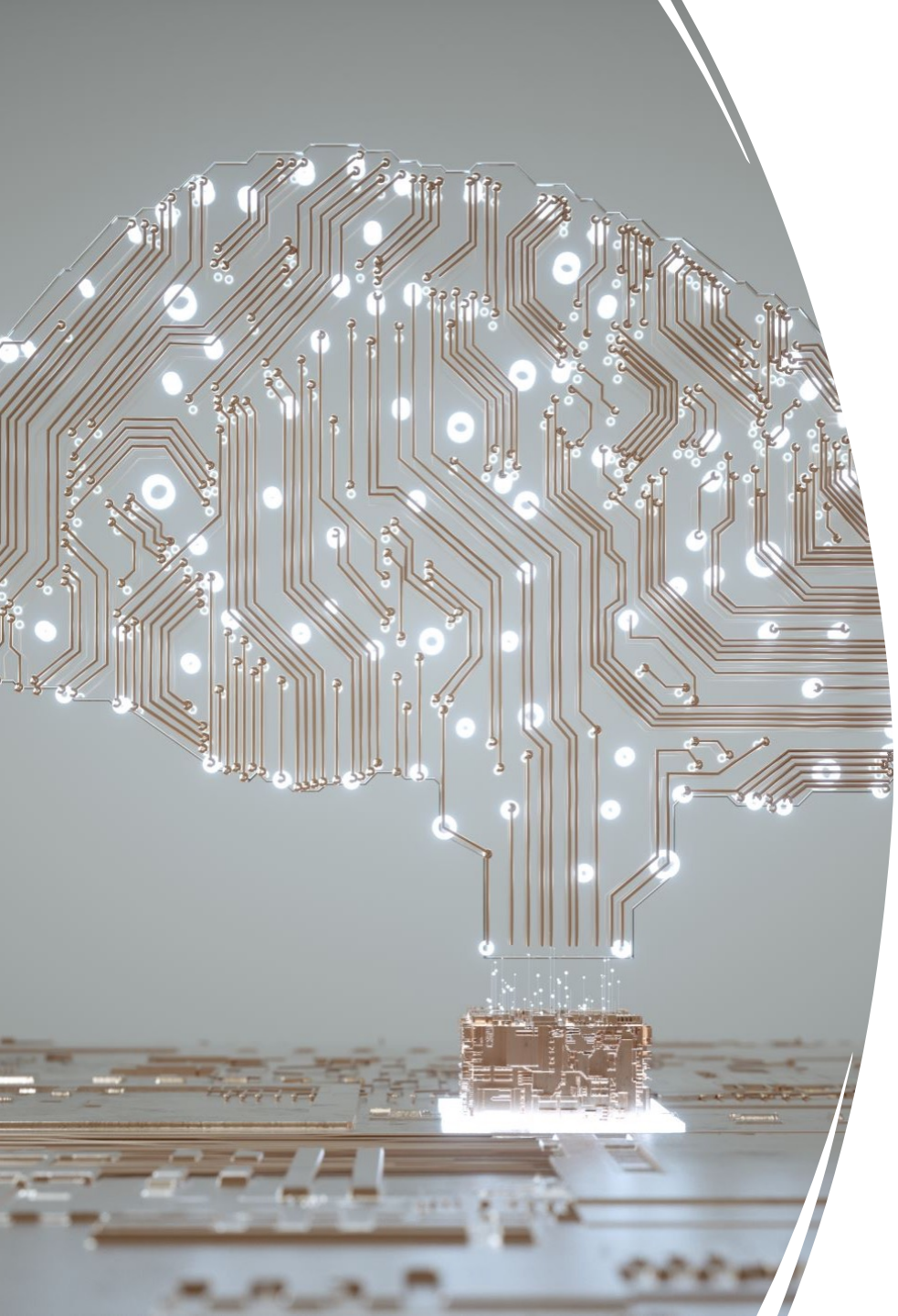


Artificial Intelligence? Augmented Intelligence? Augmented Humanity?

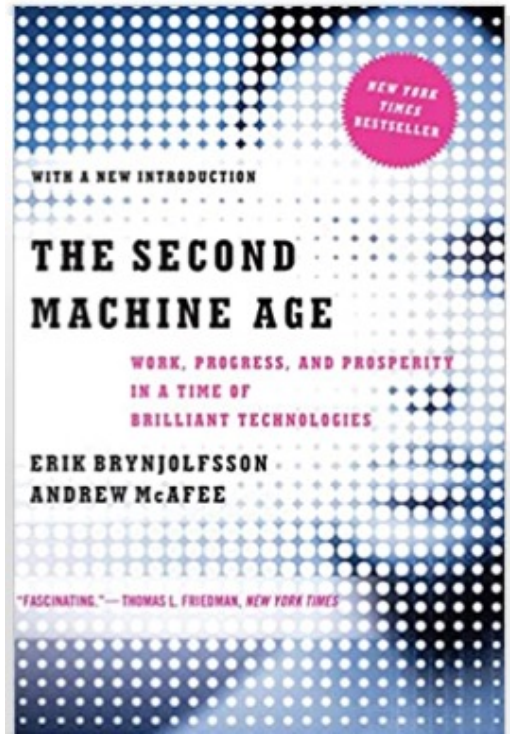
Norbert Wiener, MIT, ***God and Golem, Inc: A comment on certain points where cybernetics impinges on religion****, 1964, The world of the future will be an ever more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lay down to be waited upon by our robot slaves. (U.S. National Book Award Winner)*

