

Jussara Rowland
(INESC-ID)
jussararowland@inesc-id.pt

Generative AI in Doctoral Research: Practices, Challenges, and Implications

PhD Program in
Climate Change and Sustainable
Development Policies
15/05/2026

1. Generative AI

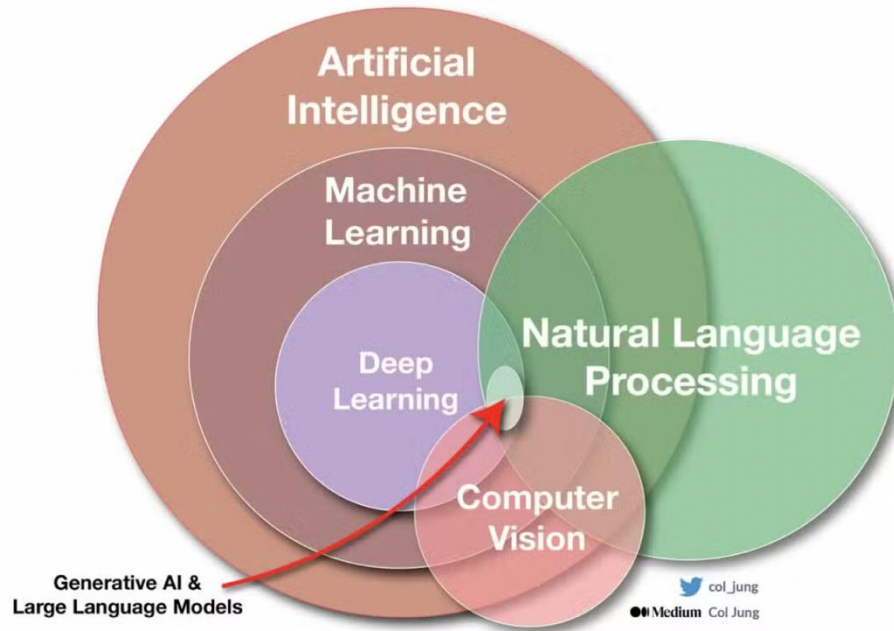
Artificial Intelligence (AI)

The definition of Artificial Intelligence (AI) is complex, and there is no universal consensus. The term "AI" is frequently used inconsistently, influenced by historical factors, market trends, and the specific capabilities of the technology in question.

AI is a broad concept encompassing various technologies and applications. Among the main types of AI are:

- Generative AI, which synthesizes and produces content—such as text, images, or music—by identifying and replicating patterns and styles from large datasets, typically using deep learning models (e.g., ChatGPT and DALL-E).
- Predictive AI, which uses machine learning techniques to analyze historical data and predict future outcomes, identifying patterns to anticipate trends and behaviors. It is widely used in applications such as credit scoring, demand forecasting, and recommendation systems.

Artificial Intelligence (AI)



Narratives

'Each technique [in AI] is *both* a method for designing artifacts and a thematic for narrating its operation'
(Agre, 1997, p. 7)

The language used to describe AI, with anthropomorphic terms such as "reasoning" and "understanding," can obscure its true operational processes.

These discourses tend to prioritize "understanding" over accuracy.

The use of anthropomorphic metaphors can lead to an overestimation of AI's capabilities, obscuring the fact that these are **systems based on mathematical models and statistical analysis**, relying on **vast amounts of data** and **human labor** in their creation and maintenance.

Discourses around AI tend to shape public perception and the socio-political relationships associated with the technology.

Generative AI (GenAI)

Generative AI works by "learning" patterns and features from large datasets. Examples include generative diffusion models for image generation and large language models, such as OpenAI's GPT, Google's Gemini, and Anthropic's Claude .

Training data is mostly sourced from publicly available information on the internet, curated datasets, and user-provided inputs within the tool.

A specific type of generative AI is Large Language Models (LLMs). These models rely on the statistical analysis of language: their goal is to predict the next word with the highest probability. They do not possess independent knowledge.

Access to these functionalities is typically provided through chatbots (e.g., ChatGPT, Bard, Claude), which are designed to interact naturally with users, mimicking human-like communication patterns. Unlike earlier chatbot models based on pre-programmed rules, current models generate dynamic and contextually appropriate responses to user queries (Goodlad & Stone, 2024).

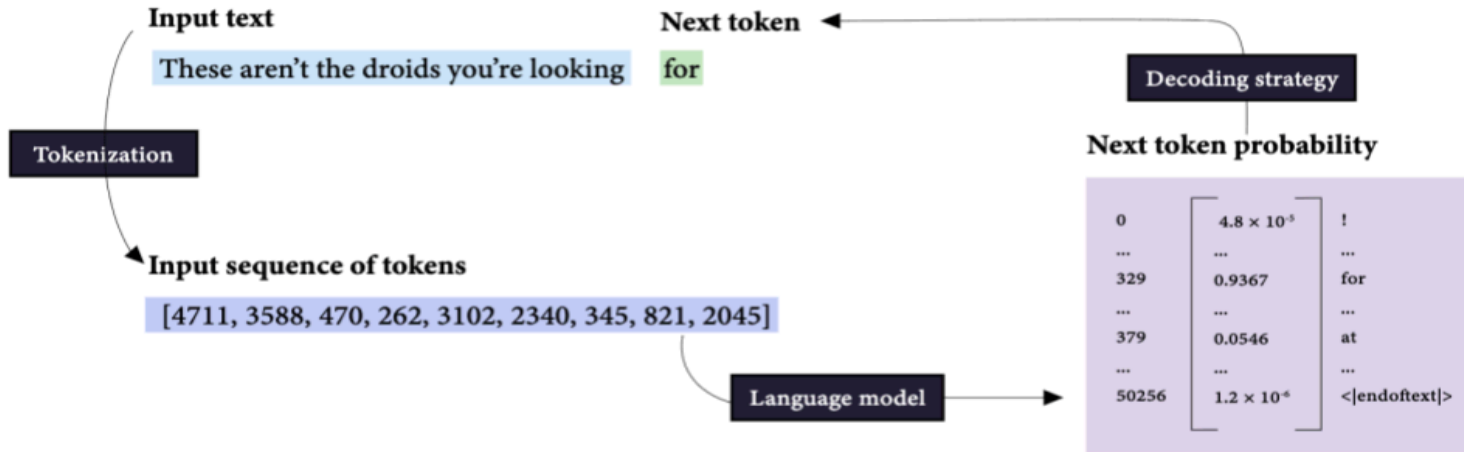
Generative AI has excelled in specific tasks, particularly in text translation, text generation, the creation of coherent images or sounds, automatic text summarization, semantic analysis, opinion detection, as well as text mining and content retrieval. These capabilities stem from its ability to identify complex patterns within large datasets.

LLMs use a process called tokenization, which breaks text into smaller units (tokens). These tokens can be entire words, parts of words, or even individual characters, depending on how the model was trained.

The Transformer architecture, a specific software design, is essential for enabling LLMs to process language efficiently. Through the "attention" mechanism, these models can analyze complex relationships between words in a text, resulting in more coherent and contextually appropriate responses.

