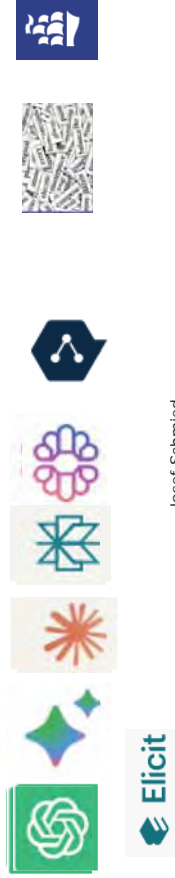


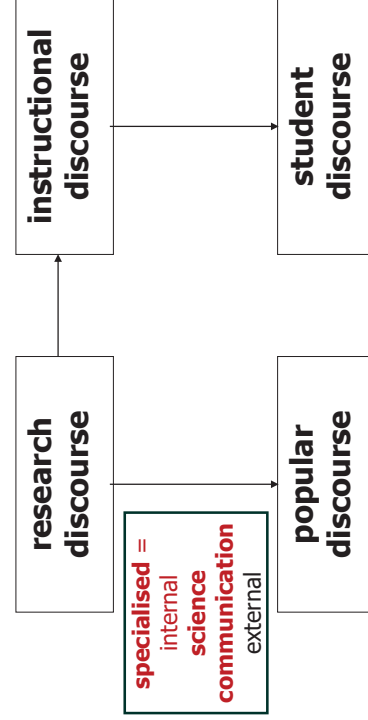
Specialised Discourse with Large Language Models? Beyond ChatGPT in Academic English



Josef Schmieid
English Language & Linguistics
Chemnitz University of Technology
https://www.tu-chemnitz.de/phil/english/sections/ling/presentations_js.php
josef.schmieid@phil.tu-chemnitz.de
<http://tinyurl.com/Josef-Schmieid>



1. Introduction: Academic Discourses



1. Introduction: Academic Discourse Genres (traditional)

“(staple) genres” as conventionalised practices (cf. Schmieid 2015)

- research “output”**
- research article (title,abstract,keywords)
 - book reviews
 - project proposals
 - conference presentations (abstract)

new genres

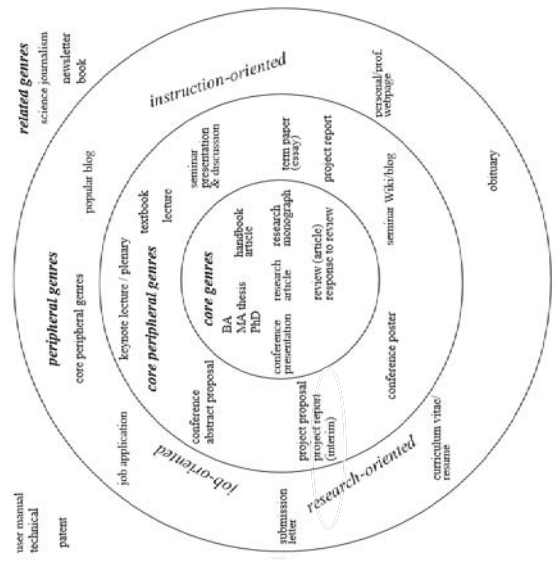
discipline-specific culture-specific

- science “journalism”**
- popular science articles
 - popular blogs (David Crystal)
 - popular science films (Horizon BBC 1964)
 - popular science books
 - science slam

- teacher genres**
- presentations
 - lectures (ppt)
 - seminar discussions
 - textbooks
 - Moodle/Wikis/www pages

- student genres (“literacy”)**
- fieldwork notes / essays
 - seminar presentations
 - BA/MA thesis
- “Novice Academic English”

author-specific culture-specific



(Schmied 2015)

social media
chatbots

„Artificial Intelligence“

„Office“

Large Language Models (LLMs)

Teacher

Lecture Planning Course-Curriculum-Translation

Skills Practice „Rules“, etc.

