





The Future of Education in the Age of Generative Artificial Intelligence

Dragan Gašević

June 26th, 2023 Faculty of Education University of Hong Kong



Generative artificial intelligence

Generating data* in response to prompts

*text, image, video, program code



Generative artificial intelligence

Many debates to sensemake uncertainty linked to generative AI



Fearless, independent, reader-funded

Support us \rightarrow

News Opinion

Sport Culture

Lifestyle

Australia Coronavirus World AU politics Environment Football Indigenous Australia Immigration Media Business Science Tech

Artificial intelligence Explainer

What is AI chatbot phenomenon ChatGPT and could it replace humans?

The tool has impressed experts with its writing ability, proficiency at complex tasks and ease of use

Samantha Lock

✓@Samantha_Lock
Mon 5 Dec 2022 17.13 AEDT

f) 🕑 🖾



Chat GPT is being hailed as a potential game-changer in the world of Al Photograph: Andriy Popov/Alamy

The New York Times

THE SHIFT

An

More~

The Brilliance and Weirdness of ChatGPT

A new chatbot from OpenAI is inspiring awe, fear, stunt attempts to circumvent its guardrails.



Here is what DALL-E 2 produced when given the prompt, "A distributed linguistic superbrain that takes the form of an A.I. chatbot." Kevin Roose, via DALL-E



Dec. 5, 2022

An Al chatbot went viral. Some say it's better than Google; others worry it's problematic.

An Al chatbot went viral. Some say it's better than Google; others worry it's problematic.

ChatGPT has become social media's latest favorite tool for getting answers online. But some have warned bad actors could take advantage of the bot.



Dec. 3, 2022, 8:58 AM AEDT By Kalhan Rosenblatt



THE CONVERSATION

Academic rigour, journalistic flair

Arts + Culture Books + Ideas Business + Economy Education Environment + Energy Health Politics + Society Science + Tech

ChatGPT is the push higher education needs to rethink assessment

Published: March 12, 2023 6.42pm AEDT



With proper teaching, students can use ChatGPT to develop their arguments and build their essays. Gorodenkoff/Shutterstock

Authors Sioux McKenna Director of Centre for Postgraduate Studies, Rhodes University & Visiting Research Professor in Center for International Higher Education, Boston College, Rhodes University Dan Dixon Adjunct Lecturer, University of Sydney **Daniel Oppenheimer** Professor of Decision Science and Psychology, Carnegie Mellon University Margaret Blackie Associate Professor, Rhodes University Sam Illingworth Associate Professor, Edinburgh Napier University Disclosure statement

> The authors do not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and have disclosed no relevant affiliations beyond their academic appointment.

Q Search analysis, research, academics.

THE Times Higher Education

Campus Resources for academics and university staff

ChatGPT and the future of university

assessment

Artificial intelligence-powered tools like ChatGPT are forcing a much-needed opportunity to reimagine the role of education in the 21st century, says Alex Sims





Debates about assessment in the age of artificial intelligence





Artificial Intelligence and the Future of Teaching and Learning

Insights and Recommendations







https://tech.ed.gov/ai-future-of-teaching-and-learning/

https://tech.ed.gov/ai-future-of-teaching-and-learning/



Today's talk

What are promises, concerns, and needs for education in the age of AI?



Key takeaway 1

Much promise for guiding, teaching support, and assessment/feedback



Key takeaway 2

Significant concerns about reliability, bias, workload, and consent



Key takeaway 3

Transition from commentaries to research to understand implications



PROMISE – CONCERN – NEED – FINAL REMARKS



Uses of LLMs in education

Categories	Educational Tasks
Profiling and Labelling	Forum post classification, dialogue act classification, classification of learn- ing designs, review sentiment analysis, topic modelling, pedagogical classifi- cation of MOOCs, collaborative problem-solving modelling, paraphrase qual- ity, speech tagging, labelling educational content with knowledge components, key sentence and keyword extraction, reflective writing analysis, multimodal representational thinking, discipline similarity, concept classification, cognitive level classification, essay arguments segmentation
Detection	Semantic analyses, detecting off-task messages, confusion detection, urgency detection, conversational intent detection, teachers' behaviour detection
Assessment and Grading	Formative and summative assessment grading, short answer grading, essay grading, subjective question grading, student self-explanation
Teaching Support	Classroom teaching, learning community support, online learning conversation agent, intelligent question-answering, teacher activity recognition

Categories	Educational Tasks
Prediction	Student performance prediction, student dropout prediction, emotional and cognitive engagement detection, growth and development indicators for college students, at-risk student identification
Knowledge Representa- tion	Knowledge graph construction, knowledge entity recognition, knowledge trac- ing, cause-effect relation extraction
Feedback	Real-time feedback, post-hoc feedback, aggregated feedback, feedback on feedback (peer-review comments)
Content Generation	MCQs generation, open-ended question generation, code generation, reply (natural language) generation
Recommendation	English reference selection and recommendation, resource recommendation, course recommendation

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D. (2023). Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review. *Submitted for review*. https://arxiv.org/abs/2303.13379





Large language models for guiding and tutoring





Singer, N. (2023, June 8). Not Just Math Quizzes: Khan Academy's Tutoring Bot Offers Playful Features. *The New York Times*. https://www.nytimes.com/2023/06/08/business/khanmigo-tutor-chat.html



Writing, debugging and explaining code





Explanation:

- The function `first_10_primes` initializes an empty list `prime_numbers` to store the first 10 prime numbers.
- 2. It then sets the starting value of `num` to 2, the smallest prime number.
- 3. The function uses a `while` loop to run until the length of `prime_numbers` is equal to



Provide step-bystep solutions to questions from many disciplines



\$

The equation $x^2 - 3|x - 2| - 4x = -6$ can be solved by considering two separate \bigcirc \bigcirc cases: when x < 2 and when x >= 2.

