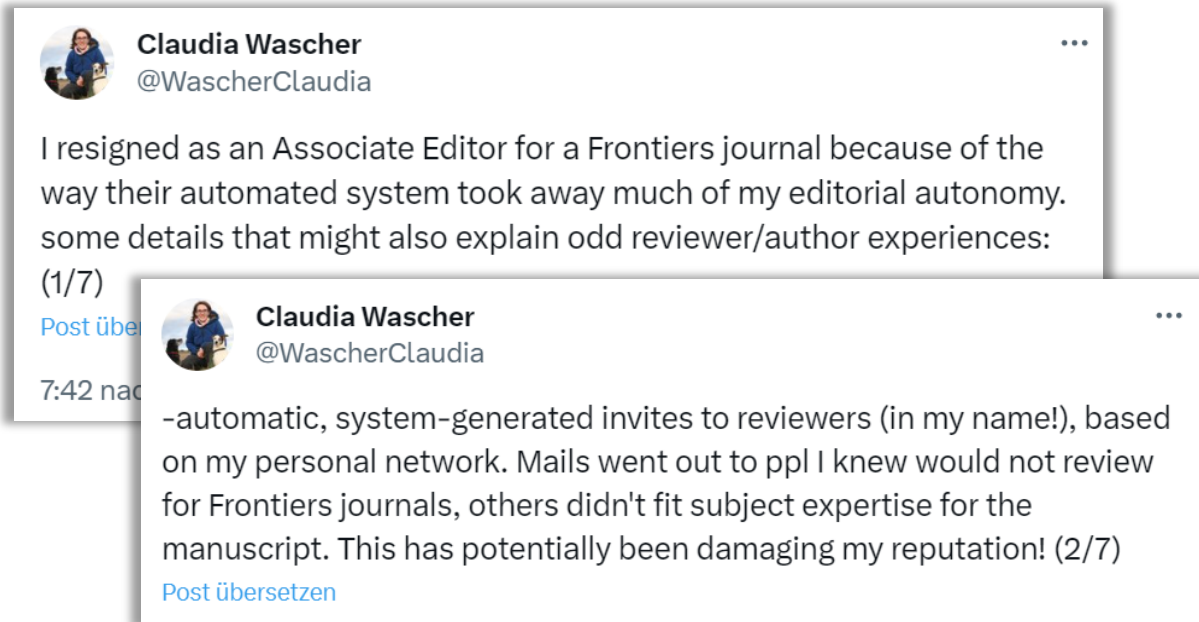
Peer Review in der Realität: Beispiel Frontiers

Sorgfalt? Autonomie? Qualität?

Default-Zeitraum: 7 Tage, 21 mit Verlängerung

Übersteuerung der Gutachter-Auswahl durch automatische Einladungen nach 3 Tagen

Basis: Topics/Similarity + Personendaten



Claudia Wascher @WascherClaudia

I resigned as an Associate Editor for a Frontiers journal because of the way their automated system took away much of my editorial autonomy. some details that might also explain odd reviewer/author experiences: (1/7)

Post übersetzen

7:42 nach

Claudia Wascher @WascherClaudia

-automatic, system-generated invites to reviewers (in my name!), based on my personal network. Mails went out to ppl I knew would not review for Frontiers journals, others didn't fit subject expertise for the manuscript. This has potentially been damaging my reputation! (2/7)

Post übersetzen



Wolf Hautz @wolfhautz · 6. Mai

Same story why I resigned from an editors position at a BMC journal. This practice hurts science.

10 5.009

Serge Horbach, Michael Ochsner, und Wolfgang Kaltenbrunner. „Reflections on Guest Editing a Frontiers Journal“. Leiden Madtrics (blog), 31. Oktober 2022. <https://www.leidenmadtrics.nl/articles/reflections-on-guest-editing-a-frontiers-journal>

Claudia Wascher [@WascherClaudia]. „I Resigned as an Associate Editor for a Frontiers Journal Because of the Way Their Automated System Took Away Much of My Editorial Autonomy. Some Details That Might Also Explain Odd Reviewer/Author Experiences: (1/7)“. Tweet. *Twitter*, 5. Mai 2023. <https://twitter.com/WascherClaudia/status/1654542216218411008>.