Tokenization

LLM tokenization



Generative AI

LLMs work based on probability, not facts.

Rather than “knowing” information, they generate text by predicting likely sequences of tokens based on statistical patterns learned during training. These predictions depend on how frequently and in which contexts words and phrases tend to occur together.

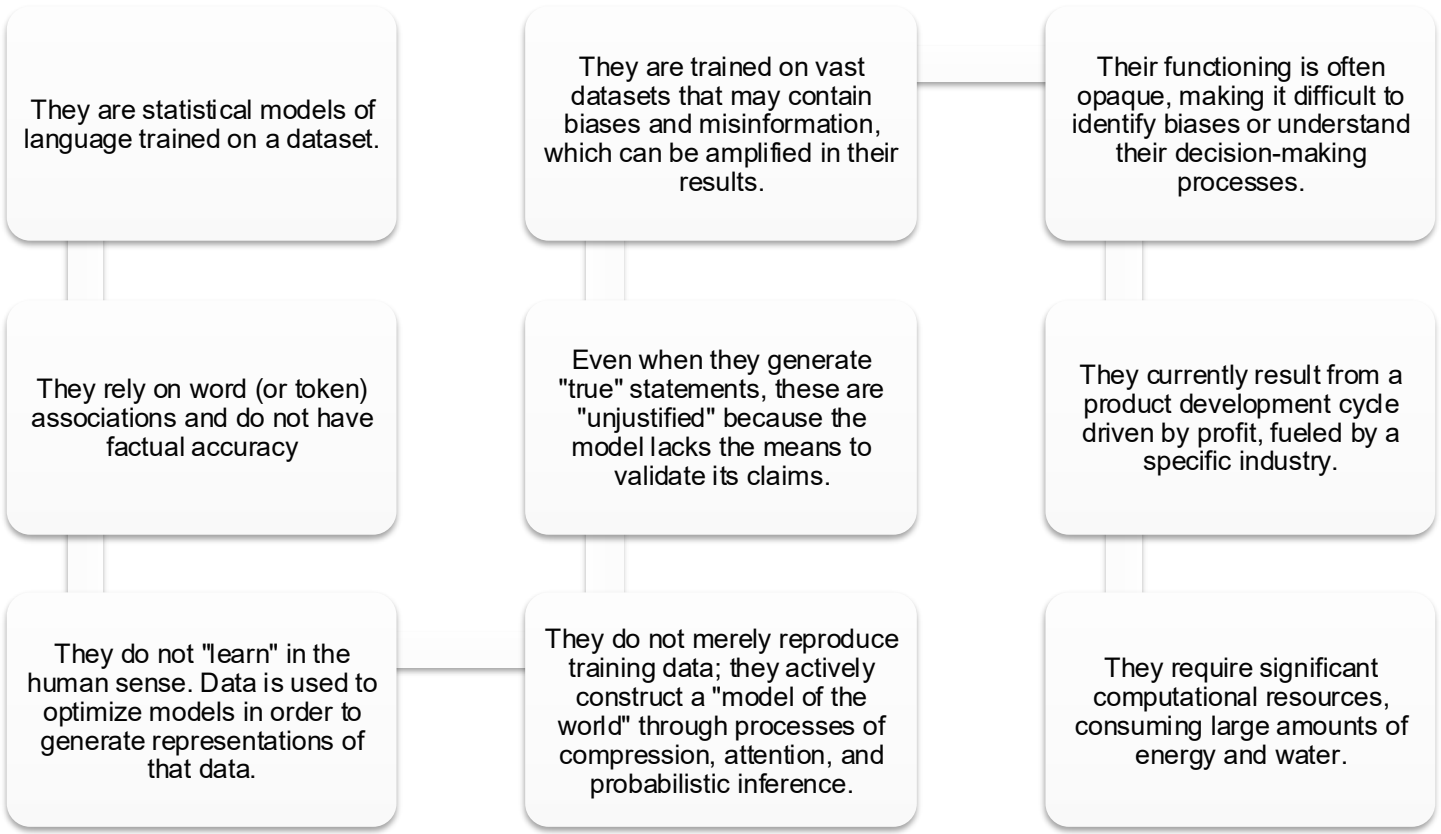
This means LLMs can produce fluent, confident-sounding text that is factually wrong, a phenomenon that has been called "hallucination."

The same question can yield different answers depending on how it is phrased, because the model responds to surface patterns in the input, not to an underlying meaning it has "understood."

There is no internal fact-checker. LLMs have no mechanism to verify claims against reality before generating a response.

Because these patterns are learned from vast and uneven datasets, and shaped by design choices about how information is weighted, they reproduce existing biases (around gender, race, language, and culture) often invisibly and at scale

Large Language Models



Environmental footprint

GenAI requires significant computational resources, far beyond those of a traditional internet search.

This has a material cost: data centers running these models consume large amounts of energy and water.

ARTIFICIAL INTELLIGENCE

AI is poised to drive 160% increase in data center power demand

May 14, 2024 Share <

AI's Energy Demands Are Out of Control. Welcome to the Internet's Hyper-Consumption Era

Generative artificial intelligence tools, now part of the everyday user experience online, are causing stress on local power grids and mass water evaporation.

Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse?

Emissions from in-house data centers of Google, Microsoft, Meta and Apple may be 7.62 times higher than official tally

Green | ESG & Investing

Google Is No Longer Claiming to Be Carbon Neutral

The tech giant, which has seen its planet-warming emissions rise because of artificial intelligence, has stopped buying cheap offsets behind the neutrality claim. The company now aims to reach net-zero carbon by 2030.

TECHNOLOGY

Microsoft's Hypocrisy on AI

Can artificial intelligence really enrich fossil-fuel companies and fight climate change at the same time? The tech giant says yes.

By Karen Hao

It is not a search engine, although it may retrieve and generate information in ways that resemble search.

It does not process information or produce meaning in the same way humans do, despite generating outputs that may appear meaningful.

It can reveal patterns and relationships that are not immediately evident, which can be useful and lead to new discoveries.

It can personalize user experiences, potentially promoting inclusion and accessibility.

It creates a "model of the world" (Amoore et al., 2024) based on statistical models, which both reflects and influences how we perceive, classify, and act upon that world.

It is associated with various risks and negative impacts, such as the amplification of biases, copyright and privacy violations, environmental impact, labor exploitation, and the degradation of information quality (Critical AI Explainer, 2024).

It also raises concerns about power concentration, opacity, and accountability in the development and deployment of these systems.

2. AI and Higher Education

AI and Higher Education

Generative AI, especially LLMs, has brought new challenges to education, including university and postgraduate studies.

- Academic Integrity Issues: AI raises concerns about authorship and plagiarism, requiring a rethink of assessment methods and ethics.
- Loss of Skills: Relying on AI can weaken critical thinking and original writing, leading to shallow learning.
- Business Interests vs. Education: AI automation often prioritizes efficiency and profit over quality education and meaningful teacher-student interactions.



Narratives and marketing

The use of GenAI in education has led to contradictory narratives:

- Some universities completely ban its use and rely on (problematic) "AI detectors"...
- While others actively encourage its use at various levels.

