

AI detecting:

Features of human-written and ChatGPT-generated essays



Josef Schmied

English Language & Linguistics

Chemnitz University of Technology

https://www.tu-chemnitz.de/phil/english/sections/ling/presentations_js.php

josef.schmied@phil.tu-chemnitz.de

<http://tinyurl.com/josef-schmied>



TECHNISCHE UNIVERSITÄT
CHEMNITZ

IntroductionConcepts: LLMsStudiesDetectionCorrectionConclusion

2/34

1. Introduction: LLM and Language Teaching & Learning

social media

„Artificial Intelligence“

„Office“:
spell/grammar/style checker

chatbots

Large Language Models (LLMs)

Teacher

Language Teaching/Learning (+critical thinking)

NOT Data Processing & Analysing, etc.
(cf. Ray 2023, Farrokhnia et al. 2024)

Lecture Planning
Course-
Curriculum-

Skills Practice
„Rules“, etc.
Translation

Grammar
Rules & Usage

Lexical Choices
Collocations

Idiomatcity
Style

Speaking & Listening

Reading

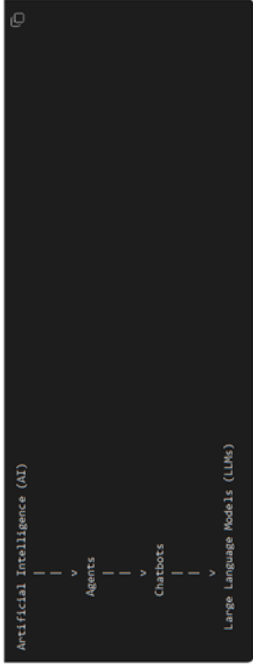
Writing

Translation

= the same integrity/ethical issues as with distance online teaching during the pandemic

→ 2 innovation waves in Language Teaching & Learning

2. Clarification of Concepts: Artificial Intelligence (AI), Agents, Chatbots, Large Language Models (LLMs)

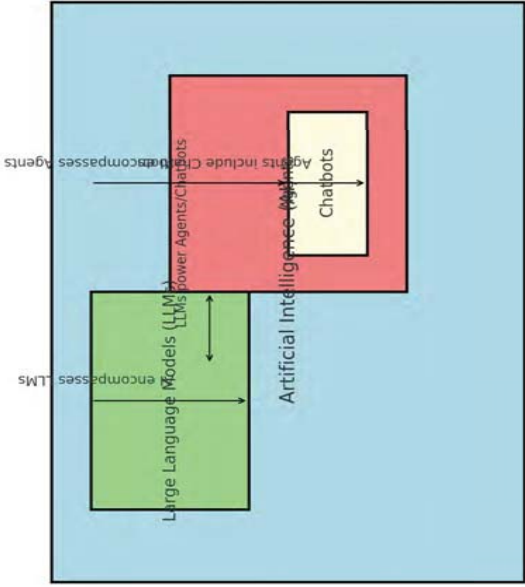


Explanation

- AI is the overarching field that includes all the other components.
- Agents are a type of AI system designed to perform tasks autonomously.
- Chatbots are a specific type of agent designed for conversational interactions.
- LLMs are advanced AI models that can be used to enhance the capabilities of chatbots by providing more sophisticated language understanding and generation.

In summary, LLMs are a type of AI model that can be used to power chatbots, which are a type of agent, all of which fall under the broader category of AI.

