Current Psychiatry Reports (2023) 25:839–846
https://doi.org/10.1007/s11920-023-01477-x

An Introduction to Generative Artificial Intelligence in Mental Health Care: Considerations and Guidance

Darlene R. King[1] · Guransh Nanda[2] · Joel Stoddard[3,4] · Allison Dempsey[4] · Sarah Hergert[1] · Jay H. Shore[4] · John Torous[5,6]

Accepted: 21 November 2023 / Published online: 30 November 2023

Abstract

Purpose of Review This paper provides an overview of generative artificial intelligence (AI) and the possible implications in the delivery of mental health care.

Recent Findings Generative AI is a powerful technology that is changing rapidly. As psychiatrists, it is important for us to understand generative AI technology and how it may impact our patients and our practice of medicine.

Summary This paper aims to build this understanding by focusing on GPT-4 and its potential impact on mental health care delivery. We first introduce key concepts and terminology describing how the technology works and various novel uses of it. We then dive into key considerations for GPT-4 and other large language models (LLMs) and wrap up with suggested future directions and initial guidance to the field.

Keywords Generative artificial intelligence · Artificial intelligence (AI) · Augmented intelligence (AI) · Mental health AI · ChatGPT in mental health

Introduction

Today, artificial intelligence (AI) is more relevant to the world than ever before. In 2022, Chat GPT, a user interface created by OpenAI, transformed access to AI models by allowing millions of users to interact with generative large language models (GPT 3.5, GPT-4). ChatGPT became the fastest-growing app in the world when it reached 100 million users within 2 months of launch [1]. The OpenAI website has been visited nearly 2 billion times so far and gets 25 million daily visits [1]. More than ever before, the world is abuzz with talk of AI. The rapid rate of proliferation, uptake, and the sheer power of this technology has raised many concerns. In 2023, top AI researchers advocated for increased regulatory oversight and open discussion to "mitigate the risk of extinction from AI" and "other societal-scale risks such as pandemics and nuclear war" [2]. Many researchers agree that there is a need for a set of standards that must be followed by all users and providers of AI [3, 4]. Important questions to be asked are the following: Is the current level of concern warranted? What are the real current and future risks of AI? What is the potential of AI and how will it impact the field of psychiatry? Why is it so important for

✉ Darlene R. King
Darlene.King@UTSouthwestern.edu

1 Department of Psychiatry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, MC 8849, Dallas TX 75390-8849, USA

2 The University of Texas, Southwestern Medical School, 5323 Harry Hines Blvd, Dallas, TX 75390-8830, USA

3 Children's Hospital Colorado Anschutz Medical Campus, Child and Adolescent Psychiatry, University of Colorado Anschutz Medical Campus, 13123 East 16th Ave, Aurora, CO 80045, USA

4 Department of Psychiatry and Family Medicine, School of Medicine, Centers for American Indian and Alaska Native Health, Colorado School of Public Health, Telemedicine Helen and Arthur E. Johnson Depression Center, University of Colorado Anschutz Medical Campus, Mail Stop F800, 13055 East 17th Avenue, Aurora CO 80045, USA

5 Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

6 Massachusetts Mental Health Center, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

psychiatrists to understand the risks, benefits, and state of artificial intelligence?

Brief Note on Artificial Intelligence vs. Augmented Intelligence

The American Medical Association (AMA) House of Delegates uses the term augmented intelligence (AI) as a conceptualization of artificial intelligence that focuses on an assistive role, emphasizing that its design enhances human intelligence rather than replaces it. In this article, we will use the two terms interchangeably.

Considerations for Generative AI in Psychiatry and Patient Care

More than half of US counties lack a single psychiatrist, and more than 150 million people live in federally designated mental health professional shortage areas (defined as more than 30,000 residents per clinician). Prior to the COVID-19 pandemic, 11% of adults reported symptoms of anxiety or depression. At the height of the pandemic, this percentage jumped to 40% as "powerful stressors eroded the well-being of communities across the country" [5]. Today, the percentage of adults reporting symptoms of anxiety and depression remains high (32% in Nov 2023) [5]. Access to appropriate treatment poses an ever-expanding challenge to both clinicians and patients. On the clinician side, the demand and need for care must be balanced with documentation, regulatory, and administrative tasks. Across all medical professions, physicians regularly spend more time on documentation in electronic health records than with patients [6]. This can be especially true in psychiatry where documentation includes unique aspects of a patient's symptoms and narrative that are not easily captured with one-size-fits-all documentation aids. This documentation burden has proven to have such a negative impact on health care that the American Medical Informatics Association (AMIA) has made it their goal to reduce documentation burden to 25% of current state in 5 years [6].