EU Conference on Health and Social Security

March 6th, 2024

Koenraad Debackere, KU Leuven

Digital evolutions and revolutions?



sensors

MDPI

Systematic Review

Augmented Humanity: A Systematic Mapping Review

Graciela Guerrero [1,2], Fernando José Mateus da Silva [3], Antonio Fernández-Caballero [2,4,5] and António Pereira [3,6,*]

[1] Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Sangolquí 171103, Ecuador; rggaerrero@espe.edu.ec
[2] Instituto de Investigación en Informática de Albacete, 02071 Albacete, Spain; Antonio.Fdez@uclm.es
[3] Computer Science and Communication Research Center, School of Technology and Management, Polytechnic of Leiria, 2411-901 Leiria, Portugal; fernando.silva@ipleiria.pt
[4] Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, 02071 Albacete, Spain
[5] CIBERSAM (Biomedical Research Networking Centre in Mental Health), 28029 Madrid, Spain
[6] Information and Communications Technologies Unit, INOV INESC Innovation, Delegation Office at Leiria, 2411-901 Leiria, Portugal
* Correspondence: apereira@ipleiria.pt; Tel.: +351-244-820-300

Abstract: Augmented humanity (AH) is a term that has been mentioned in several research papers. However, these papers differ in their definitions of AH. The number of publications dealing with the topic of AH is represented by a growing number of publications that increase over time, being high impact factor scientific contributions. However, this terminology is used without being formally defined. The aim of this paper is to carry out a systematic mapping review of the different existing definitions of AH and its possible application areas. Publications from 2009 to 2020 were searched in Scopus, IEEE and ACM databases, using search terms "augmented human", "human augmentation" and "human 2.0". Of the 16,914 initially obtained publications, a final number of 133 was finally selected. The mapping results show a growing focus on works based on AH, with computer vision being the index term with the highest number of published articles. Other index terms are wearable computing, augmented reality, human–robot interaction, smart devices and mixed reality. In the different domains where AH is present, there are works in computer science, engineering, robotics, automation and control systems and telecommunications. This review demonstrates that it is necessary to formalize the definition of AH and also the areas of work with greater openness to the use of such concept. This is why the following definition is proposed: "Augmented humanity is a human–computer integration technology that proposes to improve capacity and productivity by changing or increasing the normal ranges of human function through the restoration or extension of human physical, intellectual and social capabilities".

Keywords: systematic mapping review; augmented humanity; wearable computing; mixed reality; human–robot interaction; smart devices

check for updates

Citation: Guerrero, G.; da Silva, F.J.M.; Fernández-Caballero, A.; Pereira, A. Augmented Humanity: A Systematic Mapping Review. *Sensors* **2022**, *22*, 514. https://doi.org/10.3390/s22020514

Academic Editor: Uichin Lee

Received: 19 November 2021
Accepted: 6 January 2022
Published: 10 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Humans are increasingly dependent on technology. Technology has changed not only humans' behavior and values but also the way they think, communicate and act [1]. However, recent scientific discoveries and inventions have demonstrated that technology is also beginning to modify human capabilities, pushing them beyond their natural limits [2–4]. With the advance of technology, the interaction between humans and machines has been "improved", "augmented" or even "redesigned" [5–8]. This has made it not only interesting and intriguing but also viable and arising as a serious concept of scientific research and development [9,10]. A term related to this technology advancement is Augmented Humanity (AH).

The term AH was coined in 2010 at the Internationale Funk Ausstellung conference [11,12], indicating that different devices which at first glance seem unconnected

KU LEUVEN

KU Leuven, Faculty of Economics and Business
Dept. of Management, Strategy and Innovation (MSI)

Artificial Intelligence and Firm-level Productivity

Dirk Czarnitzki, Gastón P. Fernández and Christian Rammer

MSI Discussion Paper No. 2203

AI: foundations?

- 1950: Alan Turing, *Mind,* « Can a machine think ? » ➔ « Can a machine be linguistically indistinguishable from a human ? »
- 1956: Herbert Simon (Nobel laureate)
  - Bounded rationality versus rational decision-making (“satisficing” versus “optimising”)
  - The Logic Theory Machine
  - Heuristics and rule-based programming
- The power of computing power, data, transmission, IoT, …
- Deduction versus induction, complementary and reinforcing
- Tversky & Kahneman (Nobel laureate), Prospect Theory (in essence a theory on risk aversion, preferring certainty rather than probability, asymmetric risk perceptions with fear of loss dominant)

AI: what?