Specialised Concepts abstract title key words metadiscourse

Academic Style

Researcher for Ac. Writing (+critical thinking)
NOT Data Processing & Analysis, Outreach, etc.
(cf. Ray 2023, Farrokhnia et al. 2024)

Specialised Discourse

“new genres” as emerging practices

internal research discourse:
(cf. Plo-Alastrué et al. Eds. 2023):

- 3-minute talks/pitches
- open reviews
- multimodal: video abstracts
- Famelab scientific talks
- science blogs
- research group videos

= “broader concept” like Illingworth & Allen (2024)

external knowledge dissemination:
infodemic (cf. Metag et al. Eds. 2024)

- X/Twitter Tweets for projects
- Facebook for advertising
- Instagram/TikTok personal accounts
- Youtube channel

teacher genres

- ...

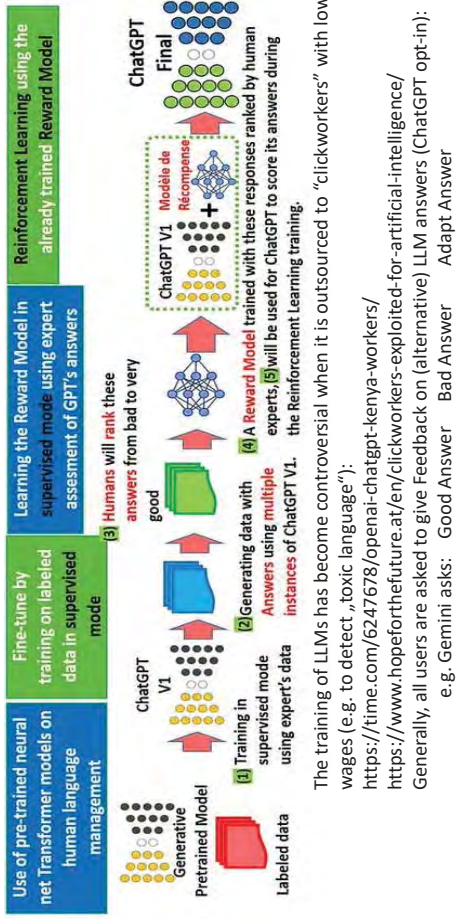
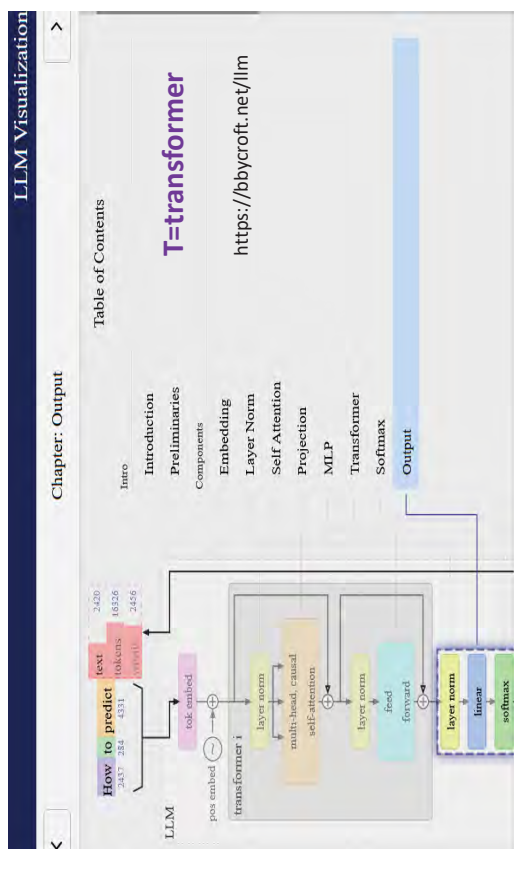
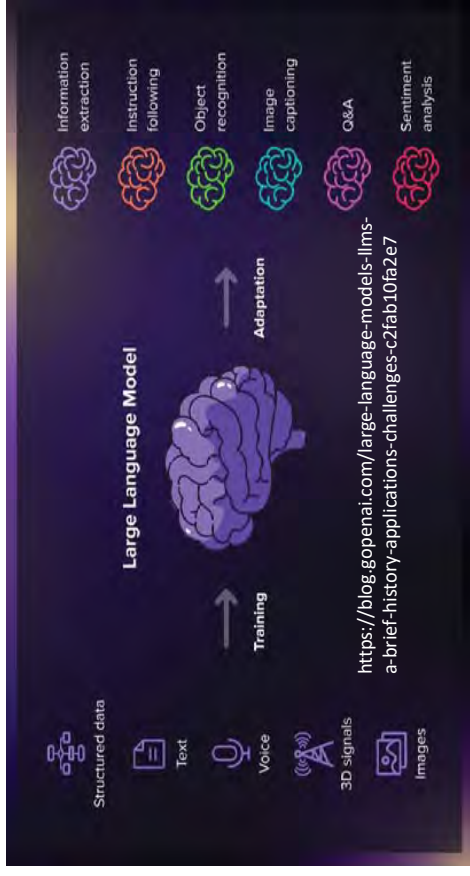
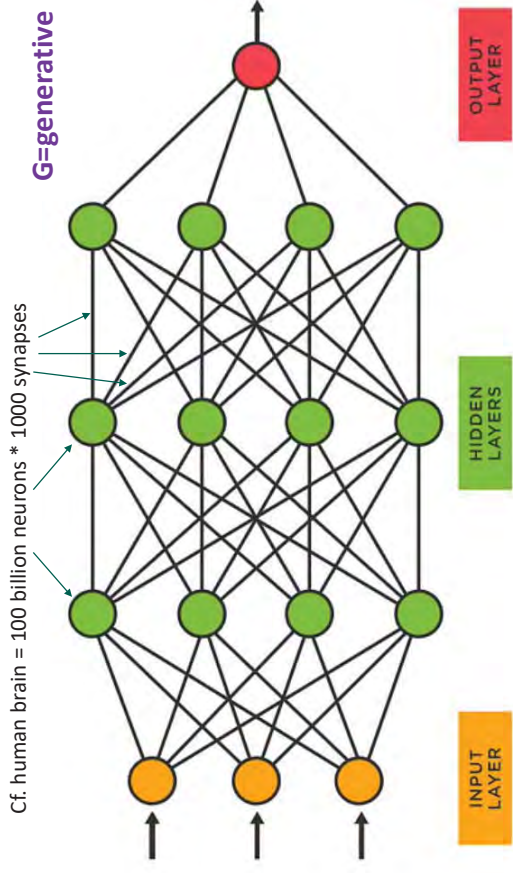
student genres (“literacy”)