When x < 2, the absolute value |x - 2| becomes 2 - x, and the equation becomes:



When $x \ge 2$, the absolute value |x - 2| becomes x - 2, and the equation becomes:



Khorsavi, H. (2023). ChatGPT and the Future of Education. GRAILE Webinar, https://www.youtube.com/watch?v=m7F6VhOkoYc





Large language models for content generation and design support

Support learning designers

I would like you to act as a grammar language subject matter expert that also has learning designer skills and knowledge of markdown. You know about the grammatical errors that first year university students make. You also know how to design online modules, you can break up the content into chunks, write quizes to test the knowledge of students and come up with active learning strategies to use in a face-to-face zoom virtual lecture. Could you please come up with the sections for an online module on Grammar for first year university students in Markdown format (i.e. using # and ## to indicate headings) in the code editor so that I can copy the Markdown?

Sure, here is an example of what the sections of an online module on grammar for first year university students in Markdown format could look like: 6 P



Bakharia, A. (2023). Act as a Learning Designer — Getting ChatGPT to Generate an Online Module. https://aneesha.medium.com/act-as-a-learning-designer-getting-chatgpt-to-generate-an-online-module-8a16a2813bd6

Α

\$



Generate items for assessments

C()LAM



want a lesson to help students learn how to identify, summarize, or explain the main idea using supporting details in an expository text. That's a great topic! Have your students had any previous lessons on identifying the main idea or summarizing expository texts?

7ء

No

7

♦ Clear chat

Teacher development https://www.khanacademy.org/khan-labs

Singer, N. (2023, June 8). Not Just Math Quizzes: Khan Academy's Tutoring Bot Offers Playful Features. The New York Times. https://www.nytimes.com/2023/06/08/business/khanmigo-tutor-chat.html



GPTeach – teacher development

Session Description You're a TA for an introductory computer science course at	Concerne Contraction of the second se	Stanford Code in Plac	ce	Teach	¢٠	≞ * ¥*	► ∰
They have a question regarding how the following line works: new_string = input_string[-1]	2 student(s) present: Jia, Lakisha				Sectio	n Chat	C 0
Learning Goals Understand student misconceptions and hidden blockers Create a safe, inclusive learning environment When you feel you've achieved the learning goals, move on to	I'm having some trouble understanding this line of code: new_string = input_string[:-1]. I'm feeling a bit apprehensive about this class and I'm hoping you can help me understand it better. Lakisha		600			to today's sections started for today everyone share from lecture that excited about, a	on! To get y, can one thing at you're and one thing
Next Session	Hi. I'm having trouble understanding this line of code: new_string = input_string[:-1]. I'm really interested in this class and I'm eager to get started, but I'm a bit worried that this assignment will be too difficult for me.	STUDENT	Logan	Tatiana	Logan Hi everyone, excited about	you still find cor I'm Logan. I'm understanding	nfusing?
	either of you tell me what an index of -1 means? Jia I'm not sure. I know that an index of 0 typically means the beginning of a string, but I'm not sure what -1 would mean.	FORUM	(\cdot)		how to write for I'm still a bit of how to debug Tatiana	unctions, but onfused on	
	Lakisha I think -1 represents the last character in the string. That's exactly correct! An index of -1 represents the last character. What about the :? What does that mean in this line of code?	V STORIES	Shyla	Julia	Hi everyone, excited about classes and h them, but I'm on how to cre	I'm Tatiana. I'm the concept of now to use still confused ate a module.	
Chudu Dragmaaa	•••	ABOUT			Shyla		
Study Progress	Undo	2 760 online					4

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). GPTeach: Interactive TA Training with GPT Based Students. In *Proceedings of the 10th ACM Conference on Learning at Scale* (in press). ACM Press. https://doi.org/10.35542/osf.io/r23bu





Large language models for automated feedback



Providing feedback on written products

More readable than instructor-generated feedback Limited alignment with human instructor (positive vs negative) Provides traces of process level feedback

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y-S., Gašević, D., Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. In *Proceedings of the 23rd IEEE International Conference on Advanced Learning Technologies* (In press). IEEE.