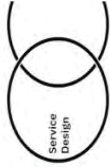
2. Clarification of Concepts: Artificial Intelligence (AI), Agents, Chatbots, Large Language Models (LLMs)



2. Clarification of Concepts: Artificial Intelligence (AI), Agents, Chatbots, Large Language Models (LLMs)



Here's a simple diagram illustrating the relationship between artificial intelligence, agents, chatbots, and LLMs:

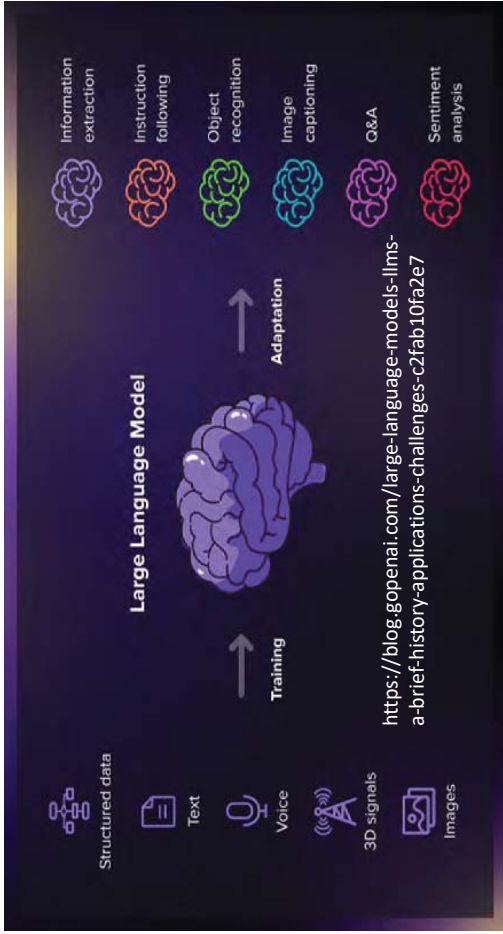


Venn diagram with four overlapping circles. The largest circle is labeled Artificial Intelligence (AI). Inside the AI circle, there are three smaller circles labeled Agents, Chatbots, and Large Language Models (LLMs). The Agents and Chatbots circles overlap, and the Chatbots and LLMs circles overlap.

Explanation of the Diagram

- **Artificial Intelligence (AI):** This is the broadest category, encompassing any technique that allows computers to mimic human intelligence. It includes a wide range of approaches, from simple rule-based systems to complex machine learning models.
 - **Agents:** These are AI systems designed to perceive their environment and take actions to achieve specific goals. They can be simple or complex, and may or may not involve natural language processing.
 - **Chatbots:** These are AI systems designed to simulate conversation with human users. They typically rely on natural language processing to understand and respond to user input.
 - **LLMs (Large Language Models):** These are a type of AI model that is trained on vast amounts of text data to generate human-like text. They are often used in chatbots to provide more natural and engaging conversations.
- Key Points**
- Chatbots and LLMs are both subsets of AI.
 - Chatbots often use LLMs to enhance their conversational abilities.
 - Agents can be chatbots, but they can also take other forms, such as robots or autonomous vehicles.
 - Not all agents are chatbots, and not all chatbots use LLMs.