Which youtube channels should I follow to understand research in ML. I am currently a sophomore in college. so kinda overwhelmed [D]

What Are Large Language Models?

An LLM is a machine learning model that processes and generates human language from training on massive volumes of data. It uses a deep learning algorithm to learn the patterns and nuances of language and provide human-like responses to a wide variety of queries and prompts. LLMs can also learn new skills such as creative writing or sentiment analysis.

The “large” in its name refers to the size of the datasets on which it’s trained. The goal of an LLM at its core is to infer what comes next in a sequence of texts. ChatGPT is a conversational agent built on top of an LLM that can generate text in a way that’s extremely similar to humans. When it comes to expanding the search experience, Coveo is testing generative AI capabilities that surface answers directly within search results.



The training of LLMs has become controversial when it is outsourced to “clickworkers”, with low wages (e.g. to detect „toxic language“):
<https://time.com/6247678/openai-chatgpt-kenya-workers/>
<https://www.hopeforthefuture.at/en/clickworkers-exploited-for-artificial-intelligence/>
 Generally, all users are asked to give feedback on (alternative) LLM answers (ChatGPT opt-in):
 e.g. Gemini asks: Good Answer Bad Answer Adapt Answer



P=pre-trained

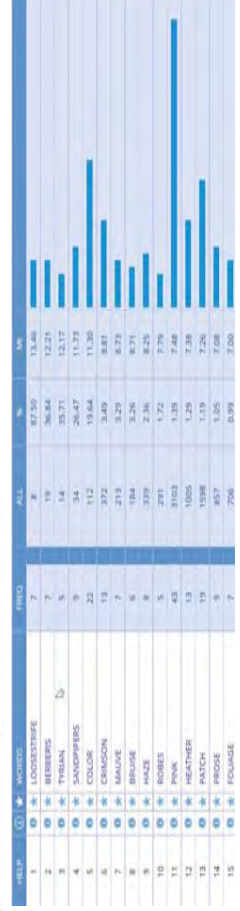
2. Large Language Models: Linguistic Principles „MI“



In our corpora, Mutual Information is calculated as follows:
 $MI = \log \left(\frac{AB * size(Corpus)}{(A * B * span)} \right) / \log(2)$
<https://www.english-corpora.org/mutualinformation.asp>

Suppose we are calculating the MI for the collocate *color* near *purple* in BNC.
 A = frequency of node word (e.g. *purple*); 12,68
 B = frequency of collocate (e.g. *color*); 1,12
 AB = frequency of collocate near the node word (e.g. *color* near *purple*); 32.
 sizeCorpus = size of corpus (# of words, in this case the BNC); 96,263,399
 span = span of words (e.g. 3 to left and 3 to right of base word); 6
 log(2) is literally the natural log of the number 2; 0.6931

BNCweb



2. Large Language Models: Linguistic Principles „collocates“



alcohol (NOUN) 2536

See in iWeb Collocates Clusters Topics Texts KWIC HELP ACCOUNT

TOPICS (more)

drinking, drug, drink, liquor, substance, cocaine, abuse, smoking, addiction, marijuana, heroin, smoking, consumption, drink, beer, drunk, treatment, excessive, use, harmful

COLLOCATES (more)

NOUN drug, use, abuse, consumption, tobacco, problem, blood, level
 drink, consume, rub, avoid, abuse, smell, smoke, associate
VERB feel, excessive, illegal, heavy, moderate, illicit, smoking, related
ADJ eg. ie, prenatally, voluntarily, negatively, excessively, moderately, capita
ADV

RELATED WORDS
 alcoholic, alcoholic, alcoholism, alcohol-related, alcohol-free, alcohol-based, blood-alcohol

1, a liquor or brew containing alcohol as the active agent 2, any of a series of volatile hydroxy) compounds that are made from hydrocarbons by distillation

CONTEXT
 BLOG WEB TV/M SPOK PIC MAG NEWS ACAD

WORD
 YouDish PlayPhrase Yarn
 DE Google WordNet Reverse Linguist

2. Large Language Models: SWOT Analysis ChatGPT in education



3. Brainstorming: Awareness of AI since ChatGPT

Mizumoto/Eguchi (2023):

The world was taken by surprise by the release of ChatGPT (<https://chat.openai.com/>) at the end of the year 2022. ChatGPT is a chatbot that answers human questions with an AI that seems to have a perfect understanding of the language. ...

The emergence of an AI tool capable of comprehending and producing a language like ChatGPT marks the beginning of a new era in which humans and AI coexist in L2 language learning and teaching, as well as in its research.

However, the use of ChatGPT is currently being evaluated, and some academic journals have cautioned against its irresponsible application (e.g., Nature Editorial, 2023). Despite criticisms from scholars like Chomsky, who denounce ChatGPT as a form of "high-tech plagiarism" (EduKitchen, 2023), it is evident that AI language models are here to stay, and it is our responsibility to educate students on their ethical use because "the future is now" (Pavlik, 2023, p. 9). At the same time, learners, teachers, and researchers can benefit from acquiring effective methods for utilizing GPT in their work, particularly non-native English speakers, who can be linguistically empowered. Language models such as GPT can be best understood as an ever-present, logical assistant, as AI is not currently a substitute for human expertise.