LMMs (classification) for writing analytics

🛆 🔿 🛛 🗕 🛛 Edit mode 🔘 FLORA Home My courses Site administration × 3: Future Learning / General Instructions **General Instructions** General Instructions All this is Rubric Page Settings Morebased on ≚ <u>1: Artificial Intelligence i</u>... 1.1 Definition of Artificial I.. LLM In this learning session, the goal is to write a vision essay that describes the future of education. Please 1.2 History of Artificial Intel. describe, in 200 to 400 words, how you envision learning in a school in 2035. classification 1.3 How does AI work? Please consult the materials in this learning environment that provide information about three important topics for envisioning the future of education in 2035. 1.4 Ethics and risks of deve... Checklist Tool Essav Writing 222 words 1.5 Supervised machine lea... Ο Normal \Rightarrow B I \bigcup \circledast $x_2 x^2 \boxminus$ \boxminus \boxminus \blacksquare \blacksquare \blacksquare \blacksquare Basic Academic Originality Integration and elaboration 1.6 Unsupervised machine ... Nowadays, the role of AI in education is limited in some tangible educational products. More and more data are 1.7 Reinforcement learning needed to make Al work better. And the scaffolding is that the work is shared alone between the learner and some Essay Sentence: Nowadays, the role of AI in education is limited in some tangible more knowledgeable other or agent. In addition, it also enables learning from the experience. Scaffolding to optimize educational products. learning needs to acquire more knowledge. Besides, as we all know, differentiation is a combination of careful Integration Level: Remember and understanding 1.8 Deep Learning progress monitoring and adapting instruction in response. By this, students or learners could be divided into different groups by their teache according to their competence degree. In these groups, students can learn knowledge that they don't understand and make progress belonging to themselves. Differentiation practices in the classroom will help Essay Sentence: More and more data are needed to make AI work better. students in their group process. Integration Level: Evaluate and create Integration of three topics is very essential and all three are indispensable. In the future education, the three 2.1 What is Differentiation? should complement each other. Respectively, there are great expectations for AI in education to help students and teachers work better. To facilitate our independent and appropriate use in education in the future, we should learn 2.2 Using differentiation to ... more on how the scaffolding component skills work and so on. Furthermore, differentiation should be paid more Essay Sentence: And the scaffolding is that the work is shared alone between the attention in education. Beyond that, the three should continue to innovate, so as to better serve the education and learner and some more knowledgeable other or agent. other industries in the future 2.3 Standards for teaching Integration Level: Evaluate and create ≚ <u>3: Scaffolding in Education</u> Essay Sentence: In addition, it also enables learning from the experience. 3.1 The development of sc... Integration Level: Apply and analyse 3.2 What is cognitive appr... Essay Sentence: Scaffolding to optimize learning needs to acquire more knowledge 3.3 What is scaffolding This analysis may take around 5-10 min to process. 3.4 Applications of scaffold. 3.5 Applications of cognitiv...

http://floraproject.org

LMMs (classification) for writing analytics

FLORA Home My courses Site administration



C L A M

×

Rubric

3.5 Applications of cognitiv...

Chatbot to provide writing guidance

🗘 🔎 AU – Edit mode 🌑 FLORA Home My courses Site administration ◀ 3: Future Learning / General Instructions Q B **General Instructions** General Instructions Page Settings More-這 ≚ <u>1: Artificial Intelligence i</u>... 6 1.1 Definition of Artificial I.. 0 In this learning session, the goal is to write a vision essay that describes the future of education. Please 1.2 History of Artificial Intel. describe, in 200 to 400 words, how you envision learning in a school in 2035. 1.3 How does AI work? Please consult the materials in this learning environment that provide information about three important topics for envisioning the future of education in 2035. 1.4 Ethics and risks of deve... a la sue e constante × Checklist Tool 1.5 Supervised machine lea... Essay Writing 226 words 1.6 Unsupervised machine ... Normal \Rightarrow B I U \otimes X₂ X² \equiv \equiv = \rightarrow ¶ \equiv \equiv Basic Academic Originality Integration and elaboration 1.7 Reinforcement learning Nowadays, the role of AI in education is limited in some tangible educational products. More and more data are Sentence Text: Nowadays, the role of AI in education is limited in some tangible needed to make AI work better. And the scaffolding is that the work is shared alone between the learner and some more knowledgeable other or agent. In addition, it also enables learning from the experience. Scaffolding to optimize educational products. 1.8 Deep Learning learning needs to acquire more knowledge. Besides, as we all know, differentiation is a combination of careful Error Position: limited in progress monitoring and adapting instruction in response. By this, students or learners could be divided into different Error Correction: limited to groups by their teache according to their competence degree. In these groups, students can learn knowledge that Error Explanation: The preposition "in" is incorrect in this context. The correct they don't understand and make progress belonging to themselves. Differentiation practices in the classroom will help preposition is "to" when referring to limitations. 2.1 What is Differentiation? students in their group process Error Type: preposition Integration of three topics is very essential and all three are indispensable. In the future education, the three should complement each other. Respectively, there are great expectations for AI in education to help students and 2.2 Using differentiation to ... teachers work better. To facilitate our independent and appropriate use in education in the future, we should learn more on how the scaffolding component skills work and so on. Furthermore, differentiation should be paid more Sentence Text: More and more data are needed to make AI work better. 2.3 Standards for teaching attention in education. Beyond that, the three should continue to innovate, so as to better serve the education and Error Position: no error other industries in the future. Ai is good now. Error Correction: No error ≚ <u>3: Scaffolding in Education</u> Error Explanation: No error Error Type: No error 3.1 The development of sc... Sentence Text: And the scaffolding is that the work is shared alone between the 3.2 What is cognitive appr... learner and some more knowledgeable other or agent. Error Position: And the scaffolding is 3.3 What is scaffolding This analysis may take around 5-10 min to process. 3.4 Applications of scaffold.