Enter ChatGPT: a technology which responds naturalistically to prompts, holds conversations, and generates professionally written text responses in a matter of seconds—work that would have taken sometimes hours for a human to research, synthesize, and produce. In the current medical landscape, the sensible course of action seems to be the following: how do we implement this technology to provide solutions we need? It is easy to imagine such a tool assisting with documentation as a digital scribe, formulating a preliminary differential diagnosis, assisting with personalized treatment plans, summarizing a patient's medical records, chatting with patients between visits to answer simple questions, and even providing patients with supportive therapy. Like other preceding technologies such as mental health apps, the potential for generative AI or LLMs to increase access to care, improve clinical outcomes, and strengthen the therapeutic alliance makes generative AI an exciting addition, but it also presents risks. Just as we must understand the risks and benefits of a medication or treatment, so must we understand the risks and benefits of generative AI, especially if we wish to incorporate it into clinical practice. We will address three major areas for risk consideration: Privacy/Security, Clinical Foundation/Treatment Goals, and Ethics/Legal Considerations. The goal of this discussion is to serve as a primer for future considerations and approaches to the utilization of generative AI.

History of Artificial Intelligence

Artificial intelligence has had a fascinating history over the past century with progressive and exciting advancements happening over time. We have come a long way from the first mentions of robots in 1921 in Karel Capek's Czech play "Rossum's Universal Robots" to deep learning chatbots such as Chat GPT today [7, 8]. Rapid advancements were made in the 1960s with several different physical robots and ELIZA [8]. With the explosion of innovation came inflated public expectations and increased public and private funding. However, progress slowed, and funding correspondingly dwindled leading to the "AI Winters" of the 1970s, 1980s, and early 1990s [9]. Despite the lack of public interest, slow and steady development was ongoing. Substantial technical progress was made through modeling human neurocognition and language as well as developments in data representation and probability theory that paved the way for today's AI technology.

Public interest was reignited in the late 1990s with inter-human–computer competitions such as IBM's Deep Blue v. chess grandmaster Gary Kasparov [9]. These early models of AI were too simplistic to model human learning and adapting, which led to the development of artificial neural networks. Artificial neural networks underlie deep learning (DL) and function, at the simplest level, in a similar way to our own neurons—multiple inputs gather into a single output decision or behavior [9]. Today, DL is embedded in our everyday lives including social media using DL algorithms to feed consumers targeted ads, automatic photo tagging, and DL chatbots such as ChatGPT [10]. Figure 1 shows an overview of the history of AI.

Current Uses and Public Response to Generative AI

Artificial intelligence has long been a topic of public debate, but since the release of ChatGPT, there has been a flood of news stories and even lawmaker involvement in the