3. Brainstorming: Ask ChatGPT, who?! can do anything

ChatGPT

Sure, here are some examples of things you can tell me to do:

1. Generate a poem about nature.
2. Summarize the plot of "To Kill a Mockingbird" by Harper Lee.
3. Provide a list of the top 10 tourist attractions in Paris.
4. Explain the concept of quantum entanglement.
5. Create a simple recipe for chocolate chip cookies.
6. Generate a random 5-digit number.
7. Translate the phrase "Hello, how are you?" into Spanish.
8. Summarize the key points of a TED Talk on artificial intelligence.
9. Describe the process of photosynthesis in plants.
10. Provide a brief biography of Albert Einstein.

Which task would you like to start with?



3. Brainstorming: Ask Claude 3 Sonnet specifically

As an AI assistant, I can help you with various aspects of writing a research article:

1. **Literature Review**
 2. **Structuring the Article:** I can provide guidance on the typical structure of a research article, including the introduction, literature review, methodology, results, discussion, and conclusion sections. I can help you organize your ideas and ensure a logical flow of information.
 3. **Writing and Editing:** I can help you with the actual writing process, from drafting paragraphs and sections to improving clarity, coherence, and language usage. I can also assist with editing and proofreading to ensure your article is well-written and free of errors.
 4. **Formatting and Referencing**
 5. **Data Analysis and Visualization**
 6. **Manuscript Submission**
 7. **Collaboration and Feedback**
- It's important to note that while I can provide valuable assistance, I cannot replace your subject-matter expertise or critical thinking. My role is to support you throughout the writing process, but the research, analysis, and conclusions should ultimately come from you and your co-authors.

3. Brainstorming: Ask Gemini ethical issues

Large Language Models (LLMs) are powerful tools, but their use in research articles raises some ethical concerns. Here are some key areas to consider:

Authorship and Credit: Who gets credit for research findings when an LLM helps generate content? Should the LLM be a co-author? How much human oversight is needed? These are open questions that can blur the lines of traditional authorship. Oxford study on LLM responsibility [invalid URL removed]

Transparency and Reproducibility: Research relies on being able to repeat and verify findings. If an LLM is involved in data analysis or literature review, how can researchers ensure transparency about the LLM's role and limitations? Can others replicate the research if they don't have access to the same LLM?

Bias and Accuracy: LLMs are trained on massive amounts of data, which can reflect societal biases. These biases can be unknowingly incorporated into research articles, leading to skewed or inaccurate findings. Researchers need to be aware of these potential biases and take steps to mitigate them.

Overtrust and Misinformation: It's easy to be impressed by an LLM's ability to generate text. However, LLMs can still produce factually incorrect or misleading content. Researchers need to critically evaluate the LLM's outputs and avoid uncritically accepting them as truth.

These are just some of the ethical considerations surrounding LLM use in research. As this technology continues to develop, researchers and publishers will need to grapple with these issues and develop best practices for ethical and responsible LLM use.

3. Brainstorming with Bard (=Gemini): writing for journals

write on digital dissemination practices for linguists for publication in Peer-Reviewed Journals

choice of options

Vorschlag 1

Sure, here are some digital dissemination practices for linguists for publication in peer-reviewed journals:

informal style

Sure, here are some digital dissemination practices for linguists for publication in peer-reviewed journals:

- Publish in open access journals: Open access journals make their articles freely available online, which can help to increase the visibility and impact of research. There are a number of open access journals that specialize in linguistics, such as the *Journal of Linguistics* and the *International Journal of Language and Communication Disorders*.
- Create and share research data: Linguists can share their data through online repositories, such as the Linguistic Data Consortium and the Open Linguistics Archive. This can make it more accessible to other researchers and the public, and can facilitate further research on the data.
- Use social media: Social media platforms like Twitter, Facebook, and LinkedIn can be used to share research findings with a wider audience. Linguists can create accounts for themselves and their research groups, and can use these accounts to post links to their articles, data, and other resources. They can also use social media to interact with other linguists and the public, and to answer questions about their research.
- Create and maintain a website: Linguists can create websites to showcase their research and make it easier for people to find their work. The

3. Brainstorming: Ask ChatGPT 4 ethical issues

Using Large Language Models (LLMs) for research articles raises several ethical considerations that researchers should carefully address. Here are some key ethical issues:

1. Bias and Fairness
2. Plagiarism
3. Authorship and Attribution
4. Privacy and Confidentiality
5. Manipulation and Misinformation
6. Impact on Academic Publishing
7. Environmental Impact
8. Equitable Access

Addressing these ethical issues requires a combination of **awareness, transparency, accountability, and proactive measures** to mitigate potential risks and promote responsible use of LLMs in research articles.

4. Research Publishing: GPT4 Research Article Structure

Research articles in digital humanities typically follow a structure that combines traditional humanities scholarship with innovative digital methodologies and tools. While the specific structure may vary depending on the nature of the research and the preferences of individual researchers or journals, here's a prototypical structure commonly used in digital humanities research articles:

1. Title Page
2. Abstract
3. Keywords
4. Introduction
5. Literature Review
6. Methodology
7. Case Study or Analysis
8. Discussion
9. Conclusion
10. References
11. Appendices (if applicable)

This structure provides a framework for organizing research articles in digital humanities, integrating traditional humanities scholarship with innovative digital methodologies and technologies. Researchers should **adapt and refine this structure to fit the specific requirements** and conventions of their discipline, as well as the aims and scope of their research projects.

4. Research Publishing: the Art of Prompting

1. **Title**
Suggest a title for this research journal abstract
2. **Abstract**
Suggest a 300 word abstract for this title: Using large Language Models to write Research articles
3. **Keywords**
Suggest 5 keywords for this research journal abstract
4. **Introduction**
Write an (historical?) introduction for this research journal abstract
5. **Literature Review**
Write a literature review for this research journal abstract
6. **Methodology**
Which methodologies can I use for this article?
7. **Case Study or Analysis**
Find 3 case studies for this issue
8. **Discussion**
9. **Conclusion**
Which conclusions may I arrive at in this article?
10. **References**
Which 10 references are crucial for this article?

4. Research Publishing: GPT4o Conclusions (without analysis?)

Based on the findings from the case studies and the literature review, several conclusions can be drawn regarding the utilization of Large Language Models (LLMs) in research article composition:

- Enhanced Writing Efficiency and Productivity:**
 - LLMs offer researchers a powerful tool for streamlining the writing process, leading to increased efficiency and productivity.
 - Researchers using LLMs can generate coherent and contextually relevant text more quickly, enabling them to produce research articles in a more timely manner.
- Facilitated Interdisciplinary Collaboration**
- Improved Writing Creativity and Exploration**
- Ethical Considerations and Challenges**
 - The use of LLMs in research article composition raises important ethical considerations related to bias mitigation, plagiarism prevention, and authorship attribution.
 - Ethical guidelines and safeguards are essential to ensure responsible use of LLMs and mitigate potential risks associated with bias, plagiarism, and data privacy.
- Opportunities for Future Research and Development**

In conclusion, the integration of LLMs into research article composition holds immense potential for enhancing writing efficiency, promoting interdisciplinary collaboration, and fostering creativity in the scholarly community. However, it is crucial for researchers to navigate the ethical considerations associated with LLM-assisted writing responsibly and transparently. By critically evaluating the opportunities and challenges of LLMs in academic writing, researchers can harness their transformative potential to advance knowledge and scholarship in the digital age.