http://floraproject.org

Chatbot to provide writing guidance

FLORA Home My courses Site administration ΔO AU mode 🔘 × 3: Future Learning / General Instructions **Checklist Tool** ✓ Instructions **General Instruc** General Instructions Integration and elaboration Basic Academic Originality Rubric Page Settings More-≚ <u>1: Artificial Intelligence i</u>... are Sentence Text: Nowadays, the role of AI in education is limited in some tangible and some 1.1 Definition of Artificial I.. In this learning session, the goalig to optimize educational products. 1.2 History of Artificial Intel. describe, in 200 to 400 words, careful Error Position: limited in 1.3 How does AI work? Please consult the materials i ed into different Error Correction: limited to for envisioning the future of owledge that Error Explanation: The preposition "in" is incorrect in this context. The correct 1.4 Ethics and risks of deve... assroom will help preposition is "to" when referring to limitations. 1.5 Supervised machine lea... Error Type: preposition tion, the three 1.6 Unsupervised machine ... students and 1.7 Reinforcement learning should learn Sentence Text: More and more data are needed to make AI work better. paid more 1.8 Deep Learning eds to acquire more kn ucation and Error Position: no error progress monitoring and adapting i groups by their teache according to the Error Correction: No error they don't understand and make prog 2.1 What is Differentiation? students in their group process Error Explanation: No error Integration of three topics is very should complement each other. Respec Error Type: No error 2.2 Using differentiation to ... eachers work better. To facilitate our indep more on how the scaffolding component skil 2.3 Standards for teaching attention in education. Beyond that, the three s other industries in the future. At is good no ≚ <u>3: Scaffolding in Education</u> Sentence Text: And the scaffolding is that the work is shared alone between the 3.1 The development of sc... learner and some more knowledgeable other or agent. Error Position: And the scaffolding is 3.2 What is cognitive appr... Error Correction: The scaffolding concent involves 3.3 What is scaffolding This analysis may take around 5-10 min to p 3.4 Applications of scaffold. 3.5 Applications of cognitiv... http://floraproject.org





Assessment in the age of artificial intelligence



Opportunities provided by AI





Opportunities provided by AI





Feasible assessment

Writing assessment

Oliveira, H., Ferreira Mello, R., Barreiros Rosa, B. A., Rakovic, M., Miranda, P., Cordeiro, T., ... & Gasevic, D. (2023). Towards explainable prediction of essay cohesion in Portuguese and English. In *Proceedings of the13th International Learning Analytics and Knowledge Conference* (pp. 509-519).



Model Performance (Prediction)



Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D. (2023). Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review. Submitted for review. https://arxiv.org/abs/2303.13379

Assessment support tools

Rubric Score

Dashboard

Dashboard

Essay Analysis

Essay

Nowadays, the role of AI in education is limited in some tangible educational products. More and more data are needed to make AI work better. And the scaffolding is that the work is shared alone between the learner and some more knowledgeable other or agent. In addition, it also enables learning from the experience. Scaffolding to optimize learning needs to acquire more knowledge. Besides, as we all know, differentiation is a combination of careful progress monitoring and adapting instruction in response. By this, students or learners could be divided into different groups by their teache according to their competence degree. In these groups, students can learn knowledge that they don't understand and make progress belonging to themselves. Differentiation practices in the classroom will help students in their group process.

Integration of three topics is very essential and all three are indispensable. In the future education, the three should complement each other. Respectively, there are great expectations for Al in education to help students and teachers work better. To facilitate our independent and appropriate use in education in the future, we should learn more on how the scaffolding component skills work and so on. Furthermore, differentiation should be paid more attention in education. Beyond that, the three should continue to innovate, so as to better serve the education and other industries in the future.

Result				
CRITERIA	WORD/ERROR	RANKING	SCORE	DATE
Word count	227	51.85	2	2023-06-11
Basic writing skills	10	90.12	0	2023-06-11
Academic writing skills	11	91.36	0	2023-06-11
Originality	2	86.42	0	2023-06-11
Total		34.57	2	2023-06-11

Rubric

• Word count: The essay consists of 200 to 400 words; Yes (2 points), No (0 points)

 Basic writing skills: The essay is clearly a mature draft, has no low-level writing mistakes, such as missing texts, 'placeholders', messy typography, many spelling and grammatical errors; Yes (2 points), Partial (1 point) No (0 point)