At the same time, we are seeing:

- The growing integration of AI features into existing digital tools.
- The launch and promotion of new tools and features specifically designed for teachers and students.



Principles and Recommendations ICS-Ulisboa

1. Recommendations for Students and Doctoral Researchers
 - Teaching, learning, and research assistance: support understanding of complex topics without replacing mastery of core concepts or critical thinking
 - Writing and translation assistants: support structuring, editing, and translating without replacing the student's own writing and research plan
 - Resources for brainstorming and creative process support: help generate initial ideas and approaches, but AI suggestions must always be critically evaluated by the student
2. Recommendations for teaching and assessment
3. Supervisor/supervisee relationship
4. General recommendations

<https://www.ics.ulisboa.pt/sites/ics.ulisboa.pt/files/basicpage/ficheiros/principios-e-recomendacoes-sobre-o-uso-de-ia-generativa-en.pdf>

Rules and guidelines

Reflexão e Guia para a utilização da Inteligência Artificial no Instituto Superior Técnico

- **Tutoring, not authorship:** AI output must be reviewed and verified; never taken at face value
- **Data privacy:** keep confidential and sensitive data out of AI prompts
- **Assessment:** favour oral presentations and group discussions; ask students to articulate their own original contribution
- **Declaration of use:** when AI is permitted, students must explain how and why they used it
- **Three permission levels:** prohibited / limited with declaration / fully permitted with declaration
- **The bottom line:** AI is not a shortcut; critical thinking, originality, and learning from mistakes cannot be outsourced











AI POLICIES

open archive of the social sciences, provides a free, non-profit, open access platform for social scientists to upload working papers, preprints, and published papers, with the option to link data and code. SocArXiv is dedicated to opening up social science, to reach more people more effectively, to improve research, and build the future of scholarly communication.



- There are acceptable uses for LLMs and other machine-assistance tools in research on our site, as long as they are disclosed and documented. Failure to disclose, or implausible disclosures, are grounds for rejection. These are acceptable:
 - Language translation (we may ask for the original to review)
 - Pre-writing work such as literature searches, idea generation and organization, when the paper also reports on original research and the author attests to thorough human supervision (e.g., verifying sources).
 - Copy-editing and formatting
 - Machine-assisted content or data analysis
 - Dictation software
- These are unacceptable uses:
 - Generating text which is used verbatim (including whole paragraphs and sections)
 - Generating fake human subjects data (including simulated, *in silico* samples)
 - LLM as co-authors or interlocutors as if human (e.g., “interviewing” or “dialogue” with LLMs)
 - Generating false information to mislead moderators (or readers)
 - Submitting AI-generated content which the author has not thoroughly reviewed and confirmed (this includes reviewing the cited sources to confirm they exist and are accurately characterized, and verifying images)
 - Entire papers produced by AI generation with no human-generated components beyond prompts

Resources

	
Resources for Teaching and Generative AI DESIGN JUSTICE LABS	
NOME	ATUALIZADO ↓
 Selections from Special Issue: Generative AI and the Rise of Chat...	4 de set. de 2024 por Lauren Go...
 Teaching and Gen AI 9-16-2024.pptx	16 de set. de 2024 por Lauren G...
 Critical AI student guide 9-7 reformatted.pdf	7 de set. de 2024 por Lauren Go...
 Teach your students to audit.pptx	30 de ago. de 2024 por Daniel E...
 The How and Why of Generative AI .pptx	27 de ago. de 2024 por Matthe...
 Critical AI Literacies in the Language Classroom.pptx	27 de ago. de 2024 por Chloa KI...
 Fostering AI Literacies through Simple Probing and Research Exer...	27 de ago. de 2024 por Teresa R...

Critical AI Literacies: A Guide for Students

This publication was prepared by Melissa K. High, Rutgers as part of a grant awarded to her by the National Endowment for the Humanities. It is not an official Rutgers University publication and should not be used to represent Rutgers University or its policies. For more information, see <https://www.rutgers.edu/ai-literacies>.



"Generative" artificial intelligence (AI) tools are now embedded in popular software, tempting you to regard these commercial applications as reliable "copilots" and tutors. While these tools generate content in seconds, they rely on immense troves of human-generated "training data" and the energy-intensive computing resources necessary to statistically "model" it. Generative AI is the product of humans' work, including the centuries of writing used in training (often without consent); the work of technologists in fields like computer science and linguistics; and the poorly compensated labor of human annotators whose ongoing work is necessary to make such "AI" seem more reliable and human-like.

Generative AI leverages a data-driven technique called "machine learning"—a process of statistical pattern-finding through which models "learn" to "optimize" for useful predictions. (They do this by updating the weights in a statistical message-passing "architecture".) Although the human-like text and images these systems produce can be hard to distinguish from content created by people, machine learning systems are disembodied statistical models that do not understand language or images in the way humans do. Machine "intelligence" excels at pattern-matching at vast scales, enabling generative AI tools to work probabilistically—by predicting the most likely responses to a user's prompt. However, generative AI tools are incapable of anything like a full-bodied "experience", as disembodied statistical models they have no access to the world beyond a user's inputs, no ability to update themselves between rounds of training, and (for all these reasons) no means of assuring the truthfulness of their outputs. That is why generative AI requires so much oversight from human workers, including users. *These systems should not be confused with search engines and cannot be trusted to generate accurate writing without considerable proofreading and fact-checking.*

Discussing the ethics of generative AI (and "AI" more generally), is a complicated subject that you should reflect on after getting the facts (see below). Many people believe that it is impossible to use generative tools ethically. The underlying models on which generative AI is built recapitulate biases and stereotypes; infringe on copyright protections; surveil users (including students); leak data; expend enormous amounts of energy, water, and investment; and concentrate enormous power and resources in the hands of a tiny elite. The authors of this document, in accordance with Kathryn Conrad's "[Blueprint](#)" for AI rights, recommend that all students be given the right to opt out of any assignment that involves using such tools, even if no subscription is required. At the same time, since generative applications are now widely available, it is important to understand how they work and, for those who wish to use them, to learn how to avoid the most concerning academic and intellectual pitfalls. These may include the violation of your institution's code of conduct and/or of your instructor's course policies, or accusations of academic misconduct (which may result because of faulty AI "detection" tools). Above all, students who rely on chatbots may lose crucial opportunities to develop the foundational skills in writing, research, and critical thinking that make higher education valuable in the first place. Building critical AI literacies is a process of empowerment that enables students and citizens to exercise independent judgment about whether or if to use this very new and largely untested commercial technology.