4. Research Publishing: Improve AI Proposals

The screenshot shows a DeepL AI writing tool interface. The main text area contains German text about AI development and its applications in research. A sidebar on the right provides AI-generated suggestions for improving the text, such as using hedging, engagement markers, and formal language. The interface includes a search bar, a language selector (English), and a sidebar with 'Anmerkungen anzeigen' (Show annotations) and '5/1'.

4. Research Publishing: Evaluate AI Proposals

Style academic

The screenshot shows the same DeepL AI writing tool interface as in the previous slide, but with several red circles highlighting specific terms and phrases. The highlighted terms are: 'specialised', 'specific positive', 'formal tentative', 'specific verbose?', 'less personal', 'explicit', and 'fundamental concepts'. The text in the main area is the same as in the previous slide, but with some words underlined in blue.

4. Research Publishing: Personalising ChatGPT advice

good concepts
typical examples
no references

how can I express my academic identity in my academic English through metadiscourse?

Expressing your academic identity in your academic writing, especially through metadiscourse, is crucial for establishing credibility, engaging readers, and conveying your authority on the subject. Metadiscourse refers to the language used to guide readers through your text and convey your attitude and stance toward the material. Here are some strategies for using metadiscourse effectively to express your academic identity in academic English:

- Hedging and Boosting:** Use hedging to indicate uncertainty or caution, and boosting to emphasize your confidence in your claims. For example:
 - Hedging: "It appears that the results may suggest a correlation..."
 - Boosting: "The robust findings unequivocally support the hypothesis..."
- Engagement Markers:** Engage with your readers by addressing them directly or signaling your awareness of their presence. This can foster a sense of connection and collaboration. For instance:
 - "In this paper, we will demonstrate..."
 - "As you can see from the data..."
- Epistemic Modality:** Indicate your degree of certainty or belief in the information you're presenting. Use adverbs and modal verbs like "clearly," "possibly," "likely," or "certainly" to express your stance:
 - "This finding clearly demonstrates..."
 - "It is likely that this phenomenon occurs because..."

5. Outlook: new criteria for good academic writing?

so far: individuality → learn conventionalised practices stricter culture/discipline-specific genre expectations?

Are conventionalised practices reinforced by LLMs? or Are unconventional practices favoured as they signal human, non-AI?

subject- and discipline-specific adaptations become more important? → new assessment criteria for qualification texts?

6. Conclusion: Ethical and legal issues

In the end the AI user decides – and is responsible!

- Copyright issues are still unclear?
- <https://noyb.eu/en/chatgpt>-provides-false-information-about-people-and-openai-cant-correct-it in academic writing and publishing all aids and resources used must be specified clearly!



After the launch of ChatGPT, the Pandora box opened. We are now observing a technological shift in the ways we do work. People are creating websites, apps, and even writing novels using ChatGPT.

With all the hype and introduction of AI generative tools, we have seen a rise in bad actors. If you are following the latest news, you must have heard that ChatGPT has passed the Wharton MBA exam. There is a whole list of exams passed by ChatGPT, from medical to a law degree - List: [Here](#). [See the Exams ChatGPT Has Passed so Far](#).

Apart from exams, students are using it to submit assignments, writers are submitting generative content, and researchers are producing high-quality papers just by typing prompts.

To counter the abuse of generative content, I am introducing you to 5 free AI content detector tools. You can use them to check the validity of the content and improve your SEO ranking.

<https://www.kdnuggets.com/2023/02/5-free-tools-detecting-chatgpt-gpt3-gpt2.html>

6. Conclusion: AI detection tools

100% AI likeness score

Top sentences driving AI probability

- 1 It helps to soften assertions and make them more diplomatic. **0.41**
- 2 In business texts, hedging is commonly employed to maintain a sense of professionalism and objectivity, and to acknowledge uncertainty or variability in certain situations. **0.30**
- 3 Hedging is a linguistic phenomenon where cautious or non-committal expressions are used to avoid making absolute statements or to mitigate potential risks or uncertainties associated with statements. In business texts, hedging is commonly employed to maintain a sense of professionalism and objectivity, and to acknowledge uncertainty or variability in certain situations. It helps to soften assertions and make them more diplomatic. **0.30**

6. Conclusion: AI detection and un-detection – the race is on

AI Detector Score

READS LIKE AI!