Academic writing skills: The writing of this essay should conform to the norms of academic writing, such as using appropriate logic structure, good flow and linkers usage, correct verbs and tenses and voices, consistent with academic writing style; Yes (4 points), Partial (1-3 point) No (0 point)
Originality: Your writing should be your own opinion elaborated in your own words, not simply copy-pasted sentences from the material; Yes (2 points), Partial (1 point) No (0 point)

http://floraproject.org



Continuous assessment

Process measurement is the dark side of assessment

Chatbot to provide writing guidance

FLORA Home My courses Site administration

∧ ∧ All - Edit mode

my courses site administration		
	3: Future Learning / General Instructions	,
	PAGE	
	General Instructions	
	Page Settings More -	
	In this learning session, the goal is to write a vision essay that describes the future of education. Please describe, in 200 to 400 words, how you envision learning in a school in 2035.	
	Please consult the materials in this learning environment that provide information about three important for	pics
	for envisioning the future of education in 2035.	
	Essay Writing 236 words × ChatGP	'T Tool
	Normal \Rightarrow B I U \Rightarrow x ₂ x ² \models \Rightarrow \uparrow \equiv \Rightarrow \uparrow \uparrow \equiv \Rightarrow	Can you give more advice on how to
	Al in Education. Ai plays an important role in Education.	18:20:01
	needed to make Al work better. And the scaffolding is that the work is shared alone between the learner and some more knowledeable other or agent In addition, it also enables learning from the experience. Scaffolding to optimize Certainly!	Here are some additional
	learning needs to acquire more knowledge. Besides, as we all know, differentiation is a combination of careful suggestion progress monitoring and adapting instruction in response. By this, students or learners could be divided into different	ns to improve your essay:
	groups by their teache according to their competence degree. In these groups, students can learn knowledge that they don't understand and make progress belonging to themselves. Differentiation practices in the classroom will help	on the role of AI in education:
	students in their group process. Discuss sp Integration of three topics is very essential and all three are indispensable. In the future education, the three be applied	d to learning, such as
	should complement each other. Kespectively, there are great expectations for Al in education to help students and personaliz teachers work better. To facilitate our independent and appropriate use in education in the future, we should learn more on but the scoreding component cilling work and co on Europerson differentiation should be additionant assessment assessment and the score additionant and the score additionant and the score additionant assessment assessment and the score additionant additiona	ed learning, adaptive nts, and intelligent tutoring
	attention in education. Beyond that, the three should continue to innovate, so as to better serve the education and systems. E other industries in the fullure Bai is grand how.	ixplain how these applications
	education education	in 2035.
	2. Elabora	te on scaffolding to optimize
	learning: F	rovide examples of how
	support le	arners in acquiring new
	Ask a	question
	Save Essay	Sen

http://floraproject.org

Chatbot to provide writing guidance

FLoRA Home My courses Site administration




Mapping trace data to processes



van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Bannert, M., ... & Molenaar, I. (2021). Do instrumentation tools capture self-regulated learning?. In Proceedings of the 11th International Learning Analytics and Knowledge Conference (pp. 438-448).





Process assessment in essay scoring



Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., ... & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning, 39*, 3, 703-718.



Process assessment



Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., ... & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning, 39*, 3, 703-718.



Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., ... & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning, 39*, 3, 703-718.



Process assessment



Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., ... & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning*, *39*, 3, 703-718.



Modern assessment

What skills should we measure in the times of generative AI?

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075.



Modern assessment

Computers and Education: Artificial Intelligence 4 (2023) 100140



Computers and Education: Artificial Intelligence

Contents lists available at ScienceDirect

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Large language models can create written products that score highly

Can large language models write reflectively

Yuheng Li^{a,1}, Lele Sha^{a,1}, Lixiang Yan^a, Jionghao Lin^a, Mladen Raković^a, Kirsten Galbraith^b, Kayley Lyons^c, Dragan Gašević^a, Guanliang Chen^{a,*}

^a Centre for Learning Analytics, Monash University, Australia

^b Experiential Development and Graduate Education, Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Australia

° Centre for Digital Transformation of Health, University of Melbourne, Australia

Criteria (max. mark)	Description	Human	ChatGPT
Overall (6)	Overall mark of all 6 criterias.	2.900 (1.7242)	4.300 (0.9966)
Returning to Experience (1)	Experience was clearly described (e.g., chronological information or personal judgements).	0.6500 (0.4810)	0.9500 (0.2198)
Attending to Feelings (1)	Personal feelings were described with judgements/reasons provided.	0.6167 (0.4903)	0.7833 (0.4155)
Association (1)	Links between prior knowledge, feelings or attitudes, and newly acquired knowledge.	0.6500 (0.4810)	0.9333 (0.2515)
Integration (1)	Association between prior and new knowledge, feelings or attitude, and new insights.	0.5500 (0.5017)	0.8667 (0.3428)
Validation (1)	Self-assessment of the new insights provided, with reference to prior experience.	0.0167 (0.1291)	0.0500 (0.2198)
Appropriation (1)	New insights related to current life and/or future development.	0.4167 (0.4972)	0.7167 (0.4544)

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., ... & Chen, G. (2023). Can Large Language Models Write Reflectively. *Computers and Education: Artificial Intelligence*, 4, 100140.



PROMISE – CONCERN – NEED – FINAL REMARKS







Limitations

Generative AI techniques are not flawless

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1-38.



No reasoning and planning, no sense of truthfulness, no temporal and spatial awareness, no casual inference, no common sense, no comprehension



Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.