AI peer reviewers unleashed to ease publishing grind

A suite of automated reviewers that humans are still in the

ReviewerGPT? An Exploratory Study on Using Large Language Models for Paper Reviewing

Ryan Liu, Nihar B. Shah

A handful of academic publishers are piloting AI tools to do anything from selecting reviewers to checking statistics and summarizing a paper's findings.

Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review

[Mohammad Hosseini](#) ✉ & [Serge P. J. M. Horbach](#)

Heaven, Douglas. „AI Peer Reviewers Unleashed to Ease Publishing Grind“. *Nature* 563, Nr. 7733 (22. November 2018): 609–10. <https://doi.org/10.1038/d41586-018-07245-9>.
Liu, Ryan, and Nihar B. Shah. „ReviewerGPT? An exploratory study on using large language models for paper reviewing“. *arXiv*, 1. Juni 2023. <https://doi.org/10.48550/arXiv.2306.00622>.
Hosseini, Mohammad, and Serge P. J. M. Horbach. „Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review“. *Research Integrity and Peer Review* 8, Nr. 1 (18. Mai 2023): 4. <https://doi.org/10.1186/s41073-023-00133-5>.

GenAI im Review-Prozess: Chancen

- bessere Steuerung des Reviews
 - Prüfung von formalen Kriterien
 - qualitative Vorselektion?
- **Effizienz- und Volumensteigerung**

*Through supporting both actors in efficiently writing constructive reports or decision letters, **LLMs can facilitate higher quality review and address issues of review shortage.***

GenAI im Review-Prozess: Risiken

“klassische” Unzulänglichkeiten von LLM wie Halluzinationen, Biases

ethische + rechtliche Fragen:

- **Vertraulichkeit**
- **Vertrauenswürdigkeit**
- **Verantwortung**

*However, the **fundamental opacity** of LLMs’ training data, inner workings, data handling, and development processes raise concerns about **potential biases, confidentiality and the reproducibility** of review reports.*

Gen AI Policies, Beispiel Sage

Für Reviewer:

- **Nicht erlaubt** für die Erstellung von Gutachten

Für Editoren:

- **Nicht erlaubt** Triagierung von Manuskripten, für Zusammenfassungen + Decision Letters
- **Erlaubt** für Reviewer-Suche

*We ask that **Editors ensure the reviewers invited are aware of the confidentiality issues presented by generating a review report using language models or generative AI.***

Editors and reviewers should **evaluate the appropriateness of the use of LLMs...**

SAGE Publications. „ChatGPT and Generative AI“, 27. Januar 2023.
<https://us.sagepub.com/en-us/nam/chatgpt-and-generative-ai-0>.

SAGE Publications. „Using AI in Peer Review and Publishing“, 31. Juli 2023.
<https://uk.sagepub.com/en-gb/eur/using-ai-in-peer-review-and-publishing>.

Herausgeber-Sicht aus Humanities/Bioethics

Editoren

- Verantwortung + Kontrolle der **Reviewer-Auswahl**
- Verantwortung für **automatisiertes Lektorat** liegt bei Autoren + Editoren
- **Zugang zu Tools und Strategien** zur Prüfung auf AI-Content

*Editors and reviewers **should not rely solely** on generative AI to review submitted papers.*