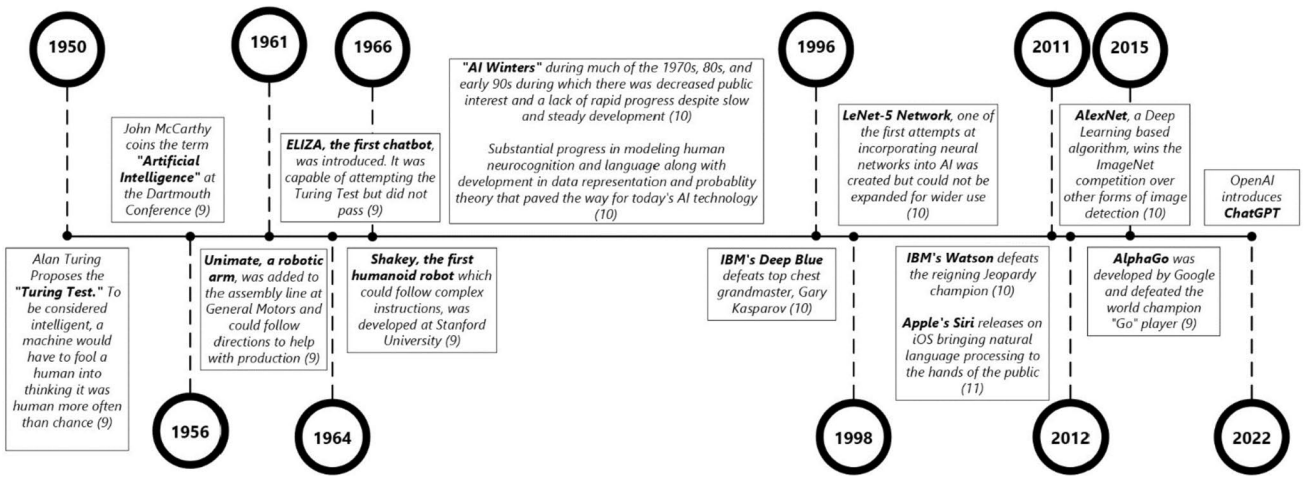

Fig. 1 History of artificial intelligence

regulation and use of AI. A recent Pew Research Center poll found that Americans are only partially aware of the involvement of AI in their day-to-day lives. When asked about six specific areas in which AI are involved (fitness trackers, customer service chatbots, product recommendations, security camera alerts, music playlist recommendations, and spam email filtering), only 30% could accurately label all six as uses of AI [11]. Additionally, the same survey found that about 84% of Americans are at least as equally concerned as they are excited about the possible uses of AI with only 14% more excited than concerned about its use [11]. Currently, generative AI technology is being used in three main categories: language/text, visual, and auditory as shown in Table 1. Google Gemini, an upcoming generative AI technology, takes the potential even further with the integration of several different modalities. Programs such as ChatGPT, Dall-E 2 (art), and Jukebox (music) are available for use by the public. The potential uses of this technology in improving workflow, assisting with creative thinking, and accessing knowledge are as wide as the individual's imagination.

Over the past 2 years, the European Union (EU) has been attempting to pass regulations on AI. Since the release of ChatGPT and other technologies, the EU has been working to get these regulations finalized governing the rules around facial recognition, biometric surveillance, and other uses [12]. The EU's AI Act was passed by European parliament on May 11, 2023, which stratifies AI technology into levels of risk and includes such unacceptable risks as manipulations, exploitation, predicting criminal behavior, and more [13]. In the USA, lawmakers also have their eyes on AI. On May 16, 2023, Sam Altman, the former CEO of OpenAI, was called before the Senate to testify about the future of AI [14]. He discussed his belief that a regulatory agency should be created, that certain jobs will likely become obsolete, and other potential issues that could arise with this technology [14]. The push for governmental regulation of AI has been spurred by reports citing the use of generative AI to create malware, scam emails, ransom/kidnapping calls, bots on social media with a realistic presence and manipulative agenda, and tragically, a report of a Belgian man who committed suicide to reduce climate change after a chatbot encouraged him to do so [15, 16]. The public is understandably wary of the potential power and possible tragedies from the flurry of AI technologies.

Table 1 Current uses of generative AI

	Description	Examples
Language/text	Language-based functions including writing, coding, generating human-like responses	ChatGPT, Google Bard, Bing Chat, Github Copilot, Anthropic's Claude
Visual	Creating images, pictures, videos, and assisting with graphic design	Generative Adversarial Networks, Dall-E 2, Bing Image Creator, Make-A-Video by Meta
Auditory	Creation of music, speech, and other sounds that can be utilized in video clips and music albums	AudioLM, Jukebox
Multimodal	Combines different modalities together for integrative generative potential and/or integrates directly with pre-existing software	Google Gemini (currently not released), Microsoft 365 Copilot

Technology and Terminology for Generative AI

Technological Foundations of ChatGPT

Artificial intelligence has three classification tiers: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence (ASI). ANI operates with a narrow focus to complete specific tasks such as with Apple's Siri or Google Translate [17]. Most forms of AI we have seen until now have been ANI, but GPT-4 and other DL interfaces have shown signs of AGI which is marked by human-like reasoning, planning, and learning [17, 18]. Microsoft recently published a paper claiming that GPT-4 shows "sparks" of AGI and represents a step towards AGI given its numerous abilities spanning multiple domains and its "performance on a wide spectrum of tasks at or beyond human-level," though this paper has been criticized due to its subjective assessment and potential bias as Microsoft may have commercial reasons to overestimate its functionality [19]. Finally, ASI encompasses theoretical technology that could operate at a level beyond human intelligence and cognition [18]. The three categories ANI, AGI, and ASI are summarized in Table 2.