<https://rutgers.app.box.com/s/v1ywq3701cb6pvnewz5ptuhojwppmsg>

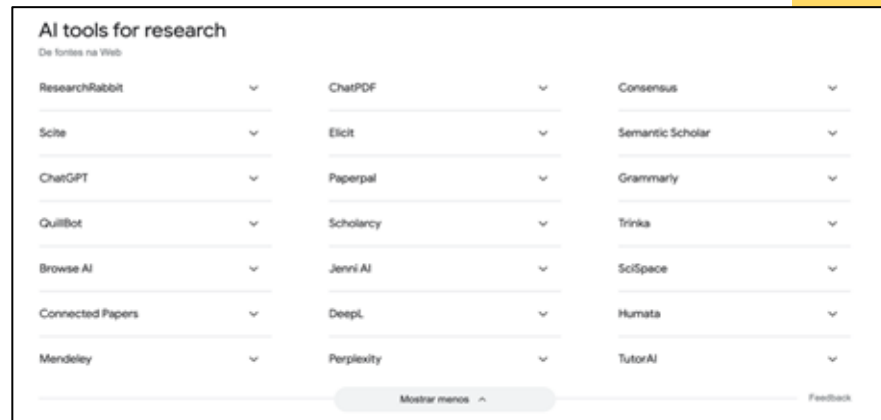
3. AI in Scientific Research

Growing number of AI applications for research

The launch of an increasing number of AI applications for research.

AI systems and applications are not limited to specific functions or fields; their expansion is ongoing, integrating into various tools and stages of scientific research.

AI is present in widely used tools like ChatGPT, as well as in research-specific platforms such as Scite.ai, Scholarcy, Perplexity, Paperpal, ResearchRabbit, and others.



AI technology is also integrated into everyday tools, such as word processors, as well as specialized software, including qualitative analysis programs like MaxQDA, NVivo, and Atlas.ti.

Growing Role in the Social Sciences



The start of the registration for the advanced training program
"Artificial Intelligence in Social Research"



Course start: **may 21, 2024**

LSE Department of Methodology


Generative AI in Social Science Research

Join colleagues from the Department of Methodology at LSE for a full day event showcasing cutting edge research on, and using, generative AI in social science research from internationally-renowned scholars at LSE, Oxford, Princeton, Brigham Young, and Tübingen.

Keynote speech from Prof. Arthur Spirling (Princeton University)

Sign up on Eventbrite!

LSE Marshall Building
Friday 7 June 2024



Call for papers

GENERATIVE METHODS

AI as collaborator and companion in the social sciences and humanities

Conference dates: **December 6-8, 2023**
Location: Copenhagen, Aalborg University Campus

The advent of generative AI, such as large language models, chat bots, and text-to-image generation, has presented both opportunities and challenges for society. This conference seeks to explore the multifaceted impact of generative AI as instruments and objects of research. We invite contributions that demonstrate, speculate, and/or discuss the ways in which generative machine learning models will transform research methodologies and reshape research problems across SSH disciplines. How will generative models impact our understanding of data, analysis, and research ethics? What novel forms of research communication and stakeholder engagement could emerge? What technical and conceptual interventions will be necessary to repurpose them for different epistemic traditions? And how do we audit or establish rapport with them?

We encourage researchers and practitioners to submit their contributions in two formats:

Papers We invite abstracts for papers that address the conference theme. Abstracts should summarize the main argument and not exceed 200 words.	Generative experiments We also welcome proposals for experiments with generative AI that can be exhibited at the conference. Proposals should pitch the concept, describe the practical requirements, and not exceed 200 words.
---	---

Important dates and information

- Call for papers and experiments closes August 15 (acceptance by August 31)
- Registration opens September 1 and closes October 1 - participation is free
- The conference is organized by MASSHINE, AAU's Hub for Computational SSH
- Submission, registration, and more info: www.ssh.aau.dk/MASSHINE23

MASSHINE

Growing Role in the Social Sciences

25, 26, 27, 29
May 2026
4 pm - 7 pm



Summer School
**Exploring Data,
Networks and
Narratives**
Computational methodologies for advanced
textual data analytics



Methods Seminar

A series of seminars focused on advanced methodological approaches, particularly in Natural Language Processing (NLP), Artificial Intelligence (AI), video and image analysis, and multimodal analysis.

Speaker



Date: May 28th, 11:00H CET
University of Namur - EMCP faculty (E11) & Online

Computational approaches to meaning change

Barbara McGillivray
Senior Lecturer in Digital and Computational Humanities at King's College London

In this seminar, we will address:

- ✓ Semantic change across temporal, cultural, and textual dimensions
- ✓ Computational approaches for analysing semantic change in historical texts
- ✓ Combination of computational and human-centred approaches

Coordinators:
Nelson Santos
Auguste Debroise
Jérémy Dodgeine
Guilhem Cassan



To receive all the details:
<https://marketing.ego1.page/1e2e84ny/signup>

AI tools can be leveraged for various purposes throughout the research process, including:

- Idea development and research design
- Content development and structuring
- Literature review and synthesis
- Data management and analysis
- Support for editing, review, and publication
- Communication, dissemination, and ethical compliance

They are often promoted as a means to enhance productivity and objectivity (Messeri & Crockett, 2024).

(Critical) perspective on AI in Research (based on a review of the current literature)

Table 1 | Visions of AI across the research pipeline

Vision	Research stage	Limits to overcome	Vision
AI as Oracle	Study design	There is too much literature to digest; scientific publications vary in quality; readers are biased; too many research paths to choose from	Tools that objectively and efficiently search, evaluate and summarize scientific literature and generate new hypotheses
AI as Surrogate	Data collection	Data are too difficult, time consuming or expensive to obtain	Tools that accurately and tractably generate surrogate data points from natural complex systems, including human participants
AI as Quant	Data analysis	Data are too large or complex to curate and analyse	Tools that surpass the limits of human intellect in curating and analysing vast and complex datasets to produce new knowledge
AI as Arbiter	Peer review	There are too many papers and proposals to review; reviewers are biased	Tools that objectively and efficiently evaluate scientific merit and the replicability of findings

Fonte: Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002), p.53.

Risks Associated with Generative AI

- Potential generation of factually incorrect information (false answers, "hallucinations," and bias)
- Possible disclosure of sensitive data and confidentiality risks
- Lack of references for provided information and potential violation of intellectual property and copyright
- Limited ability to provide the same answer to identical questions
- Potential replication of results
- Variable stability and reliability of Generative AI models for critical and time-sensitive processes
- Amplification of cognitive biases

Data Protection and GenAI

The use of GenAI tools raises important concerns regarding data protection, confidentiality and compliance with ethical and legal frameworks.

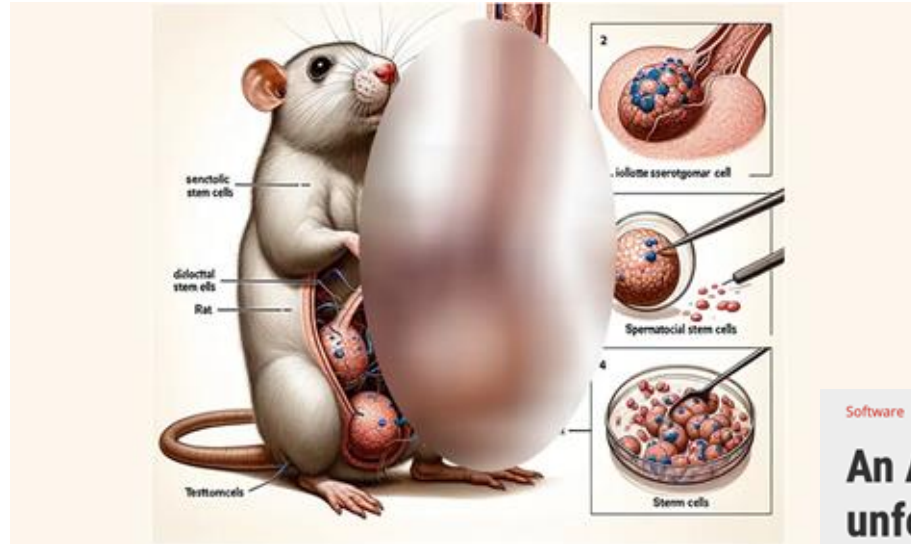
These issues are particularly relevant in research involving human participants and sensitive materials.