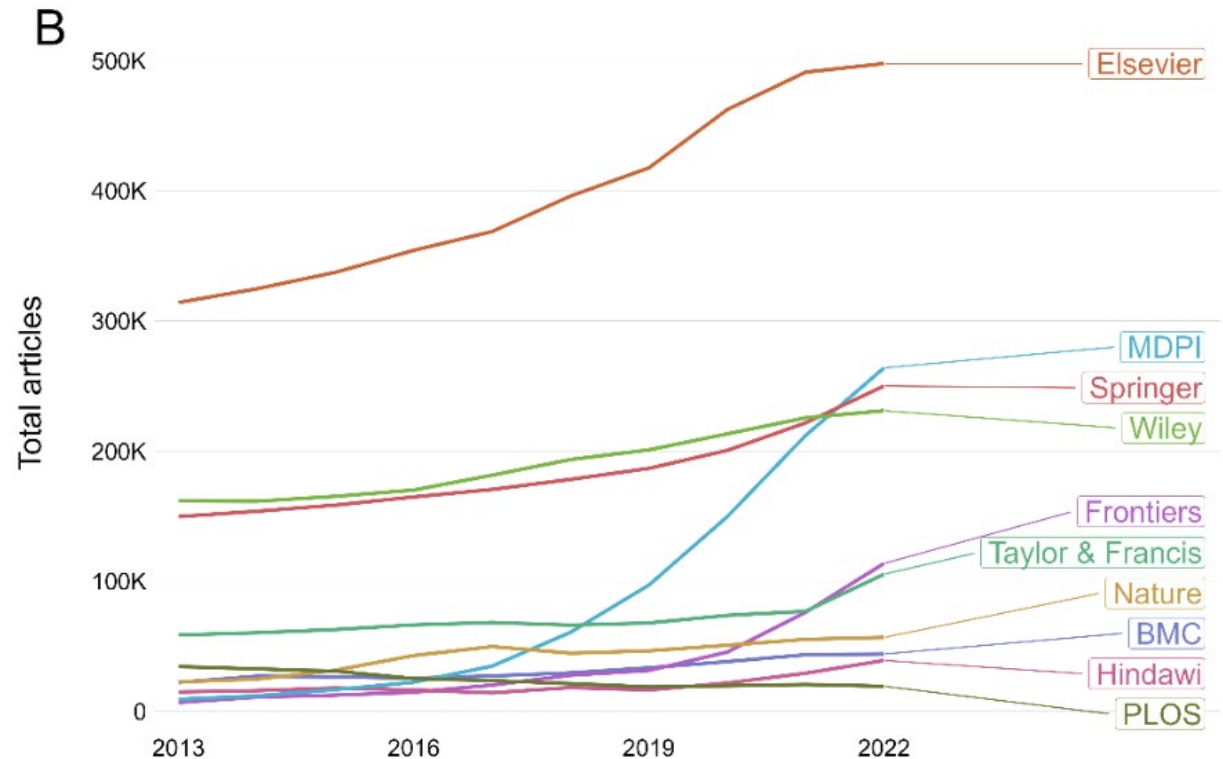
*...**given the current state of development** [...], we do not believe that they are adequate as reviewers.*

*Again, using AI as a decision-support tool may be beneficial and save time. But **replacing this editorial function** with AI seems unwarranted, **except under exceptional circumstances.***

u^b Das Business

WissKomm...ein interessantes Business!

- Exp. Wachstum: + 47% Artikel in WoS/Scopus 2016–2022
- (Fehl-)Anreize: Publish-or-Perish, Metriken, (OA)
- Folgen: enormer Workload, Betrug, Ausnutzung OA-Transformation
- Unternehmenskonzentration, Oligopol
- enorme Gewinnüberschüsse
- Plattform-, Data-Analytics-Unternehmen



Hanson, Mark A., Pablo Gómez Barreiro, Paolo Crosetto, und Dan Brockington. „The strain on scientific publishing“. arXiv, 27. September 2023. <https://doi.org/10.48550/arXiv.2309.15884>.

Produkte für Forschungsevaluation + -steuerung

- Produktivität von Personen, Einheiten
- Analyse der Forschungslandschaft, Trends, Forschungslücken
- Kollaborationen, Funding/Grants, Institutionen-Benchmarking, PPPs
- Daten per APIs, Analyse-Dashboards, (Verknüpfung mit institutionellen) FIS
- descriptive + predictive + prescriptive analytics

What Dimensions API can do for you

- Competitive intelligence
- Horizon-scanning and emerging trends
- Innovation landscape mapping
- Academic and industry partnerships and collaboration networks
- Key Opinion Leader (KOL) identification
- Recruitment and talent
- Performance and benchmarking
- Tracking funding dollar flows and citation patterns
- Literature gap analysis
- Marketing and communication strategy
- Social and economic impact of research

<https://www.dimensions.ai/products/all-products/dimensions-api/>

Überwachung des Informationsverhaltens

First party data = Identifizierung von Individuen

- möglichst personalisierte Zugangsauthentifizierung
- Hashing jedes Downloads = Zuorden- + Rückverfolgbarkeit aller Kopien

Third party data = Verknüpfung und Analyse mit **Drittanbietern**

- Einbindung von AdTech wie Audience-Tools, Finger Printer, Bid Streaming
- Verknüpfung mit „privater“ Webnutzung, z.T. sogar Echtzeit- und Offlinedaten
- Bsp. Nature-Aufsatz: ca. 70 Analyse- und Profiling-Tools von Verlagen + Third Parties

Siems, Renke. „Das Lesen der Anderen: Die Auswirkungen von User Tracking auf Bibliotheken“. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 9, Nr. 1 (15. März 2022): 1–25. <https://doi.org/10.5282/o-bib/5797>.