Classification and prediction

E.g. phenotyping patients in view of personalised medical treatments like immune therapy

E.g. predicting the value of derivatives

Supervised versus non-supervised machine learning

E.g. training and learning with known data, categories and labels

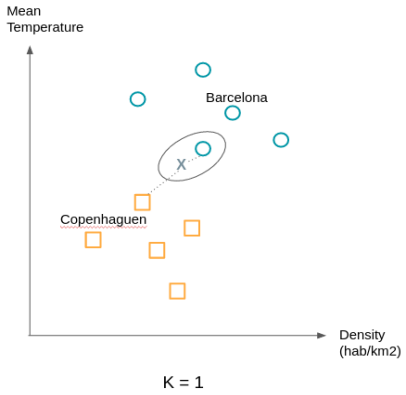
E.g. training and learning with missing data, categories and labels (“emerging”)

Algorithms: Machine Learning, Neural Networks (single/multilayered), deep learning

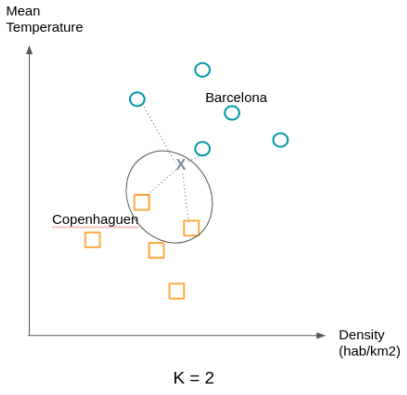
In search of “Explainable AI”, requiring design methodologies for AI applications

Algorithms, multiple and multifaceted

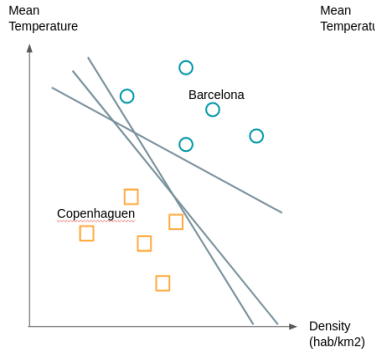
- Decision trees
- Random forests
- Gradient boosting
- Logistic regression
- Clustering, K-means
- K nearest neighbours
- Support Vector Machines
- Genetic algorithms
- ……

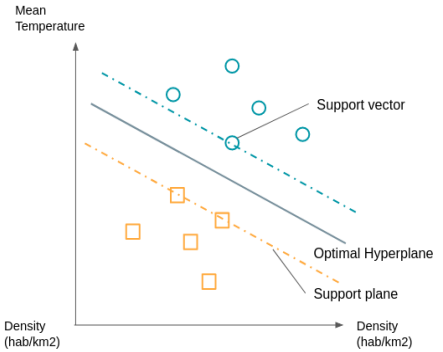


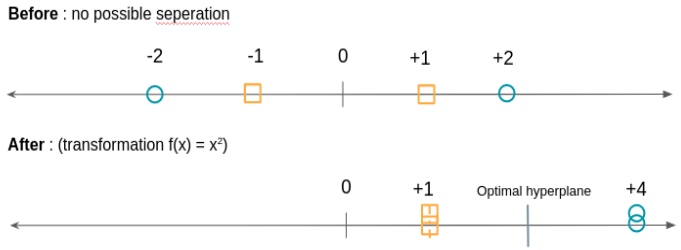
K = 1



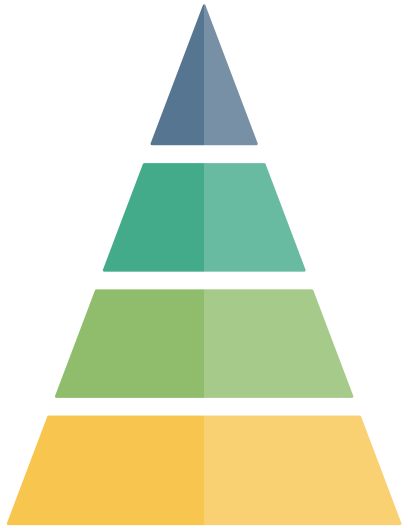
K = 2







European AI Act, challenge to identify and qualify results



Unacceptable risk

Limited set of AI uses that violate fundamental rights, are prohibited such as social scoring by governments, exploitation of vulnerabilities of children, use of subliminal techniques and biometric identification systems.

High

AI systems creating an adverse impact on fundamental rights and safety. Mandatory requirements are proposed to ensure trust and high level of protection of safety and fundamental rights.

Limited

Specific transparency requirements are imposed. Users should for instance be aware when they are interacting with a chatbot.

Minimal risk

All other AI systems can be developed and used subject to the existing legislation without additional legal obligations. Providers may voluntarily choose to apply the requirements for trustworthy AI.

ANNEX III

HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)

High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

1. Biometric identification and categorisation of natural persons:

(a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons;

...

4. Employment, workers management and access to self-employment:

(a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;

(b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.

5. Access to and enjoyment of essential private services and public services and benefits:

(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;

From "black box" to "explainable AI", understanding results obtained, a methodological necessity.

Robert Musil, 1930, Der Mann ohne Eigenschaften, meaning and humility.

3 questions

Legal?

Compliant?

Right?