1. GenAI systems may store or reuse user inputs, depending on platform policies.
2. There is limited transparency about how data is retained, processed, and shared by commercial providers.
3. Sharing personal or sensitive data in cloud-based models can lead to GDPR violations.
4. Research materials such as interview transcripts, fieldnotes, draft papers or internal documents must be handled with particular care.
5. Responsible practice requires a cautious, critical approach to what is shared, how it is processed, and which tools are used.

See: Utilização de inteligência artificial (IA) em gestão de dados e investigação (Centro para a Gestão de Dados de Investigação iRe: Search da Ulisboa)

Impacts of Generative AI in Academia

Misuse/Inappropriate Use:



Software > AI

An AI-generated image of a rat with unfeasibly large genitals made it into a peer-reviewed article, along with the caption 'dck'

Impacts of Generative AI in Academia

Misuse/Inappropriate Use:

ChatGPT "fingerprints"

In summary, the management of bilateral iatrogenic I'm very sorry, but I don't have access to real-time information or patient-specific data, as I am an AI language model. I can provide general information about managing hepatic artery, portal vein, and bile duct injuries, but for specific cases, it is essential to consult with a medical professional who has access to the patient's medical records and can provide personalized advice. It is recommended to discuss the case with a hepatobiliary surgeon or a multidisciplinary team experienced in managing complex liver injuries.

Conclusion

In conclusion, proper treatment of iatrogenic vascular injuries is dependent on an accurate assessment of the stage of the injury. The injury should be recognized quickly. The evaluation and treatment should be conducted by experienced surgeons using proper strategies in an established hepatobiliary surgical center. Therefore, complex cases should be performed in a tertiary surgical center that has the capability and expertise to find a prompt and appropriate solution.

Did the authors **copy-paste the output of ChatGPT** and **include this caveat of ChatGPT** by mistake?

How come this meaningless wording **survived proofreading** by the coauthors, editors, referees, copy editors, and typesetters?



The phrase "Certainly! Here is..." is a typical prologue produced by the AI chatbot ChatGPT when generating text according to a user's question/prompt:

1. Introduction

Certainly, here is a possible introduction for your topic: Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can affect the safety and life of the battery [3-9]. Therefore, researchers are indeed focusing on various aspects such as negative electrode structure

Impact of Generative AI on Academic Literature: Introduction of Linguistic Biases

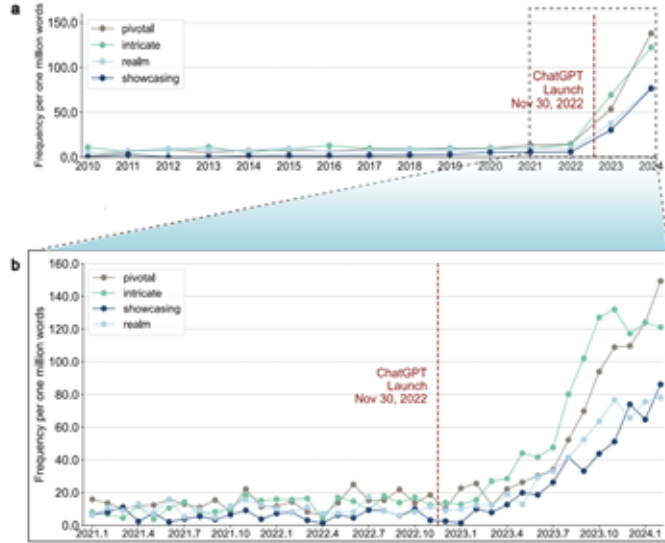
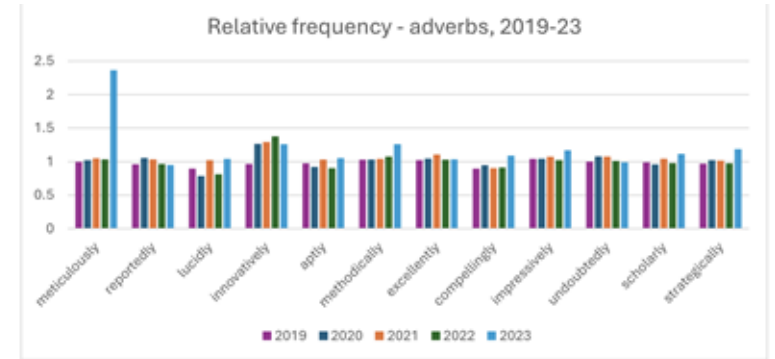
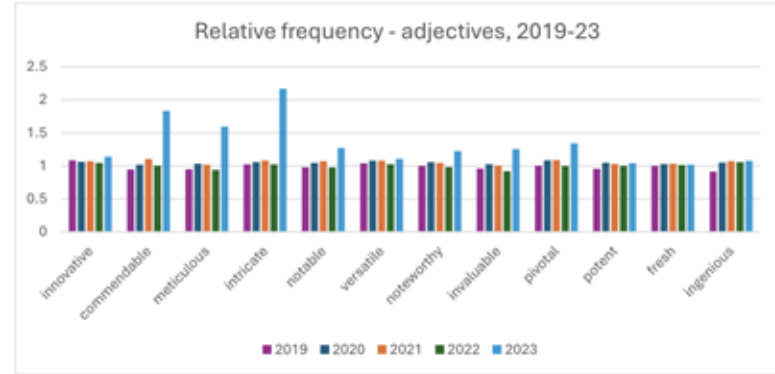


Figure 2: **Word Frequency Shift in arXiv Computer Science abstracts over 14 years (2010-2024).** The plot shows the frequency over time for the top 4 words most disproportionately used by LLM compared to humans, as measured by the log odds ratio. The words are: *realm*, *intricate*, *showcasing*, *pivotal*. These terms maintained a consistently low frequency in arXiv CS abstracts over more than a decade (2010–2022) but experienced a sudden surge in usage starting in 2023.



Impact of Generative AI on Academic Literature: Zombie Citations

Zombie citations are fake or non-existent references that are increasingly produced and amplified by generative AI.

The Haunting of Academia: Unmasking Ghost References

What is a Ghost Reference?
Ghost Reference: A citation to a work that does not exist.
This is distinct from 'citation verbiage', where a real paper is cited but doesn't support the claim being made.