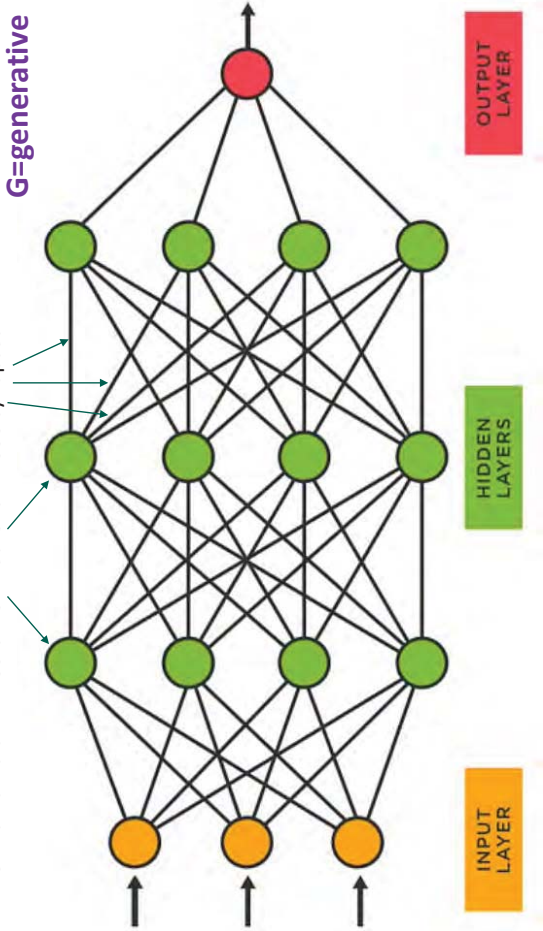
2. Large Language Models: Principles



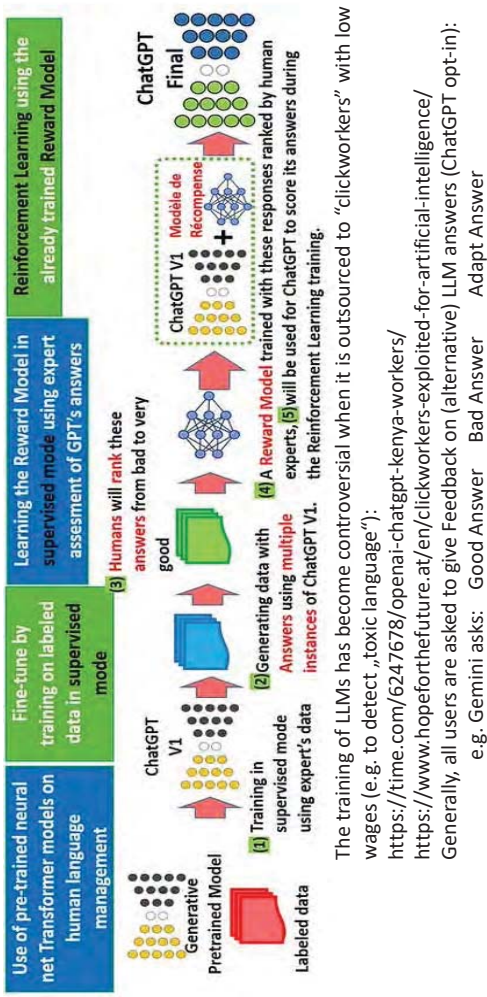
<https://blog.gopenai.com/large-language-models-llms-a-brief-history-applications-challenges-c2fab10fa2e7>