Try the AI Content Detector.
Our AI Checker is trained on blog posts, wikipedia entries, essays, and more articles found online and across multiple large language models (LLMs).

Then AI Humanizer Has You Covered!
The AI Humanizer uses a proprietary mix of 3 AI engines, NLP and semantic analysis algorithms to rewrite sentences and paraphrase paragraphs in a way that is so humanlike it bypasses even the toughest and most accurate AI detection tools – like this one!

Did you write this yourself or are you going to get grade optimized?

Predicted: **Fully Human**

Rewrite All AI Sentences

Check For AI Content

Users are used to avoid making absolute statements in business texts. Hedging is commonly employed to acknowledge uncertainty or variability in certain situations. [It helps to](#)

limit.
e. If you write like a robot, content by learning how to writing.

6. Conclusion: LLMs in specialised academic studies

Future research questions:

- Which LLM applications can be included in (non-native) graduate teaching? (e.g. Schmied 2023)
 - literature review (more than search engines)
 - (previous and new) research questions
 - academic style adaptations
 - visualization of quantitative data

Hello, Anthropic Team team@email.anthropic.com 31-10-24
 We are introducing the analysis tool, a new built-in feature for Claude.ai that enables Claude to write and run JavaScript code. Claude can now process data, conduct analysis, and produce real-time insights.

Hi there! This week, we're bringing you two improvements to "Chat with Papers," one of our most popular features. Researchers use Chat to ask questions about specific papers, create summaries, and more.

- Are ethical issues considered in LLM applications?

YES, at least in all instructions they are included

References

- Aktaş, S., S. Gök, D. Uzunoglu (2023). ChatGPT in education: *Türk Akademik Yayınlar Dergisi (TAY Journal)* 7(2), 378-406.
- Brabová, K., R. Skarmitz (2018). Foreign or native-like? The attitudes of Czech EFL learners towards accents of English and their use as pronunciation models. *Studie z aplikované lingvistiky. Studies in Applied Linguistics* 1, 38-50.
- Claude AI (2024). How to write a research paper with Claude AI. <https://claudeai.uk/how-to-write-a-research-paper-with-claude-ai/>
- Crosthwaite, P., & Balsa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 3(3), 100066.
- Farrokhi, M., Banihashemi, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International* 61(3), 460-474. <https://doi.org/10.1080/14703297.2023.2195846>
- Ivanova, M., Albrecht, S., & Schmied, J. (f.c.), Learning outcomes with Text-to-Speech synthesis of native and non-native English varieties. *Johannesson, P. (2024). Writing Your Thesis with ChatGPT: Research, Scholarship and Academic Writing in the Age of Generative AI.* <https://writingyourthesiswithchatgpt.wordpress.com/>.
- Mizumoto, A., M. Eguchi (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics* 2, 100050
- Ray, P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 3, 121-154, ISSN 2667-3452, <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- Skrabut, St. (2023). 80 Ways to Use ChatGPT in the Classroom: Using AI to Enhance Teaching and Learning. *StansKrabut. Change in Academic Writing. Special Issue. TOKEN: A Journal of English Linguistics* 16. <https://token.ujk.edu.pl/volumes/volume-16/>
- Schmied, J. (2023). Confidence and trust in online academic discourse: Integrating new technologies into teaching and learning. Schmied, J. & M. Ivanova (eds. 2023). *Comparing Confidence and Trust On- and Offline*. Göttingen: Cuvillier, 1-14.
- Zappavigna, M. (2023). Hack your corpus analysis: How AI can assist corpus linguists deal with messy social media data. *Applied Corpus Linguistics* 3(4), 100067.