A Categorical Archive of ChatGPT Failures

Ali Borji Quintic AI aliborji@gmail.com

April 5, 2023

Abstract

Large language models have been demonstrated to be valuable in different fields. ChatGPT, developed by OpenAI, has been trained using massive amounts of data and simulates human conversation by comprehending context and generating appropriate responses. It has garnered significant attention due to its ability to effectively answer a broad range of human inquiries, with fluent and comprehensive answers surpassing prior public chatbots in both security and usefulness. However, a comprehensive analysis of ChatGPT's failures is lacking, which is the focus of this study. Eleven categories of failures, including reasoning, factual errors, math, coding, and bias, are presented and discussed. The risks, limitations, and societal implications of ChatGPT are also highlighted. The goal of this study is to assist researchers and developers in enhancing future language models and chatbots. Please refer to here for the list of questions.

Borji, A. (2023). A Categorical Archive of ChatGPT Failures. arXiv:2302.03494v7





Workload implications of artificial intelligence in education



Aspiration

Al will increase productivity to focus on things that matter the most



TECHNOLOGY

ChatGPT Is About to Dump More Work on Everyone

Productivity != Workload

How do we deal with increased complexity of working and learning with AI?

Bogost, I. (2023, Feb 3). ChatGPT Is About to Dump More Work on Everyone, The Atlantic, https://www.theatlantic.com/technology/archive/2023/02/chatgptai-detector-machine-learning-technology-bureaucracy/672927/



The Washington Post Democracy Dies in Darkness

Subscribe

Can ChatGPT help me at the office? We put the AI chatbot to the test.

The AI chatbot ChatGPT can produce conversational responses to questions, but how well does it do with workplace tasks?



<u>=9</u>

Updated February 2, 2023 at 7:20 p.m. EST | Published February 2, 2023 at 6:00 a.m. EST

"It helped, but sometimes its errors caused more work than doing the task manually."

Abril, D. (2023, Feb 2). Can ChatGPT help me at the office? We put the AI chatbot to the test. Washington Post, https://www.washingtonpost.com/technology/2023/02/02/chatgpt-work-ai-uses/





Subscribe | Media | Open Calls

ISSUE DATE April 2023

 Research
 Programs & Projects
 Conferences
 Affiliated Scholars
 NBER News
 Career Resources
 About

 Home
 >
 Research
 >
 Working Papers
 >
 Generative AI at Work

Generative AI at Work

Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond

WORKING PAPER 31161 DOI 10.3386/w31161

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.



Workload

Productivity != Workload

Automatic feedback is not associated with decreased teaching load

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2, 100027.





Ethical implications of artificial intelligence in education



Privacy issues

LLMs can negatively affect student learning experiences

Deferring decisions made by such models Labelling AI-generated content with a warning message

Schneider, J, Richner, R., & Riser, M. (2023). Towards trustworthy autograding of short, multi-lingual, multi-type answers. *International Journal of Artificial Intelligence in Education*, 33(1), 88–118.



LLMs-based System Transparency



Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D. (2023). Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review. *Submitted for review*. https://arxiv.org/abs/2303.13379



Consent

No good mechanism to track where and how data are used

Giving students control

Baker, R. S., Hutt, S., Bosch, N., Ocumpaugh, J., Biswas, G., Paquette, L., Munshi, A., & Nasiar, N. (2023). Data-Driven Classroom Interviewing: Data Science to Target Qualitative Research. Edarxiv pre-print. https://doi.org/10.35542/osf.io/xsb4a

COLAM

Consent



Q Search analysis, research, academics...

Academic rigour, journalistic flair

Arts + Culture Books + Ideas Business + Economy Education Environment + Energy Health Politics + Society Science + Tech

ChatGPT: lessons learned from Italy's temporary ban of the Al chatbot Published: April 20, 2023 9.54pm AEST

The Italian data protection authority used an emergency procedure to impose the temporary ban. Shutterstock / Diego Thomazini

- In March 2023, Italy became the <u>first western country</u> to block
- Twitter 31 the advanced chatbot known as ChatGPT.
- Facebook 97
- in LinkedIn

Print

- The Italian data protection authority, <u>Garante</u>, cited concerns over the <u>protection of personal data</u> when making this decision.
 - It has given <u>OpenAI</u>, the California-based company that created <u>ChatGPT</u>, until the end of April to comply with its demands.

Authors



Oreste Pollicino Professor of Constitutional Law, Bocconi University



Giovanni De Gregorio

PLMJ chair in law and technology at Católica Global School of Law and Católica Lisbon School of Law and academic fellow, Bocconi

Gregorio, G. D., & Pollicino, O. (2023, April 20). *ChatGPT: Lessons learned from Italy's temporary ban of the AI chatbot*. The Conversation. <u>http://theconversation.com/chatgpt-lessons-learned-from-italys-temporary-ban-of-the-ai-chatbot-203206</u>



Consent

Hard to *unlearn* something once private data are used for training LLMs



Consent

Difficulty to know when something private may come up





Fairness

Are large language models fair?

Sha, L., Li, Y., Gasevic, D., & Chen, G. (2022). Bigger Data or Fairer Data? Augmenting BERT via Active Sampling for Educational Text Classification. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 1275-1285).