2. Large Language Models: G=Neural Network

Cf. human brain = 100 billion neurons * 1000 synapses



2. Large Language Models: P=Reward Model Training



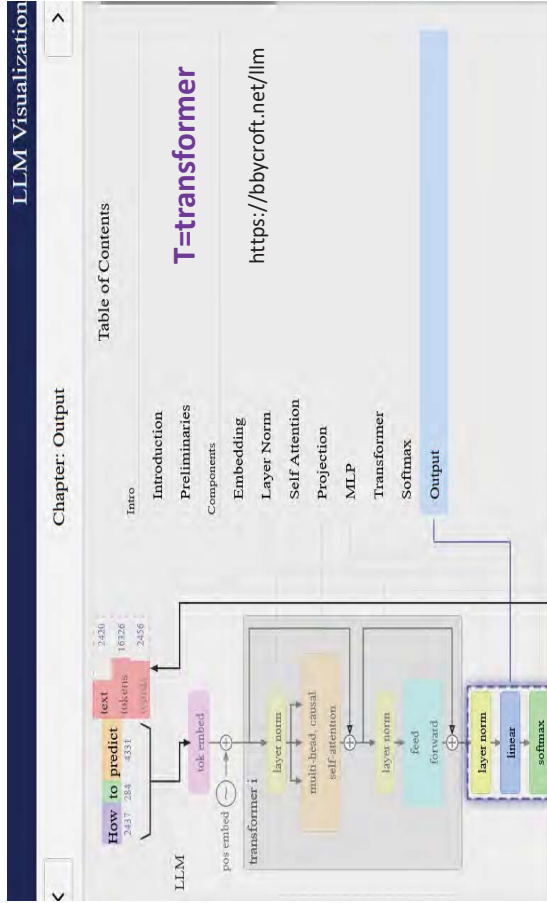
P=pre-trained

👍

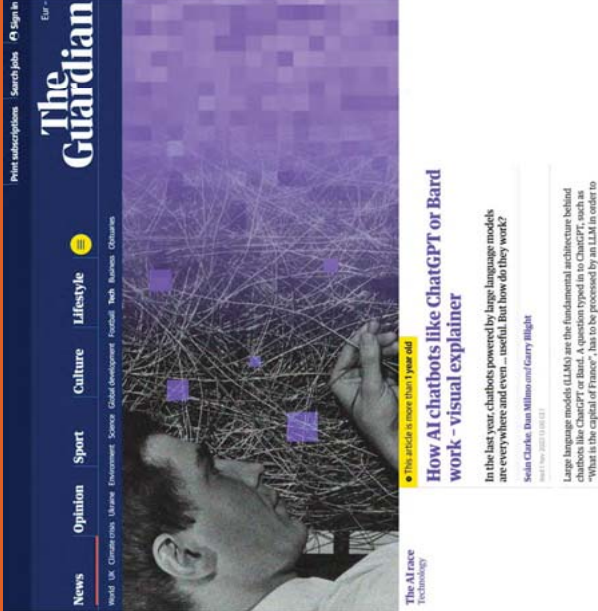
💬

🚫

2. Large Language Models: T=Transformer Visualisation



2. Large Language Models: Visualisation



<https://www.theguardian.com/technology/interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer>

2. Large Language Models: Linguistic Principles „MI“



In our corpora, Mutual Information is calculated as follows:
 $MI = \log \left(\frac{AB}{sizeCorpus} \right) / \left(\frac{A}{sizeCorpus} \right) / \left(\frac{B}{sizeCorpus} \right)$

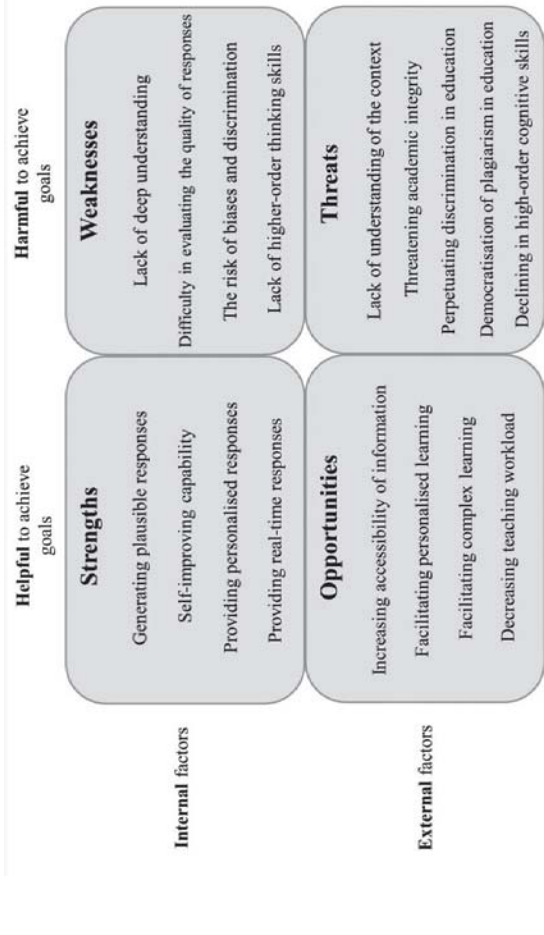
Suppose we are calculating the MI for the collocate *color* near *purple* in BNC.