It's an Old Problem, Not a New AI Flaw
 Citations reference → 2019 → 2021
Predates Widespread ChatGPT

The Human Origin of Ghosts
 Careless Copying
 Citing Without Reading

The Core Vulnerability: How Ghosts Materialize

- Vulnerability 1: The 'Citogenesis' Feedback Loop**
 - Fake reference published online
 - Discovered by a human and an AI
 - AI
 - More citations and making ghost appear legitimate.
- Vulnerability 2: Google Scholar's '[CITATION]' Records Create a for unattached references.**
- '[CITATION]' Stubs Give Ghosts Legitimacy**
 - Once a ghost gets a '[CITATION]' record and citations, it's hard to distinguish from a real but hard-to-find paper.

The AI Factor: Amplifier vs. Solution

General AI (RAG + Web Search) Fails
 General AI (RAG + Web Search)
 Search of web data (full of fake data and ghosts)
 Fodder by pre-existing pollution of ghost references, perpetuating the problem.

Academic RAG Systems Succeed
 Specialized AI Tool
 Clean & curated Database (OpenAlex, Pubmed)
 How Academic RAG Guarantees Real Citations

PROCESS FLOW

- Assign a unique ID to every document chunk.
- Pre-train LLM to generate text using ONLY the unique ID.
- Verify ID is real and retrieved.
- Programmatically replace ID with full context citation.

A Shared Responsibility: Exorcising the Ghosts

- For Researchers & Librarians**
 Treat Google Scholar '[CITATION]' records with suspicion. Always verify source existence through multiple authoritative channels before citing.
- For Tool & Index Developers**
 A key design choice for integrity is to NOT create public searchable records for unattached references.
- For Publishers & Editors**
 Peer review and editorial process is the last line of defense. More robust reference verification is needed.

© Notionist/LLM

Google Scholar

"education governance and datafication"

Articles 3 results (0.03 sec)

Any time
 Since 2026
 Since 2025
 Since 2022
 Custom range...
 2022 — 2023
 Search

[CITATION] Education governance and datafication
 B Williamson, N Piattoeva - Education and Information Technologies, 2022
 ☆ Save 📄 Cite Cited by 65 Related articles

[CITATION] Education governance and datafication: Critical perspectives
 B Williamson, N Piattoeva - Bloomsbury Academic, 2023
 ☆ Save 📄 Cite Cited by 3 Related articles

[CITATION] Education governance and datafication: Reimagining learning, surveillance and automation
 B Williamson - Critical Studies in Education, 2022
 ☆ Save 📄 Cite Cited by 2 Related articles

Sort by relevance
 Sort by date

Fonte: <https://codeactsineducation.wordpress.com/2026/01/30/tracing-the-social-half-life-of-a-zombie-citation/>

Impact of Generative AI : Historic information

1. History becomes "written by data prevalence" rather than verified sources. Whatever dominated training data is presented as authoritative fact.
2. LLMs fabricate content: invented battles, false narratives, events misdated by decades.
3. Language bias shapes perspective: Vietnamese-language queries about the Vietnam War return American viewpoints, due to English-dominated training data.
4. Students treat AI as neutral and objective, unlike social media, which lowers critical resistance to misinformation.
5. The risk is generational: each year without intervention, another cohort builds foundational historical knowledge shaped by algorithmic distortion rather than verified sources.

KEOUGH SCHOOL OF GLOBAL AFFAIRS | **Lucy Family Institute for Data & Society** | **Policy Brief** January 27, 2025

How Artificial Intelligence Can Change Our Past: Urgent Action on AI Distortion of Historical Information

By Nina Martin, Andre' Searl, Peter Coker, James (Jack) Miller, Steven Reinhardt, Miguel Cardona, Tomislav Stokich

Introduction

Large language models (LLMs) are a type of artificial intelligence (AI) model built on massive amounts of textual data, capable of generating human-like content, performing translation, summarization, and question-answering tasks on a variety of domains.¹ These systems have emerged as transformative technologies that fundamentally alter how society processes, creates, and consumes information. Recent research reveals that LLMs are being increasingly incorporated into many domains, “whether to trend toward entirely the accuracy of LLM responses when provided with default explanations. The rapid integration of LLMs into information ecosystems poses substantial and immediate challenges for maintaining accuracy and reliability in knowledge dissemination. An explicit focus on the capabilities of large language models and the broad accessibility of tools powered by this technology has led faculty members and centers regarding their use.”² If we fail to systematically evaluate if we “reput on historical information becomes increasingly urgent. The widespread adoption of these systems as authoritative sources—particularly in educational contexts—raises fundamental questions about information integrity and the preservation of accurate historical understanding in democratic societies.

Findings: Problems Requiring Action

Can Large Language Models (LLMs) change our past? Until recently, historians were the only programmatic producers and consumers of history. That is no longer true. As the national research team analyzed how ChatGPT presents historical content about the US-Vietnam War, the Yugoslav Wars, the Portuguese Colonial Wars, and what is commonly referred to as

Summary

Students are rapidly transferring low-quality, unvetted, and often fabricated information. Their widespread use—especially in education—is raising concerns about accuracy, reliability, and the preservation of historical integrity.

How artificial intelligence (AI) systems fabricate information, replicate problems seen in social media, and distort historical facts through data preferences and language manipulation capabilities.

Without immediate action, future generations might be born into an era where the historical record is distorted and unreliable. This report provides a framework for policymakers to address these challenges and ensure the integrity of historical information.

- **partially possible** developing open-source, community-driven alternatives to proprietary LLMs to ensure transparency and accountability in historical content.
- **partially possible** establishing an independent review board to monitor and address inaccuracies and biases in AI-generated historical content.
- **partially possible** creating a standards body to establish minimum requirements for AI-generated historical content, including transparency, accuracy, and accountability.

The following is a list of the authors' research findings. A full report based on a pre-determined set of research questions was developed to build on the work of the team. ChatGPT was used to generate these findings. This included inquiries on the events mentioned above, and the current situation of these conflicts, building narratives from what is known and the sources. The application

This policy brief is part of the Through-Orchard Policy Brief Series and was published in partnership with the Lucy Family Institute for Data & Society.

Epistemic risks

Current perspectives on AI in research are linked to a range of epistemic risks, potentially leading us to believe we know more about the world than we actually do.

- Promise: Enhancing productivity and objectivity by overcoming human limitations.
- Risk: By obscuring the processes through which certain methods, questions, and perspectives come to dominate over alternative approaches, AI may make science less innovative and more prone to errors.

These risks will persist even as AI technologies become more advanced.

Epistemic risks

Illusion of Explanatory Depth:

Scientists believe they deeply understand a phenomenon modeled by AI, when in reality, their understanding is superficial.

Illusion of Exploratory Breadth:

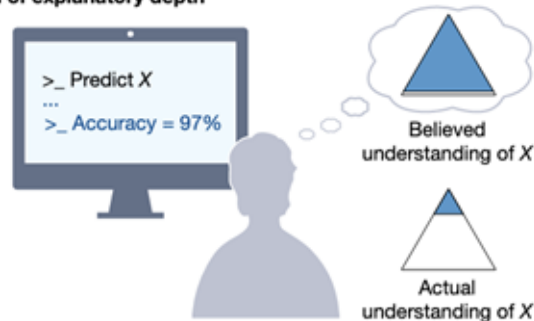
The mistaken perception that all testable hypotheses are being explored, when in fact, the exploration is limited to those testable only with AI tools.