BRIEF REPORT

AI-synthesized faces are indistinguishable from real faces and more trustworthy

Sophie J. Nightingale[a,1] and Hany Farid[b]

[a]Department of Psychology, Lancaster University, Lancaster LA1 4YW, United Kingdom; and [b]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by William Press, Computer Sciences and Integrative Biology, University of Texas at Austin, Austin, TX; received November 11, 2021; accepted December 20, 2021

Artificial intelligence (AI)-synthesized text, audio, image, and video are being weaponized for the purposes of nonconsensual intimate imagery, financial fraud, and disinformation campaigns. Our evaluation of the photorealism of AI-synthesized faces indicates that synthesis engines have passed through the uncanny valley and are capable of creating faces that are indistinguishable—and more trustworthy—than real faces.

deep fakes | face perception

Artificial intelligence (AI)-powered audio, image, and video synthesis—so-called deep fakes—has democratized access to previously exclusive Hollywood-grade, special effects technology. From synthesizing speech in anyone's voice (1) to synthesizing an image of a fictional person (2) and swapping one person's identity with another or altering what they are saying in a video (3), AI-synthesized content holds the power to entertain but also deceive.

Generative adversarial networks (GANs) are popular mechanisms for synthesizing content. A GAN pits two neural networks—a generator and discriminator—against each other. To synthesize an image of a fictional person, the generator starts with a random array of pixels and iteratively learns to synthesize a realistic face. On each iteration, the discriminator learns to distinguish the synthesized face from a corpus of real faces; if the synthesized face is distinguishable from the real faces, then the discriminator penalizes the generator. Over multiple iterations, the generator learns to synthesize increasingly more realistic faces until the discriminator is unable to distinguish it from real faces (see Fig. 1 for example real and synthetic faces).

Much has been written in the popular press about the potential threats of deep fakes, including the creation of nonconsensual intimate imagery (more commonly referred to by the misnomer "revenge porn"), small- to large-scale fraud, and adding jet fuel to already dangerous disinformation campaigns. Perhaps most pernicious is the consequence that, in a digital world in which any image or video can be faked, the authenticity of any inconvenient or unwelcome recording can be called into question.

Although progress has been made in developing automatic techniques to detect deep-fake content (e.g., refs. 4–6), current techniques are not efficient or accurate enough to contend with the torrent of daily uploads (7). The average consumer of online content, therefore, must contend with sorting out the real from the fake. We performed a series of perceptual studies to determine whether human participants can distinguish state-of-the-art GAN-synthesized faces from real faces and what level of trust the faces evoked.

Results

Experiment 1. In this study, 315 participants classified, one at a time, 128 of the 800 faces as real or synthesized. Shown in Fig. 2*A* is the distribution of participant accuracy (blue bars). The average accuracy is 48.2% (95% CI [47.1%, 49.2%]), close to chance performance of 50%, with no response bias: $d' = -0.09$; $\beta = 0.99$. Two repeated-measures binary logistic regression analyses were conducted—one for real and one for synthetic faces—to examine the effect of stimuli gender and race on accuracy. For real faces, there was a significant gender × race interaction, $\chi^2(3, N = 315) = 95.03, P < 0.001$. Post hoc Bonferroni-corrected comparisons revealed that mean accuracy was higher for male East Asian faces than female East Asian faces and higher for male White faces than female White faces. For synthetic faces, there was also a significant gender × race interaction, $\chi^2(3, N = 315) = 68.41, P < 0.001$. For both male and female synthetic faces, White faces were the least accurately classified, and male White faces were less accurately classified than female White faces. We hypothesize that White faces are more difficult to classify because they are overrepresented in the StyleGAN2 training dataset and are therefore more realistic.

Experiment 2. In this study, 219 new participants, with training and trial-by-trial feedback, classified 128 faces taken from the same 800 set of faces as in experiment 1. Shown in Fig. 2*A* is the distribution of participant accuracy (orange bars). The average accuracy improved slightly to 59.0% (95% CI [57.7%, 60.4%]), with no response bias: $d' = 0.46$; $\beta = 0.99$. Despite providing trial-by-trial feedback, there was no improvement in accuracy over time, with an average accuracy of 59.3% (95% CI [57.8%, 60.7%]) for the first set of 64 faces and 58.8% (95% CI [57.4%, 60.3%]) for the second set of 64 faces. Further analyses to examine the effect of gender and race on accuracy replicated the primary findings of experiment 1. This analysis again revealed that, for both male and female synthetic faces, White faces were the most difficult to classify.

When made aware of rendering artifacts and given feedback, there was a reliable improvement in accuracy; however, overall performance remained only slightly above chance. The lack of improvement over time suggests that the impact of feedback is limited, presumably because some synthetic faces simply do not contain perceptually detectable artifacts.

Experiment 3. Faces provide a rich source of information, with exposure of just milliseconds sufficient to make implicit inferences about individual traits such as trustworthiness (8). We wondered whether synthetic faces activate the same judgements of trustworthiness. If not, then a perception of trustworthiness could help distinguish real from synthetic faces.