A = frequency of node word (e.g. *purple*): 1246
B = frequency of collocate near the node word (e.g. *color*): 112
sizeCorpus = size of corpus (# words: in this case the BNC: 96,263,399
span = span of words (e.g. 3 to left and 3 to right of node word): 6
log (2) is literally the natural log of the number 2: 0.6931

$MI = 11.30 = \log \left(\frac{1246 * 96,263,399}{(1246 * 112 * 6)} \right) / 0.6931$

BNCweb

2. Large Language Models: SWOT Analysis ChatGPT in education



3. Studies on human vs. AI essay scoring and writing



Research Methods in Applied Linguistics
Volume 2, Issue 2, August 2023, 100050

Exploring the potential of using an AI language model for automated essay scoring

Atsushi Mizumoto a, b, c, Masaki Eguchi b

The widespread adoption of ChatGPT, an AI language model, has the potential to bring about significant changes to the research, teaching, and learning of foreign languages. The present study aims to leverage this technology to perform automated essay scoring (AES) and evaluate its reliability and accuracy. Specifically, we utilized the GPT-3 text-davinci-003 model to automatically score all 12,100 essays contained in the ETS Corpus of Non-Native Written English (TOEFL11) and compared these scores to benchmark levels. The study also explored the extent to which linguistic features influence AES with GPT. The results showed that AES using GPT has a certain level of accuracy and reliability and could provide valuable support for human evaluations. Furthermore, the analysis revealed that utilizing linguistic features could enhance the accuracy of the scoring. These findings suggest that AI language models, such as ChatGPT, can be effectively utilized as AES tools, potentially revolutionizing methods of writing evaluation and feedback in both research and practice. The paper concludes by discussing the practical implications of using GPT for AES and exploring prospective future considerations.

3. Studies on human vs. AI essay scoring and writing

Herbold et al. (2023),
https://www.nature.com/articles/s41598-023-45644-9
PROMPT: "Write an essay with about 200 words on "[topic]"
guidelines for essay assessment
established by the Ministry for Education of Lower Saxony:
zero is the worst score (e.g. no relation to the topic) and **six** is the best score (e.g. addressed the topic to a special degree).
658 ratings by 111 experienced secondary school teachers for 270 essays
(90 topics for each essay type: human-, ChatGPT-3-, ChatGPT-4-generated).

Criterion	Humans		ChatGPT-3		ChatGPT-4	
	M	SD	M	SD	M	SD
Topic and completeness	3.58	1.30	0.95	4.24	1.16	0.95
Logic and composition	3.64	1.27	0.96	4.29	1.04	0.96
Expressiveness and compr.	3.42	1.25	0.95	3.90	1.04	0.95
Language mastery	3.90	1.37	0.89	5.03	1.19	0.89
Complexity	3.72	1.26	0.92	4.20	1.14	0.92
Vocabulary and text linking	3.78	1.18	0.97	4.41	1.05	0.97
Language constructs	3.80	1.15	0.97	4.47	1.02	0.97
Overall	3.69	1.26	0.96	4.36	1.14	0.96

"clear order of the mean values, with students having the worst ratings, ChatGPT-3 in the middle rank, and ChatGPT-4 with the best performance"

3. Studies on human vs. AI essay scoring and writing



Explore contentAbout the journalPublish with us

nature>scientific reports>articles>article

Article | Open access | Published: 30 October 2023

A large-scale comparison of human-written versus ChatGPT-generated essays

Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva & Alexander Trautsch

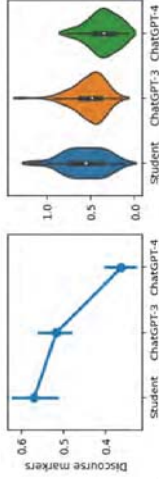
Scientific Reports 13, Article number: 18617 (2023) | Cite this article

.... We systematically assess the quality of AI-generated content through a large-scale study comparing human-written versus ChatGPT-generated argumentative student essays. We use essays that were rated by a large number of human experts (teachers). We augment the analysis by considering a set of linguistic characteristics of the generated essays. Our results demonstrate that ChatGPT generates essays that are rated higher regarding quality than human-written essays. The writing style of the AI models exhibits linguistic characteristics that are different from those of the human-written essays. Since the technology is readily available, we believe that educators must act immediately. We must re-invent homework and develop teaching concepts that utilize these AI models in the same way as math utilizes the calculator: teach the general concepts first and then use AI tools to free up time for other learning objectives.