Illusion of Objectivity:

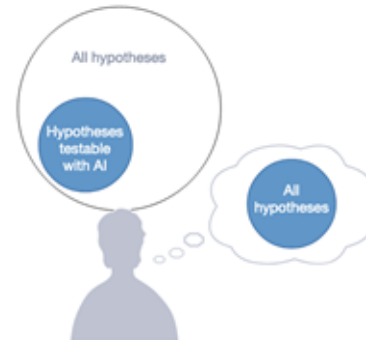
The false belief that AI tools are neutral or capable of representing all perspectives, ignoring biases in training data and the developers' influence.

Fonte: Messeri & Crockett (2024)

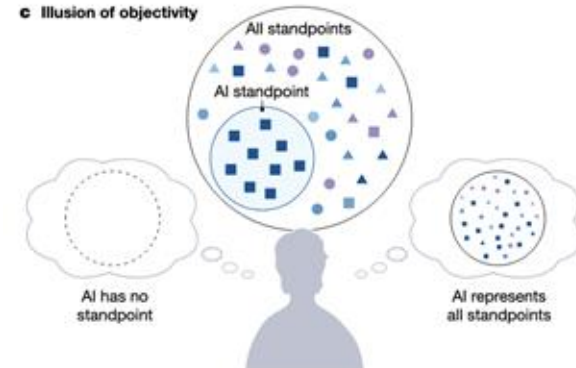
a Illusion of explanatory depth



b Illusion of exploratory breadth



c Illusion of objectivity



4. A reflexive approach to the use of AI

Reflexive use of digital tools in academia

Importance of Reflecting on the Relationship Between Technologies and Methodological Practices (Paulus & Lester, 2023)

1. Risk of viewing humans as passive elements who must simply adapt to technological changes.
2. Risk of seeing tools as purely instrumental, assigning them a passive and neutral role in the research process.

Need to consider both positive and negative consequences of integrating a given technology into the research process.

Reflexive use of AI tools in research

This type of reflection is more common at the data collection stage (e.g., online interviews, online surveys) and data analysis (e.g., the use of QDAS), but it has implications for all phases of the research process.

The integration of new AI features into the tools we use for research makes this issue even more relevant, requiring additional reflection on the implications of digital technologies and automation in our research practices.

Reflexive use of AI tools in research

The use of AI requires intentional and conscious engagement, not only with the technology itself but also with our research practices. It is essential to develop critical awareness of the assumptions we make and the methodologies we use, questioning how AI tools influence both the research process and the results obtained.

The integration of AI into academia encourages us to be increasingly aware of the role and value of our academic practices (Marc Carrigan, 2024) - from writing and literature review to the methods we employ and the disciplinary and epistemological perspectives that shape our work.

Reflexive Use of GenAI in Academic Practices

The incorporation of AI tools into academic practices requires a reflexive approach (Paulus & Lester, 2023), one that goes beyond mere technical use and includes a critical examination of ethical, social, and epistemological implications.

Critical AI literacy is essential for researchers and students to understand the capabilities and limitations of these technologies, assessing the impact of biases and the choices made during AI model development.

Doctoral students should reflect on how AI may affect their projects and impact their ability to conduct original and creative work, influencing their learning processes as well as their capacity for critical analysis and interpretation.

Example 1: writing

Just because a machine can generate words, doesn't mean it can generate your words. In fact, the words it generates may just belong to someone else. **Jane Rosenzweig, 2024**

Rule #1: Understand How LLMs Work

Understanding how LLMs work can lead to more critical, informed discussions about using AI in writing. When considering outsourcing your writing or editing to generative AI, it's important to understand both what is happening and the limitations.

Rule #2: Recognize When Writing is Thinking

Writing is an important way of thinking through a problem, question, or idea. Make sure you know what you might be missing by not doing the writing yourself.

Rule #3: Use Writing Feedback Carefully

While chatbots can provide useful writing feedback, they can also offer poor or irrelevant advice. To use chatbot feedback effectively, you need to clearly understand your writing goals and be able to discern good advice from bad.

Rule #4: Resist the Idea That Outsourcing Your Thinking is Inevitable

Writing remains a valuable tool for thinking and learning. It helps you figure things out—when there's value in the process, keep doing it yourself.

Example 2: content analysis

Generative Artificial Intelligence in Qualitative Data Analysis: Analyzing—Or Just Chatting?

Organizational Research Methods

1-37

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10944281251377154

journals.sagepub.com/home/orm



Duc Cuong Nguyen¹ and Catherine Welch²

Abstract

Researchers, engineers, and entrepreneurs are enthusiastically exploring and promoting ways to apply generative artificial intelligence (GenAI) tools to qualitative data analysis. From promises of automated coding and thematic analysis to functioning as a virtual research assistant that supports researchers in diverse interpretive and analytical tasks, the potential applications of GenAI in qualitative research appear vast. In this paper, we take a step back and ask what sort of technological artifact is GenAI and evaluate whether it is appropriate for qualitative data analysis. We provide an accessible, technologically informed analysis of GenAI, specifically large language models (LLMs), and put to the test the claimed transformative potential of using GenAI in qualitative data analysis. Our evaluation illustrates significant shortcomings that, if the technology is adopted uncritically by management researchers, will introduce unacceptable epistemic risks. We explore these epistemic risks and emphasize that the essence of qualitative data analysis lies in the interpretation of meaning, an inherently human capability.

Keywords

qualitative data analysis, scientific tools, generative artificial intelligence, epistemic risks

We reject the use of generative artificial intelligence for reflexive qualitative research

6 Pages · Posted: 29 Oct 2025

Tanisha Jowsey

Bond University

Virginia Braun

affiliation not provided to SSRN

Victoria Clarke

University of the West of England (UWE) - Department of Health and Social Sciences

Deborah Lupton

University of New South Wales (UNSW) - Faculty of Arts and Social Sciences

Michelle Fine

Carleton University, Ottawa, Canada

Date Written: October 20, 2025

Abstract

We write as 416 experienced qualitative researchers from 38 countries, to reject the use of generative artificial intelligence (GenAI) applications for Big Q Qualitative approaches, such as reflexive thematic analysis, or various phenomenological approaches.

Keywords: generative artificial intelligence, reflexive, qualitative, justice, methodological, human

Declaration of Interest

The authors declare no conflicts of interest.

Ethics Approval

This is a commentary.