COLAM

Large language models

Nationality

editors (20%) reside in the United States, followed by Germany (12%) and Russia (7%). The only country not in Europe or North America in the top 10 is India (3%).

Gender 84 / 100 The 2013 study The Wikipedia Gender Gap Revisited ☑ measured gender bias in survey completion and estimated that as of 2008, 84% of English Wikipedia editors were male. In the worldwide Wikipedia Editor Survey 2011 of all the Wikipedias, 91% of respondents were male.

COLAM

Equity

Accessibility



Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D. (2023). Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review. *Submitted for review*. https://arxiv.org/abs/2303.13379



Bias

Midjourney v4

Prompt: patriotic, dog, superhero.





Bias in LMMs

Row	Methods	T_{max}	First-language backgrounds			Gender			
ID			LR-Demo	LR-Task		LR-Demo	LR-Task		
			↓ AUC	↑ AUC	\downarrow ABROCA	\downarrow AUC	↑ AUC	↓ ABROCA	
1	w/o pretraining	-	0.686	0.869	0.086	0.591	0.882	0.057	
2	Ramdom	1	0.692 (-0.87%)	0.876 (0.81%)	0.098 (-13.95%)	0.611 (-3.38%)	0.892 (1.13%)	0.089 (-56.14%)	
3	Equal		0.670 (2.33%)	0.883 (1.66%)	0.079 (8.14%)	0.595 (-0.68%)	0.889 (0.84%)	0.066 (-15.79%)	
4	AL-QBC		0.591 (13.85%)	0.879 (1.15%)	0.105 (-22.09%)	0.559 (5.41%)	0.889 (0.77%)	0.059 (-3.51%)	
5	AL-LAL		0.589 (14.14%)	0.876 (0.85%)	0.069 (19.77%)	0.552 (6.60%)	0.898 (1.85%)	0.055 (3.51%)	
6	AL-LCC		0.573 (16.47%)	0.878 (1.01%)	0.055 (36.05%)	0.558 (5.58%)	0.891 (1.02%)	0.047 (17.54%)	
7	Ramdom		0.688 (-0.29%)	0.889 (2.30%)	0.112 (-30.23%)	0.588 (0.51%)	0.895 (1.47%)	0.072 (-26.32%)	
8	Equal	6	0.621 (9.48%)	0.889 (2.30%)	0.095 (-10.47%)	0.561 (5.08%)	0.889 (0.84%)	0.066 (-15.79%)	
9	AL-LCC		0.525 (23.47%)	0.891 (2.53%)	0.041 (52.33%)	0.534 (9.64%)	0.899 (1.96%)	0.031 (45.61%)	

AL – active learning; QBC – Query By Committee; LAL – Learning Active Learning; LCC – Least Confident Classification (LCC); T_{max} – the maximum times of additional BERT pretraining.

Sha, L., Li, Y., Gasevic, D., & Chen, G. (2022). Bigger Data or Fairer Data? Augmenting BERT via Active Sampling for Educational Text Classification. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 1275-1285).



Bias in LMMs

Row	Methods	$\left \begin{array}{c} T_{max}\end{array}\right $	First-language backgrounds			Gender			
ID			LR-Demo	LR-Task		LR-Demo	LR-Task		
			↓ AUC	↑ AUC	\downarrow ABROCA	↓ AUC	↑ AUC	↓ ABROCA	
1	w/o pretraining	-	0.686	0.869	0.086	0.591	0.882	0.057	
2	Ramdom	1	0.692 (-0.87%)	0.876 (0.81%)	0.098 (-13.95%)	0.611 (-3.38%)	0.892 (1.13%)	0.089 (-56.14%)	
3	Equal		0.670 (2.33%)	0.883 (1.66%)	0.079 (8.14%)	0.595 (-0.68%)	0.889 (0.84%)	0.066 (-15.79%)	
4	AL-QBC		0.591 (13.85%)	0.879 (1.15%)	0.105 (-22.09%)	0.559 (5.41%)	0.889 (0.77%)	0.059 (-3.51%)	
5	AL-LAL		0.589 (14.14%)	0.876 (0.85%)	0.069 (19.77%)	0.552 (6.60%)	0.898 (1.85%)	0.055 (3.51%)	
6	AL-LCC		0.573 (16.47%)	0.878 (1.01%)	0.055 (36.05%)	0.558 (5.58%)	0.891 (1.02%)	0.047 (17.54%)	
7	Ramdom	6	0.688 (-0.29%)	0.889 (2.30%)	0.112 (-30.23%)	0.588 (0.51%)	0.895 (1.47%)	0.072 (-26.32%)	
8	Equal		0.621 (9.48%)	0.889 (2.30%)	0.095 (-10.47%)	0.561 (5.08%)	0.889 (0.84%)	0.066 (-15.79%)	
9	AL-LCC		0.525 (23.47%)	0.891 (2.53%)	0.041 (52.33%)	0.534 (9.64%)	0.899 (1.96%)	0.031 (45.61%)	

AL – active learning; QBC – Query By Committee; LAL – Learning Active Learning; LCC – Least Confident Classification (LCC); T_{max} – the maximum times of additional BERT pretraining.