3. Studies on human vs. AI essay scoring and writing

Herbold et al. (2023) Table 3:

... no clear tendency regarding the direction of the differences. For instance, while the ChatGPT models write more complex sentences and use more nominalizations, humans tend to use more modals and epistemic markers instead.
The lexical diversity of humans is higher than that of ChatGPT-3 but lower than that of ChatGPT-4. While there is no difference in the use of discourse markers between humans and ChatGPT-3, ChatGPT-4 uses significantly fewer discourse markers.



Arithmetic mean (M) and standard deviation (SD) for the linguistic markers

Linguistic characteristic	Humans	ChatGPT-3	ChatGPT-4
Lexical diversity	95.72	23.50	75.68
Syntactic complexity (depth)	5.72	0.80	6.18
Syntactic complexity (clauses)	1.81	0.57	2.31
Nominalizations	1.06	0.51	1.56
Modals	10.84	5.30	8.97
Epistemic markers	0.06	0.06	0.02
Discourse markers	0.57	0.24	0.52

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!

KD

nuggets

Blog

Topics

Datasets

Events

Resources

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. Are 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>

GPTZero

SOLUTIONS

RESOURCES

PRICING

TEAM

NEWS

LOG IN

SALES

GET STARTED

RESULTDEEP SCAN SOURCES

100% AI likeness score

AI

Human

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.

Sentences most impacting the probability score (Learn more)

Top sentences driving AI probability

AI Impact

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.

0.30

GPTZero

SOLUTIONS

RESOURCES

PRICING

TEAM

LOG IN

SALES

GET STARTED

RESULTDEEP SCAN SOURCES

Learn more about this updated version of GPTZero.

Classification

We are highly confident this text is entirely human

3%

Probability AI generated

highly confident

THESE COPIES

The rapid development of AI presents new opportunities for learning and utilizing foreign languages. While tools such as ChatGPT, Bard, Perplexity, or DeepL are currently available at no cost, it is necessary to gain an understanding of their applications in order to utilize them effectively for personal or professional purposes.

This course will provide an overview of the fundamental concepts associated with the most prevalent AI tools, which are often referred to as "stochastic parrots". The course will apply these tools to a variety of English texts, including travelogues, book reviews, summaries, and others. The results generated by AI will be discussed. Rather than writing the focus will be on improving existing texts.

Rather than translating the objective will be to compare versions. This course is suitable for anyone who wants to create foreign-language texts, especially at an advanced level for personal or professional purposes.

GPTZero

SOLUTIONS

RESOURCES

PRICING

TEAM

LOG IN

SALES

GET STARTED

RESULTDEEP SCAN SOURCES

KI-Detektor wurde aktualisiert: Wir haben das neue Detector V2-Modell mit 305 Mio. Proben trainiert, um die neuesten KI-Schreibmuster deutlich besser zu erkennen!

KI-Detektor

Preise

Kontakt

Unser Blog

Funktionen

KI-Detektor

Fügen Sie Ihren Text unten ein, um auf KI-generierte Inhalte zu überprüfen.

KI-Detektor

79.5%

Dieser Text wurde wahrscheinlich geschrieben von KI

Es besteht eine 79.5% Wahrscheinlichkeit, dass dieser Text vollständig geschrieben wurde von KI

KI-Test humanisieren

5. Detection – Correction – Improvements: AI Proposals

Style improvement: DeepL write

Can you improve the following text to make it more specific academic?



ChatGPT

Certainly! Here's the improved version of the text with a more specific academic tone:

"The rapid advancement of artificial intelligence (AI) technology has heralded a new era in language acquisition and utilization. Emerging tools such as ChatGPT, Bard, Perplexity, and DeepL offer unprecedented opportunities for linguistic exploration. While these resources are presently accessible without cost, their effective integration into personal or professional contexts necessitates proficiency acquired through dedicated learning."