In this study, 223 participants rated the trustworthiness of 128 faces taken from the same set of 800 faces on a scale of 1 (very untrustworthy) to 7 (very trustworthy) (9). Shown in Fig. 2*B* is the distribution of average ratings (by averaging the ordinal ratings, we are assuming a linear rating scale). The average rating for real

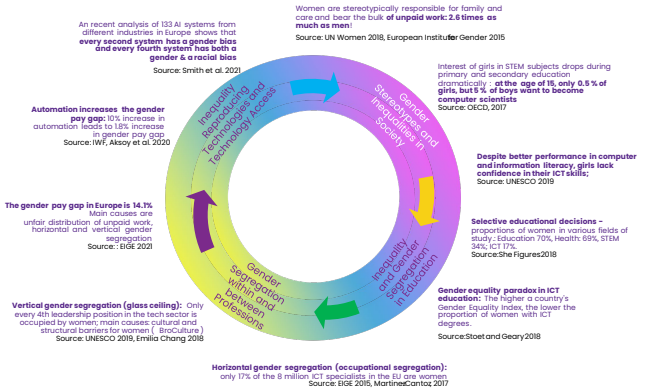
Author contributions: S.J.N. and H.F. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: s.nightingale1@lancaster.ac.uk.

Published February 14, 2022.

Figure 7: Vicious cycle of digital inequality with a summary of empirical facts



Source: Bruegel.

ABSTRACTS WRITTEN BY CHATGPT FOOL SCIENTISTS

Researchers cannot always differentiate between AI-generated and original abstracts.

By Holly Else

An artificial-intelligence (AI) chatbot can write such convincing fake research-paper abstracts that scientists are often unable to spot them, according to a preprint posted on the bioRxiv server in late December[1]. Researchers are divided over the implications for science.

"I am very worried," says Sandra Wachter, who studies technology and regulation at the University of Oxford, UK, and was not involved in the research. "If we're now in a situation where the experts are not able to determine what's true or not, we lose the middleman that we desperately need to guide us through complicated topics," she adds.

The chatbot, ChatGPT, creates realistic text in response to user prompts. It is a 'large language model', a system based on neural networks that learn to perform a task by digesting huge amounts of existing human-generated text. Software company OpenAI, based in San Francisco, California, released the tool on 30 November, and it is free to use.

Since its release, researchers have been grappling with the ethical issues surrounding its use, because much of the chatbot's output can be difficult to distinguish from human-written text. Scientists have published a preprint[2] and an editorial[3] written by ChatGPT. Now, a group led by Catherine Gao at Northwestern University in Chicago, Illinois, has used ChatGPT to generate artificial research-paper abstracts to test whether scientists can spot them.

The researchers asked the chatbot to write 50 medical-research abstracts based on a selection published in *JAMA*, *The New England Journal of Medicine*, *The BMJ*, *The Lancet* and *Nature Medicine*. They then compared these with the original abstracts by running them through a plagiarism detector and an AI-output detector, and asked a group of medical researchers to spot the fabricated abstracts.

Under the radar

The ChatGPT-generated abstracts sailed through the plagiarism checker: the median originality score was 100%, which indicates that no plagiarism was detected. The AI-output detector spotted 66% of the generated abstracts. But the human reviewers didn't do much better: they correctly identified only 68% of the generated abstracts and 86% of the genuine abstracts. They incorrectly identified 32% of the generated abstracts as being real and 14% of the genuine abstracts as being generated.

Wachter says that, if scientists can't determine whether research is true, there could be "dire consequences". As well as being problematic for researchers, who could be pulled down flawed routes of investigation, because the research they are reading has been fabricated, there are "implications for society at large because scientific research plays such a huge role in our society". For example, it could mean that research-informed policy decisions are incorrect, she adds.

But Arvind Narayanan, a computer scientist at Princeton University in New Jersey, says: "It is unlikely that any serious scientist will use ChatGPT to generate abstracts." He adds that whether generated abstracts can be detected is "irrelevant". "The question is whether the tool can generate an abstract that is accurate and compelling. It can't, and so the upside of using ChatGPT is minuscule, and the downside is significant," he says.

Irene Solaiman, who researches the social impact of AI at Hugging Face, an AI company with headquarters in New York and Paris, has fears about any reliance on large language models for scientific thinking. "These models are trained on past information and social and scientific progress can often come from thinking, or being open to thinking, differently from the past," she adds.

The authors suggest that those evaluating scientific communications, such as research papers and conference proceedings, should put policies in place to stamp out the use of AI-generated texts. If institutions choose to allow use of the technology in certain cases, they should establish clear rules around disclosure. This month, the Fortieth International Conference on Machine Learning — which will be held in Honolulu, Hawaii, in July — announced that it has banned papers written by ChatGPT and other AI language tools.

Solaiman adds that in fields where fake information can endanger people's safety, such as medicine, journals might have to take a more rigorous approach to verifying information as accurate.

Narayanan says that the solutions to these issues should not focus on the chatbot itself, "but rather the perverse incentives that lead to this behaviour, such as universities conducting hiring and promotion reviews by counting papers with no regard to their quality or impact".

"If the experts are not able to determine what's true, we lose the middleman to guide us through complicated topics."

Academics say this is just the beginning. The rising expectations of an emboldened labour movement were on full display on 23 December, when more than 35% of the members of two unions representing UC graduate students voted against accepting university officials' offer and ending the strike.

Missed opportunity?

One of the organizers of the vote-no campaign was Dylan Kupsh, a graduate researcher in computer science at UCLA. Kupsh was in close contact with union organizers at Columbia University, where student workers rejected an initial contract proposal and went on to secure further concessions after a ten-week strike that ended last January.