Funder Statement

None

Suggested Citation:

Jowsey, Tanisha and Braun, Virginia and Clarke, Victoria and Clarke, Victoria and Lupton, Deborah and Fine, Michelle, We reject the use of generative artificial intelligence for reflexive qualitative research (October 20, 2025). Available at SSRN: <https://ssrn.com/abstract=5676462> or <http://dx.doi.org/10.2139/ssrn.5676462>

Method and Protocol



AI as a Co-researcher in the Qualitative Research Workflow: Transforming Human-AI Collaboration

António Pedro Costa ¹, Grzegorz Bryda ², Prokopis A. Christou ³, and Judita Kasperuniene ⁴

Abstract

The integration of artificial intelligence (AI) into qualitative research is shifting the boundaries of knowledge production. Instead of mere tools, AI is increasingly conceptualized as a co-researcher or research partner—an epistemic agent that helps in the analysis, interpretation, and reflexivity of data. This article proposes an abductive model, called AbductivAI, for analyzing qualitative data, based on collaboration between humans and AI agents. It explores how AI functions as an analytical partner, reflexive collaborator, and participant in distributed cognition within the framework of the human-AI qualitative research workflow and data analysis approach. To build the model, we used advanced prompting techniques (Chain-of-Prompting) and used the 323 abstracts submitted to the World Conference on Qualitative Research as data. Although this data was not intended for research, its exploration allowed us to test the AbductivAI model, showing that interactions between humans and AI agents can benefit qualitative data analysis if the human recognizes and defines their role as a reflexive agent responsible for the entire process.

Keywords

agentic AI, Actor-Network Theory, sociomateriality, qualitative research, generative AI

Example 2: content analysis

AI functionalities have been introduced in various qualitative data analysis software (QDAS) available on the market.

In a study on the discourses of the companies behind these tools, Paulus & Marone (2024) found that the integration of AI into qualitative analysis platforms (QDAS) complicates the relationship between researchers and these tools.

A discourse analysis of ATLAS.ti, NVivo, and MAXQDA identified four key dilemmas:

- Automatic insight generation vs. meaning-making
- Interacting with documents vs. data analysis
- Speed vs. researcher engagement
- Innovation vs. researcher agency

The authors warn that, despite the potential benefits of these AI-powered features for content analysis, the commercial narratives surrounding these platforms may be incompatible with the epistemological foundations of qualitative research.

Example 2: content analysis

Christina Silver recently tested the integration of AI tools in qualitative analysis software, focusing on Atlas.ti and MaxQDA to assess whether AI supports or hinders critical thinking and reflexivity in qualitative research.

Potential Benefits:

- AI-generated summaries are fast and well-structured.
- AI-assisted querying aids in navigating large datasets and identifying key themes.
- AI suggests codes, offering an initial framework for analysis.
- Useful for descriptive tasks when the researcher is already familiar with the data.

Main Concerns:

- Reduced engagement with data, weakening interpretative depth.
- AI-generated coding lacks nuanced thematic connections.
- AI often produces excessive, fragmented, and redundant codes.
- Over-reliance on AI may diminish the researcher's analytical role.

Example 2: content analysis

Main Ethical and Methodological Challenges (Christina Silver)

Key Concerns:

- **Data Privacy & Consent:** AI processes and stores data, often without clear participant consent or transparency in data retention.
- **Reliability & Bias Risks:** AI-generated results vary with repeated queries and reflect biases from training data, potentially distorting insights.
- **Reflexivity & Interpretation:** AI identifies patterns but lacks critical thinking and interpretative intelligence, requiring human validation.

Future Considerations: AI can assist qualitative research but must be integrated cautiously. Researchers must retain control over interpretation, ensuring methodological rigor.

Bibliographie:

1. Amoores, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. *Political Geography*, 113, 103134. <https://doi.org/10.1016/j.polgeo.2024.103134>
2. Goodlad, L. M. E., & Stone, M. (2024). Beyond Chatbot-K: On Large Language Models, “Generative AI,” and Rise of Chatbots—An Introduction. *Critical AI*, 2(1). <https://doi.org/10.1215/2834703X-11205147>
3. Carrigan, M. (2025). *Generative AI for Academics* (1st edition). SAGE Publications Ltd.
4. Carobene, A., Padoan, A., Cabitza, F., Banfi, G., & Plebani, M. (2024). Rising adoption of artificial intelligence in scientific publishing: Evaluating the role, risks, and ethical implications in paper drafting and review process. *Clinical Chemistry and Laboratory Medicine (CCLM)*, 62(5), 835-843.
5. Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002).
6. Paulus, T. M., & Lester, J. N. (2023). Digital qualitative research workflows: A reflexivity framework for technological consequences. *International Journal of Social Research Methodology*, 1–14. <https://doi.org/10.1080/13645579.2023.2237359>
7. Schlagwein, D., & Willcocks, L. (2023). ‘ChatGPT et al.’: The ethics of using (generative) artificial intelligence in research and science. *Journal of Information Technology*, 38(3), 232-238.
8. Narayanan, A., & Kapoor, S. (2024). *AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference*. Princeton University Press.
9. Gray, A. (2024). "ChatGPT contamination": Estimating the prevalence of LLMs in the scholarly literature. arXiv preprint arXiv:2403.16887.
10. Liang, W., et al. (2024). Mapping the increasing use of LLMs in scientific papers. arXiv preprint arXiv:2404.01268.

Ressources:

1. Critical AI @ Rutgers. (2024, September 15). *Teaching Critical AI Literacies: An "Explainer" and Resources for the New Semester*.
<https://docs.google.com/document/d/1TAXqYGid8sQz8v1ngTLD1qZBx2rNKHeKn9mcfWbFzRQ/edit>
2. What is AI? (2024). *MIT Technology Review*.
<https://www.technologyreview.com/2024/07/10/1094475/what-is-artificial-intelligence-ai-definitive-guide/>
3. Salvaggio, E. (2024, August 29). Challenging The Myths of Generative AI. *Tech Policy Press*.
<https://techpolicy.press/challenging-the-myths-of-generative-ai>
4. Visual Storytelling Team and Madhumita Murgia. (2023, September 12). *Generative AI exists because of the transformer*. Financial Times. <https://ig.ft.com/generative-ai/>
5. Carrigan, M. (2024, September 15). LLMs and cognitive lock in.
<https://markcarrigan.net/2024/09/15/llms-and-cognitive-lock-in/>
6. Carrigan, M. (2024, July 30). Ten propositions about generative AI and the future of academic writing. <https://markcarrigan.net/2024/07/30/five-propositions-about-generative-ai-and-the-future-of-academic-writing/>

Ressources:

1. Lee, T. B., & Trott, S. (2023). Large language models, explained with a minimum of math and jargon. *Understanding AI*, 27, 2023.
2. European Commission and European Research Area (ERA). Living guidelines on the responsible use of generative AI in research. https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en
3. Cristina Silver. CNP webinar 046 AI Critical Thinking Reflexivity. <https://youtu.be/VbHLMfQ8zSo?si=fpJBQw62w5dufCE7>
4. Utilização de inteligência artificial (IA) em gestão de dados e investigação (Centro para a Gestão de Dados de Investigação iRe: Search da Ulisboa): https://drive.google.com/file/d/1-FBxa5y0zq3rfTaqXXPGoZWJJHibMMQ_/view?pli=1
5. David Bann and Liam Wright. Generative AI Tools for Quantitative Research: A Practical Guide: <https://www.ncrm.ac.uk/resources/online/all/?id=20859>
6. In Conversation videos (about AI in research): <https://www.ncrm.ac.uk/about/engage/inconversation.php>

Thank you!