Petra Gehring. „Das Schicksal von Open Science steht auf dem Spiel“. *Forschung und Lehre*, 2. August 2021. <https://www.forschung-und-lehre.de/politik/das-schicksal-von-open-science-steht-auf-dem-spiel-3902>.

„Nebengeschäfte“

Geschäftsfeld Risk Solutions

- Risiko-Analysen (Kunden u.a. Behörden), Betrugsprävention, Cyber-Sicherheit
- u.a. Portscanning, Bsp.: ThreatMetrix, auch nachgewiesen auf ScienceDirect
- u.a. Verhaltensbiometrie: Identifizierung durch Interaktionsmuster

Bibliotheken

- Library Analytics, Course Resources
- Bsp. Elsevier - VSNU: Hybrid OA ohne Zusatzkosten gegen FIS-Lizenzierungen + MD-Kooperation
- Bsp. Diskussion zur Installation von Spyware auf Bibliotheksseite gegen Rabatte

Siems, Renke. „Das Lesen der Anderen: Die Auswirkungen von User Tracking auf Bibliotheken“. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 9, Nr. 1 (15. März 2022): 1–25. <https://doi.org/10.5282/o-bib/5797>.

Mehta, Gautama. „Proposal to Install Spyware in University Libraries to Protect Copyrights Shocks Academics“. *Coda Story* (blog), 13. November 2020. <https://www.codastory.com/authoritarian-tech/spyware-in-libraries/>.

LexisNexis Risk Solutions. „Mit Verhaltensbiometrie Betrug bekämpfen“. Zugegriffen 30. Oktober 2023. <https://risk.lexisnexis.com/global/de/insights-resources/article/what-is-behavioral-biometrics>.

u^b

„Surveillance publisher“ (Pooley 2022)

*We can call a company a **surveillance publisher** if it derives a substantial proportion of its revenue from **prediction products, fueled by data extracted from researcher behavior.***

- vgl. Surveillance capitalism (S. Zuboff): Abschöpfung von Verhaltensdaten, Verarbeitung und Vermarktung zu Verhaltensdatenprodukten auf neuen Märkten
- „Datenkapital“: multifunktionaler, sich nicht verbrauchender Rohstoff, Netzwerkgut
- To be continued...

Lots of exciting job opportunities for Digital Science's newly formed AI Solutions team! Ping me if interested

- 1) Senior ML Ops - AI Solution Development - <https://lnkd.in/eaJAJKBm>
- 2) Senior API Engineer - AI Solution Development - <https://lnkd.in/eygN2kp7>
- 3) Senior Front End Developer - AI Solution Development - <https://lnkd.in/emcVGeGf>
- 4) API Engineer - AI Solution Development - <https://lnkd.in/e9rH9SGs>
- 5) Senior Site Reliability Engineer - AI Solution Development - <https://lnkd.in/euixZ4NE>

https://www.linkedin.com/posts/michele-pasinsts/michele-pasinsts_0a02231_lots-of-exciting-job-opportunities-for-digital-activity-7120372039688822784-8c4h

u^b Probleme des Data Business

“Garbage in, garbage out” (GIGO)

- Probleme in Ausgangsdaten übertragen sich in Produkte, Analysen, Modelle/Algorithmen!
- Biases bez. Publikationen + Personen (positive bias, Publikationssprache, Gefälle, Matthäus-Effekt etc.)

Schein-Objektivität + Wirkmächtigkeit
Legitimierung

Predictive Scoring als Fortschreibung (Matthäus-Effekt)

Grundproblem: Opazität von Daten und Algorithmen

u^b Neue Herausforderungen

Researchers need to collaborate [...] to develop best practices, standards and detection methods to ensure that the benefits of GenAI can be realized without fundamentally undermining science and its role in society.