"In this instructional program, we will elucidate the fundamental principles underlying prevalent AI tools, colloquially referred to as 'stochastic parrots'. Through practical application, participants will engage with various genres of English literature, including travelogues, literary critiques, and abstracts, leveraging AI-generated outputs for critical examination. Emphasizing iterative refinement over original composition and comparative analysis in lieu of direct translation, the curriculum is tailored to individuals seeking to refine their multilingual textual composition skills, particularly at an advanced level, to serve both personal and professional objectives."



5. Detection – Correction – Improvements: AI Proposals

Synonyms explained by ChatGPT =similar!



ChatGPT

Certainly! Here are several synonyms for "fast" along with explanations of their differences:

1. Quick: This synonym emphasizes speed and efficiency in completing a task or action. It often implies a rapid response or movement without delay. For example, "She was a quick wit" suggests that she is mentally sharp and responds rapidly in conversations.
2. Rapid: Similar to "quick," "rapid" also denotes a high speed or rate of movement. However, "rapid" tends to suggest a swiftness that is sustained over a period of time. For instance, "The company experienced rapid growth" indicates that the growth occurred swiftly and continuously.
3. Swift: "Swift" conveys a sense of speed and agility, often implying graceful or effortless movement. It is commonly used to describe actions or motions that are smooth and efficient. For example, "The swift runner crossed the finish line first" suggests that the runner moved quickly and smoothly.
4. Speedy: This synonym emphasizes the characteristic of being fast or quick, often in comparison to a standard or expectation. It suggests that something is accomplished or achieved in a short amount of time. For instance, "He provided a speedy resolution to the problem" implies that he resolved the problem quickly, perhaps faster than anticipated.
5. Fleet: "Fleet" refers to a group of vehicles or vessels moving together, but it can also be used to describe something that moves swiftly or with agility. It often implies a sense of nimbleness or quickness in motion. For example, "The fleet-footed deer darted through the forest" suggests that the deer moved swiftly and gracefully.
6. Expeditious: This synonym emphasizes the promptness and efficiency with which something is done. It conveys a sense of urgency and effectiveness in completing a task or achieving a goal. For instance, "The committee made an expeditious decision" implies that the decision was made quickly and efficiently.

5. Detection – Correction – Improvements: trad. Proposals

Online dictionary

THESAURUS

fast moving or able to move quickly
The cheetah is the fastest animal in the world.
a fast car

quick moving fast or doing something in a short time
He was much quicker than I was over the first 100 metres.
Do I have time for a quick shower?

high-speed [only before noun] designed to travel or operate very quickly
a high-speed train
high-speed Internet access

rapid *especially written* happening in a short period of time – used about changes, increases, improvement
a rapid increase in the population
the rapid expansion of the firm's business in the Middle East
a rapid decline in profits

swift *written* moving quickly or happening after only a short time
The horses ran along the track at a swift trot.
He received a swift response to his letter.

brisk quick and energetic
a brisk walk in the countryside
His manner was very brisk.

speedy [only before noun] happening after only a short time
Everyone wishes you a speedy recovery.
a speedy resolution to the problem

hurried done more quickly than usual, because you do not have much time
She ate a hurried breakfast in the cafe before catching her train.
We made a hurried departure.

hasty deciding or doing something very quickly, especially when this has bad results
It was a hasty decision, which he later regretted.
Let's not be too hasty.

6. Conclusion

LLMs

- are fascinating and can be used as tools in many routine questions
- „hallucinate“ (invent „facts“) less and less, but you must be able to check
- are a good starting point, if you can evaluate output
- can be used in addition to traditional tools, such as dictionaries, thesauri, corpora, etc.
- are not “Artificial Super Intelligence”, but the are useful tools for standard routine tasks

BUT

- we assume that students want to learn, not cheat!
- we know our students’ development-specific errors and preferences!
- we do not focus only on the final product, but the processing of language!
- ethic rules must be adhered to
- prompts have to be specific and thus trained
- force teachers to reconsider their assessment rubrics and tasks – less standard, more i?