In the end, UC graduate students received a range of new benefits, including increased childcare subsidies; protections against bullying, discrimination and harassment; and a new schedule for salaries. Incoming graduate students, for example, will see their annual salary increase from around US$22,000 to $30,500. "We could have won a lot more, and it's sad that we didn't get there," Kupsh says. "We're going to have to repeat in another 2.5 years."

For Barry Eidlin, a sociologist at McGill University in Montreal, Canada, who studies the labour movement, the scale of the vote-no campaign is yet another sign of changing expectations in academia. "In the past, academic workers have felt like they should just keep their heads down and be grateful they have a job," he says. "The idea that people now expect more, and are willing to fight for more, seems to me a welcome shift in perspective."

1. Gao, C. A. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2022.12.23.521610 (2022).
2. Blanco-Gonzalez, A. *et al.* Preprint at https://arxiv.org/abs/2212.08104 (2022).
3. O'Connor, S. & ChatGPT. *Nurse Educ. Pract.* **66**, 103537 (2023).

VMware

VMware software powers the world's complex digital infrastructure. The company's cloud, app modernization, networking, security and digital workspace offerings help customers deliver any application on any cloud across any device. The company's culture and values are expressed through the acronym EPIC2: execution, passion, integrity, customers and community. VMware celebrates employees in its annual EPIC2 achievement awards. This honour is given to employees who best exemplify these values through their actions.

Integrity and ethics are embedded in everything they do, from the company culture to its product development processes. To help operationalize ethics into the organization, VMware's ethics and compliance team is creating an ethical decision-making framework called DECIDE to help employees determine solutions when faced with ethically ambiguous situations. The DECIDE model is a systematic process to evaluate potential solutions through multiple ethical lenses, driving an appreciation of diverse perspectives, and enhancing ethical problem-solving capabilities. As with its AI code of ethics, which was created in a grass-roots manner, VMware prioritizes ethics and its EPIC2 values at every level from its leadership to its 32,000-strong global workforce.

Challenges: inequity, fraud, …

But also… a myriad of opportunities… method of invention, robot scientist, augmented learning…

www.nature.com/scientificreports

scientific reports

OPEN

Best humans still outperform artificial intelligence in a creative divergent thinking task

Mika Koivisto[1] & Simone Grassini[2,3]

Creativity has traditionally been considered an ability exclusive to human beings. However, the rapid development of artificial intelligence (AI) has resulted in generative AI chatbots that can produce high-quality artworks, raising questions about the differences between human and machine creativity. In this study, we compared the creativity of humans (n = 256) with that of three current AI chatbots using the alternate uses task (AUT), which is the most used divergent thinking task. Participants were asked to generate uncommon and creative uses for everyday objects. On average, the AI chatbots outperformed human participants. While human responses included poor-quality ideas, the chatbots generally produced more creative responses. However, the best human ideas still matched or exceed those of the chatbots. While this study highlights the potential of AI as a tool to enhance creativity, it also underscores the unique and complex nature of human creativity that may be difficult to fully replicate or surpass with AI technology. The study provides insights into the relationship between human and machine creativity, which is related to important questions about the future of creative work in the age of AI.

The development and widespread availability of generative artificial intelligence (AI) tools, such as ChatGPT (https://openai.com/) or MidJourney (https://www.midjourney.com), has sparked a lively debate about numerous aspects of their integration into society[1], as well as about the nature of creativity in humans and AI[2]. One of the key issues surrounding the implementation of AI technologies pertains to their potential impact on the job market[3]. With AI systems becoming increasingly capable of performing tasks that were once solely within the purview of humans, concerns have been raised about the potential displacement of jobs and its implications for future employment prospects[4]. In the field of education, questions have been raised about the ethical and pedagogical implications of such technologies, as well as concerns about how AI systems might reduce critical thinking skills[5]. Another aspect of the debate involves the legal and ethical ramifications of AI-generated content[6,7]. As these tools produce increasingly sophisticated works, ranging from articles to artistic creations, it raises the issue of whether AI-generated products should be granted the same legal protections as human-created works, and how to assign responsibility and credit for such creations.

At the heart of these discussions lie fundamental questions about the nature of human identity and creativity, and how this identity interfaces with AI systems that seem capable of human-like creative production[2]. As AI technologies continue to advance, they challenge traditional notions of what it means to be human and force us to reconsider the unique qualities that define our species. For example, the concept of creativity, which has historically been attributed exclusively to conscious human beings[8,9], is now being reevaluated considering AI's ability to seemingly generate original content.

AI has shown tremendous potential for greater and more enormous possibilities in areas that require reasoning and creative decision making. This is demonstrated, for example, by the rise of chess engines, neural networks, and deep learning-based chess networks, which are capable of defeating chess masters (https://builtin.com/artificial-intelligence/chess-ai). Additionally, AI seems to perform well in art-related creativity. Recent AI tools can produce high-quality art pieces that have been bought for high prices[10], as well as poetry that is indistinguishable from human-made art[11]. These findings seem to suggest that AI is capable of creating products that humans typically perceive as creative. But what exactly is creativity?