Sandra Wachter, Professor of Technology and Regulation,
Oxford Internet Institute

AI ANTICIPATIONS

Q: How useful do you think AI tools are for researchers in your field?

■ Essential ■ Very useful ■ Useful ■ Slightly useful ■ Not at all useful

Respondents who use AI in research



Respondents who don't use AI in research



Q: How useful do you think AI tools will become for researchers in your field in the next decade?

Respondents who use AI in research



Respondents who don't use AI in research



©nature

AI AND SCIENCE: WHAT 1,600 RESEARCHERS THINK

A *Nature* survey finds that scientists are concerned, as well as excited, by the increasing use of artificial-intelligence tools in research.
By Richard Van Noorden and Jeffrey M. Perkel

Think". *Nature* 621, Nr. 7980 (28. SeptemVan Noorden, Richard, und Jeffrey M. Perkel. „AI and Science: What 1,600 Researchers ber 2023): 672–75.
<https://doi.org/10.1038/d41586-023-02980-0>.

u

PROBLEMS OF GENERATIVE AI

Q: Where do you think generative AI may have **negative impacts on research?** (Choose all that apply.)

May proliferate misinformation

Makes plagiarism easier, and harder to detect

May bring mistakes or inaccuracies into research texts (papers, code)

Makes it easier to fabricate or falsify research and harder to detect

May bring biases into literature searches

Makes it harder to assess student learning

May entrench bias or inequities into research texts

Raises energy consumption and carbon footprint of research

Other

0

100%

QUALITY OF AI REVIEW IN RESEARCH PAPERS

Q: Do you think that journal editors and peer-reviewers, in general, can adequately review papers in your field that use AI?

■ Yes ■ No ■ Don't know/cannot tell

Respondents who study AI



Respondents who use AI in research



Respondents who don't use AI in research



0

20

40

60

80

100%

©nature

NEGATIVE IMPACTS OF AI

Q: Considering **machine-learning methods**, what do you think are **negative impacts** of AI in research? (Choose all that apply.)

Leads to more reliance on pattern recognition without understanding

Results can entrench bias or discrimination in data

Makes fraud easier

Ill-considered use leads to irreproducible research

Exacerbates power imbalances: only scientists at well-resourced universities or firms can be at the cutting edge

Expensive or energy-intensive tool

Other

0

100%

Herausgeforderte Wissenschaftskommunikation

Herausforderungen (DFG 2022)

1. schwindende Wahrnehmbarkeit publizierter Wissenschaft
2. Wissenschaftsbewertung auf Grundlage publikatorischer Metriken
3. Entwicklung neuer Mechanismen von Qualitätssicherung und -bewertung
4. ungünstige Marktstrukturen + Geschäftspraktiken, inkl. Datentracking in der Wissenschaft

Datafizierung = Verschärfung der Probleme

Disruption der Praktiken durch GenAI

- Publikationsflut/Betrug, damit auch Metriken
- Entwicklung des Review-Verfahrens?

Boost des Research-Analysis-Business

- Missbrauchspotential persönlicher Daten
- irregeleitete + irreführende Evaluationen

Rollen und Verantwortung von Bibliotheken?

Enabler – Broker: Beschaffung von/Zugang zu Publikations-/Informationsorganen, Daten, Tools

→ Umgang mit neuartigen Daten- und Tool-Produkten, mit „AI-Enhancement“ bestehender Angebote

Berater: insb. hinsichtlich Praktiken Schreiben, Rezipieren, u.a. Auswahl von Organen/Plattformen

→ Förderung neuer Kenntnisse + Fähigkeiten: AI Literacy, Data Ethics, neue wiss. Normen

Hüter von Governance: Empfehlungen vertreten, Regeln durchsetzen

→ Aushandlung neuer Normen und Governance beobachten, geg. mitwirken

→ Lenkungswirkung sowie Verantwortung für unabhängige Information + Schutz der Nutzer

→ **Anforderungen und Verantwortung steigen!**

u^b

Vielen Dank für Ihr Interesse!

Fragen? Anmerkungen? Diskussion!

Kathi Woitas, M.A.

Digital Scholarship Specialist

kathi.woitas@unibe.ch