If texts sound too good to be truly written by our students, we also ask ourselves how much of this elaboration, etc. we want or have to include in teaching ...

Evaluating LLM output can teach us a lot about how language works!



The focus of this lecture was not really AI detection, but what we can learn from AI to write better!! 😊

References

Atkay, S., S. Gök, D. Uzunoglu (2023). ChatGPT in education. *Türk Akademik Yayınlar Dergisi (TAY Journal)* 7(2), 378-406.

Clarke, S., D. Milmo & G. Blight (2023). How AI chatbots like ChatGPT or Bard work. A visual explainer. *The Guardian* 01/11/23. <https://www.theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer> (last accessed 12/01/25)

Grothswalte, P., & Balsa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 3(3), 100066.

Farrokhiya, M., Banhashem, 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/10.1080/147030297.2023.2195846>

Herbold, S., Hautli-Janisz, A., Heuer, U. et al. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Nature Scientific Reports* 13, 18617. <https://doi.org/10.1038/s41598-023-45644-9>

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, 2, 100050

Ray, P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and a 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>.

Skrabrut., St. (2023). 80 Ways to Use Chatgpt in the Classroom: *Using AI to Enhance Teaching and Learning*. StansKrabrut.

Schmied, J., M. Bondi, O. Doncheva Navratilova, & C. Perez-Llantada (eds.) (2023). Patterns of Language Variation and 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.

Exercise: Evaluate = accept OR reject DeepL suggestions



English (British) ▾	The lecture provided an interesting overview of the developments in digital science communication, highlighting both its opportunities and challenges. I found the discussion on the emergence of new genres, such as open peer review and 3-minute thesis presentations, particularly engaging. The shift from traditional publishing models to more transparent, interactive, and digitized forms of scientific discourse is fascinating, especially in the context of increased public participation and accessibility. The exploration of ethical concerns, such as the role of AI in research communication and the risks of misinformation, also added depth to the topic. It was insightful to see how digital platforms are reshaping the way scientific knowledge is shared and evaluated, fostering interdisciplinary collaboration while also raising questions about research integrity and audience engagement.	While the lecture touched on the tensions between accessibility and authority in science communication, I would have liked a more in-depth analysis of how different disciplines navigate this balance. Additionally, the discussion on the impact of digitalization on academic writing genres was intriguing, but I would have appreciated more concrete examples of how emerging practices are being adopted across different research fields. A deeper exploration of how AI, particularly large language models, is influencing academic writing and peer review would have been intriguing, considering the ethical and stylistic debates surrounding their use. Overall, the lecture was highly engaging, and I would love to hear more about the long-term implications of these transformations. I truly appreciate the opportunity to have been part of this insightful lecture.	1753 / 2000
English (British) ▾	The lecture provided an interesting overview of the developments in digital science communication, highlighting both its opportunities and challenges. I found the discussion on the emergence of new genres, such as open peer review and 3-minute thesis presentations, particularly engaging. The shift from traditional publishing models to more transparent, interactive, and digitized forms of scientific discourse is fascinating, especially in the context of increased public participation and accessibility. The exploration of ethical concerns, such as the role of AI in research communication and the risks of misinformation, also added depth to the topic. It was insightful to see how digital platforms are reshaping the way scientific knowledge is shared and evaluated, fostering interdisciplinary collaboration while also raising questions about research integrity and audience engagement.	While the lecture touched on the tensions between accessibility and authority in science communication, I would have liked a more in-depth analysis of how different disciplines navigate this balance. Additionally, the discussion on the impact of digitalization on academic writing genres was intriguing, but I would have appreciated more concrete examples of how emerging practices are being adopted across different research fields. A deeper exploration of how AI, particularly large language models, is influencing academic writing and peer review would have been intriguing, considering the ethical and stylistic debates surrounding their use. Overall, the lecture was highly engaging, and I would love to hear more about the long-term implications of these transformations. I truly appreciate the opportunity to have been part of this insightful lecture.	1753 / 2000