Traditionally, creativity has been defined as the ability to produce ideas that are, to some extent, both original and useful[12]. This definition allows us to evaluate the creativity of AI's ideas using the same criteria applied to human ideas. In this study, we compare the products generated by AI and humans in the context of creative

[1]Department of Psychology, University of Turku, Turku, Finland. [2]Department of Psychosocial Science, University of Bergen, Bergen, Norway. [3]Cognitive and Behavioral Neuroscience Laboratory, University of Stavanger, Stavanger, Norway. email: simone.grassini@uib.no

Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King[1], Kenneth E. Whelan[1], Ffion M. Jones[1], Philip G. K. Reiser[1], Christopher H. Bryant[2], Stephen H. Muggleton[3], Douglas B. Kell[4] & Stephen G. Oliver[5]

[1]*Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK*
[2]*School of Computing, The Robert Gordon University, Aberdeen AB10 1FR, UK*
[3]*Department of Computing, Imperial College, London SW7 2AZ, UK*
[4]*Department of Chemistry, UMIST, P.O. Box 88, Manchester M60 1QD, UK*
[5]*School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK*

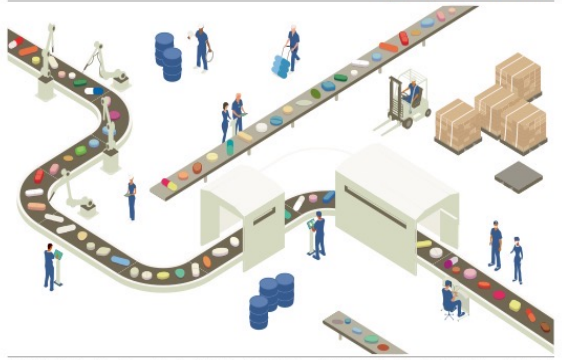
The question of whether it is possible to automate the scientific process is of both great theoretical interest[1,2] and increasing practical importance because, in many scientific areas, data are being generated much faster than they can be effectively analysed. We describe a physically implemented robotic system that applies techniques from artificial intelligence[3–8] to carry out cycles of scientific experimentation. The system automatically

Invention of a Method of Invention	General-Purpose Technology: NO	General-Purpose Technology: YES
NO	Industrial Robots (e.g. Fanuc R2000)	'Sense & React" Robots (e.g. Autonomous vehicles)
YES	Statically-coded Algorithmic Tools (e.g. fMRI)	**Deep Learning**

Source: Cockburn, Henderson & Stern, NBER, 2018

News feature

https://doi.org/10.1038/s41591-023-02361-0



INSIDE THE NASCENT INDUSTRY OF AI-DESIGNED DRUGS

Artificial intelligence tools are beginning to upend the drug discovery pipeline, with several new compounds entering clinical trials. **By Carrie Arnold**

Drug discovery is expensive, inefficient, and fraught with failure. An estimated 86% of drug candidates developed between 2000 and 2015 did not meet their stated endpoints.

Despite this challenge, the use of artificial intelligence (AI) and machine learning to understand drug targets better and synthesize chemical compounds to interact with them has not been easy to sell. Alex Zhavoronkov would know. When the CEO and founder of Insilico Medicine, with offices in Hong Kong and New York, first started trying to raise funding nearly a decade ago, he struggled to find others who shared his vision.

"It was such a grand goal, but every time I went to a venture capitalist, they never gave me money," says Zhavoronkov.

Even as recently as 5 years ago, his presentations had to explain to pharma collaborators why AI was so promising. Not anymore. Now he is at the forefront of drug discovery's AI nascent revolution.

"We've managed to get here in three years, and we didn't fail. And we did it multiple times," Zhavoronkov says.

The persistence of Zhavoronkov and a small cadre of other startup founders, including Exscientia's Andrew Hopkins and BenevolentAI's Bryn Williams-Jones, means that not only are some of the biggest players in pharma already convinced of the utility of AI in drug development, but also some of these drugs are beginning their ultimate test in clinical trials (Table 1).

"In the last couple of years, AI has gone from being hypothetically interesting to real programs moving towards the clinic," says Williams-Jones. "There's no shortcuts to drug discovery. We can have better informed ideas, but you still have to go through the rest of the [development] process."

These trials are still in their early days, says Hopkins, so it is not yet clear which compound will cross the finish line first. But he is

Challenges and opportunities reinforcing one another

Check for updates

Association of Learned and Professional Society Publishers

ORIGINAL ARTICLE

(wileyonlinelibrary.com) doi: 10.1002/leap.1582

Received: 29 April 2023 | Accepted: 29 August 2023

Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals

Brady D. Lund [1*] and K.T. Naheem [2]

Brady D. Lund

K.T. Naheem

[1]Department of Information Science, University of North Texas, Denton, Texas, USA

[2]Central Library, Pondicherry Central University, Puducherry, India

ORCID:
B. D. Lund: 0000-0002-4819-8162

*Corresponding author: Brady D. Lund, Department of Information Science, University of North Texas, Denton, TX, USA.
E-mail: brady.lund@unt.edu

Abstract: Academic publishers have quickly responded to the impact of artificial intelligence (AI) tools on authorship and academic integrity. However, there remains a lack of understanding about AI authorship policies and the attitude of academic journals towards these tools. This study aims to address this gap by examining the AI authorship policies of 300 top academic journals during the period of late-spring 2023. Over half of the journals examined have an AI authorship policy and guidelines for acknowledging AI usage in manuscript preparation. These acknowledgments are typically made in the methods or acknowledgement sections, although some journals have introduced a new, special section on AI usage. The study also found that AI authorship policies may differ depending on the publisher and discipline of the journal. Many publishers have adopted uniform AI authorship policies that are implemented across all journals that they publish. These results are useful for publishers, editors, and researchers who want to learn more about how academic journals are dealing with the emergence of large language models and other AI tools in scholarly communications.