Sha, L., Li, Y., Gasevic, D., & Chen, G. (2022). Bigger Data or Fairer Data? Augmenting BERT via Active Sampling for Educational Text Classification. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 1275-1285).



Large language models

Need to estimate bias in modern LLMs on relevant downstream tasks

Bias is prompting is particularly concerning

COLAM

PROMISE – CONCERN – NEED – FINAL REMARKS







Developing and measuring AI literacy

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 4*, 100130.





Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv* preprint arXiv:2303.10130.



Need for AI literacy

The New York Times

Here's What Happens When Your Lawyer Uses ChatGPT

A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.



May 27, 2023

The lawsuit began like so many others: A man named Roberto Mata sued the airline Avianca, saying he was injured when a metal serving cart struck his knee during a flight to Kennedy International Airport in New York.

Weiser, B. (2023). Here's What Happens When Your Lawyer Uses ChatGP. https://www.nytimes.com/2023/05/27/nyregion/aviancaairline-lawsuit-chatgpt.html

This US lawyer used ChatGPT to research a legal brief with embarrassing results. We could all learn from his error

ABC RN / By Damien Carrick and Sophie Kesteven for the Law Report, with additional reporting from Reuters. Posted Sat 24 Jun 2023 at 11:15am



Carrick, D., & Kesteven, S. (2023, June 24). 'Use with caution': How ChatGPT landed this US lawyer and his firm in hot water. *ABC News*. https://www.abc.net.au/news/2023-06-24/us-lawyer-uses-chatgpt-toresearch-case-with-embarrassing-result/102490068



Challenge

What is Al literacy for learners and educators?

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).
Conceptualizing AI literacy



Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.

The capabilities for a world with AI



Markauskaite, L., Marrone, R., Poquet, S., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., de Laat, M., Buckingham Shum, S., Gašević, D., Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?. *Computers & Education: Artificial Intelligence*, *3*, 100056.

Human and artificial cognition



Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition. *Computers and Education: Artificial Intelligence, 3*, 100107.





Automation vs control



Shneiderman, B. (2022). Human-Centered Artificial Intelligence. Oxford University Press



How do we assess AI literacy?

Going beyond self-reports

Connections of processes and technology



Siadaty, M., Gašević, D., & Hatala, M. (2016). Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning. *Computers in Human Behavior*, *59*, 469-482.







Understanding impact on learning



Fluence is not a bliss for learning

How do we deal with fluency bias and 'lazy' metacognition?

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. Annual review of psychology, 64, 417-444.



Sensemaking paradox gets even more serious

Butcher, K. R., & Sumner, T. (2011). Self-directed learning and the sensemaking paradox. Human–Computer Interaction, 26(1-2), 123-159.



Writing is an important learning tool

What happens when much of that tool is offloaded to generative AI?



Asking quality questions is an important learning tool

How do we promote development skills for question asking?

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. G. McKeown & L. Kucan (Eds.) Bringing research to life (pp. 112-141). Guilford Press.



Methodological rigor is critical

We need to see sound evidence for all the claims made about the power of generative AI in education

Challenge



No, GPT4 can't ace MIT

What follows is a critical analysis of "Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models"

Paper page - Exploring the MIT Mathematics and EECS Curriculum		Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models	
Join the discussion on this paper page	2023	Sarah J. Sanay Sanah Yari Arr Sarah Arr Sarah Sarah Sarah Sarah Sarah Sarah Ang Sarah Sara	
https://huggingface.co/papers/2306.08997	15	Keith Tyser Zad Chin Yaara Hicke Bakata Ukermity Haraad Schemith Consoling Schwarz, Schwarz Schwarz,	

- This is a joint document written by three MIT EECS seniors (Class of 2024): <u>Raunak</u> Chowdhuri, Neil Deshmukh, and David Koplow.
- On June 24th, <u>Armando Solar-Lezama</u> (Professor in EECS and COO and Associate Director of CSAIL, MIT), <u>Tonio Buonassisi</u> (Professor of Mechanical Engineering, MIT), and <u>Yoon Kim</u> (Assistant Professor in EECS and CSAIL, MIT) released a public statement regarding the paper. **Please read it below.**

Chowdhuri, R., Deshmukh, N., & Koplow, D. (n.d.). No, GPT4 can't ace MIT. Notion. Retrieved 26 June 2023, from https://bit.ly/No-GPT4-can-t-ace-MIT



Limited understanding of chatbots in education based on LLMs

Beettle II: adaptable tutorial dialogue system



Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education*, 24, 284-332.



AutoTutor family of dialogue tutors



Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24, 427-469.

Chatbots for self-regulated learning



Guan, R., Raković, M., Chen, G., Gašević, D. (2023). How Educational Chatbots Support Self-Regulated Learning? A Systematic Review of the Literature. *Submitted for review*.

PROMISE – CONCERN – NEED – FINAL REMARKS



Need to broaden understanding of learning with, about, and despite gen Al



Al is unlikely to go away and we need to lead the change

www.graile.ai

Interested in participating in shaping further developments in AI in Education?



Global Research Alliance for AI in Learning and Education

www.graile.ai







The Future of Education in the Age of Generative Artificial Intelligence

Dragan Gašević

June 26th, 2023 Faculty of Education University of Hong Kong