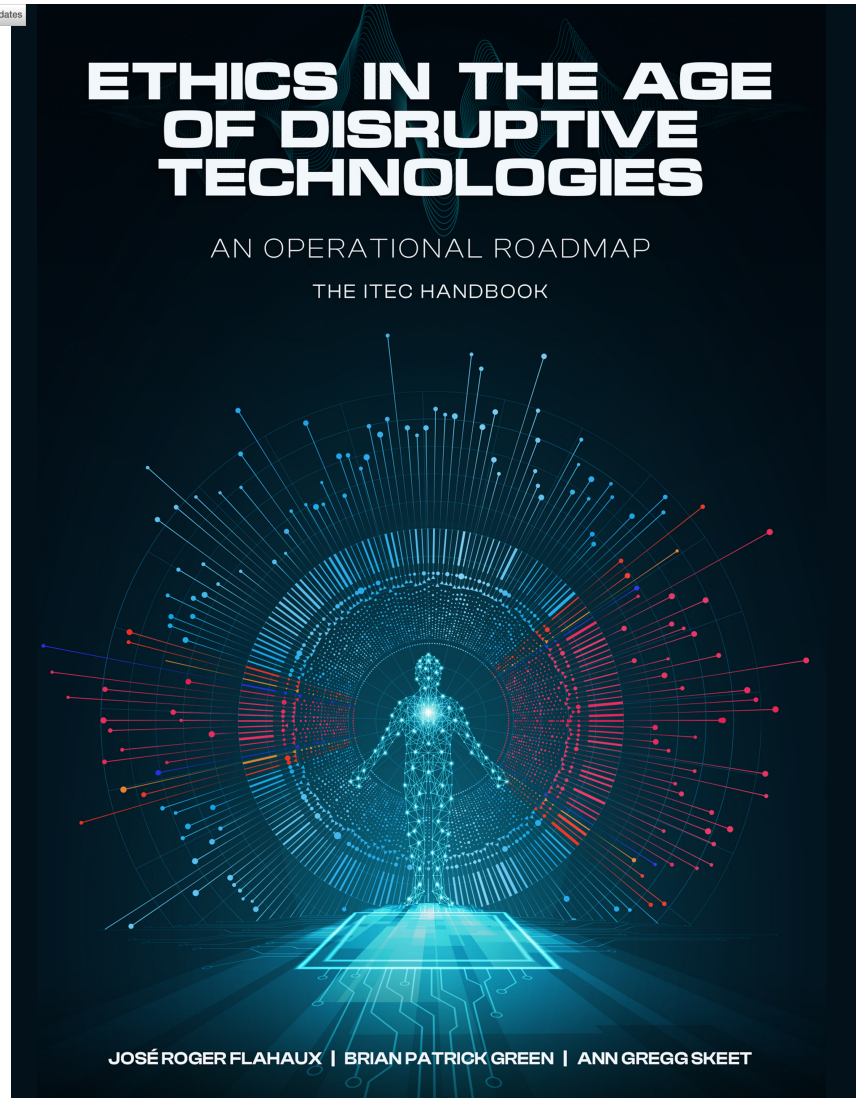
Keywords: academic journals, AI authorship, artificial intelligence, authorship policies, ChatGPT

INTRODUCTION

Use of artificial intelligence (AI) tools, especially those based on large language models like ChatGPT, has grown tremendously in the past several months. These tools have the potential to dramatically transform academic publishing, where they can be used positively to improve the quality of written works or abused to generate papers full of misinformation and phantom references (Foster, 2019). Justifiably, there is growing concern about the potential implications of these AI tools for authorship of academic manuscripts and the impact on the integrity of scientific publications. This study examines how academic journals have adjusted to the new academic reality of these AI tools, by analysing the AI authorship policies that have emerged among top publishers and how they guide the usage and acknowledgement of AI technologies. These findings should offer both clarity and guidance to other publishers, journal editors, and authors as they navigate this emerging landscape.

LITERATURE REVIEW

Generative AI—artificial intelligence applications capable of generating new content such as video, images, and text—can revolutionize scholarly writing and publishing (Liebrenz et al., 2023; Lin, 2023). The world appears to be situated on the precipice of



Working Paper 24-013

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

Fabrizio Dell'Acqua
Edward McFowland III
Ethan Mollick
Hila Lifshitz-Assaf
Katherine C. Kellogg
Saran Rajendran
Lisa Krayer
François Candelon
Karim R. Lakhani

Harvard Business School

Thank you.
Questions?

What If?

What if an AI won the Nobel prize for medicine?

Controversy ensues when the greatest prize in medical research is awarded to a non-human. An imagined scenario from 2036