To build an understanding of how GPT-4 works, let's start with its name: Generative Pretrained Transformer-4. Generative comes from the term "Generative AI" which is a type of artificial intelligence model that can *generate* new content, while pre-trained refers to AI being a pre-trained language model (PLM). Generative AI models attempt to generate new content based on what they have learned. GPT-4 has learned from reading a great swath of the internet through "neural nets" [20•]. Neural nets operate similarly to neurons in the human brain—taking multiple inputs, evaluating, and then passing that information to downstream nodes for further processing. Training neural net models involves billions of parameters or inputs which are then adjusted until a desired output is achieved for all inputs. Pre-training of the models can involve various data sets depending on the goals of the model, but most utilize data from books, webpages, Wikipedia, and even conversational text to give the model a general knowledge base with a conversational feel [21•]. Pre-training on specialized data such as multilingual text, scientific/medical text, or code can provide models with focused expertise for downstream tasks [21•].

The term pre-trained language model (PLM) represents a neural net that has been calibrated to output text. In the case of GPT-4, the inputs are concepts it has gleaned from what it has read, with the outputs being the phrases and ideas statistically most associated with input concepts.

The final word in the term GPT is "Transformer." This is a special type of neural net architecture that allows the algorithm to pay closer "attention" to some parts of the text sequence than others. This is the architecture that allows the model to be so successful and where "thinking" seems to occur [20•]. Described simply, the transformer architecture first transforms and passes incoming text through the neural net to generate probabilities for which words may come next. Probabilities for subsequent words are sequentially calculated and reorganized with certain weights which are used to list the probabilities for which text comes next and output responses [20•].

Table 2 AI-related terminology

	Description	Examples	Additional notes
Artificial narrow intelligence (ANI)	Algorithm that is designed to focus and complete a single, narrow task with limited ability of knowledge gained from doing the task to apply to other tasks	Siri, Alexa, language translation, image recognition, ChatGPT	Includes all forms of AI up until this point
Artificial general intelligence (AGI)	Algorithm can reason, plan, learn from experiences, understand human motives for actions, and have its own motivations and goals	Microsoft asserts that GPT-4 shows signs of AGI	Incorporates Deep Learning and Neural Networks to allow for analyzing multiple inputs
Artificial super intelligence (ASI)	Algorithm that surpasses human intelligence and cognitive skills	None, currently theoretical	
Pre-trained language model (PLM)	Neural network that has been calibrated to output text	GPT-2	
Large language model (LLM)	PLMs containing at least 10 billion different parameters	ChatGPT (GPT-4), Google Bard	Large number (over 10 billion) of parameters allows for performance to significantly improve and produces human-like text

Large Language Models and their Limitations

Interestingly, as the size and amount of text used to train PLMs increases above a threshold, performance significantly improves and "emergent abilities" that are not present in small-scale language models occur [21•]. Such as being able to accurately answer questions on subjects it was not trained on from patterns within the neural networks. To differentiate between models of different parameter sizes, the term large language model (LLM) was coined for PLMs containing at least 10 billion parameters [21•]. Table 2 summarizes LLMs and PLMs.

Many of these large language models have been developed in industry because of the intensive computational resources necessary in creating and training these algorithms. For example, LLaMA, a LLM by Meta AI that is pre-trained on 65 billion parameters, used 2048 A100-80G GPUs during its pre-training [21•]. Similarly, other LLMs, including many with more parameters than LLaMA, have similarly intensive pre-training and development demands. As a result of major LLMs being developed by industry, research has been limited since algorithms are not openly shared. This contributes to underlying principles of LLMs not being well explored or understood. As LLMs learn from extensive collections of human text, it can be challenging to ensure that responses are "helpful, harmless, and honest." Toxic, fictitious, or biased content can arise due to the nature of the training data [21•]. LLMs have also been known to "hallucinate" where the algorithm confabulates producing errors or risky responses in certain contexts [22]. Open AI applied an intervention called "red teaming" where GPT-4 was prompted to output biases, hateful propaganda, and other harmful responses to better understand its risks, reduce confabulations, and mitigate risk, though the risk nevertheless remains [21•, 23, 24]. Open-source algorithms without such safeguards are also available, making it easier for those interested in spreading false narratives or creating malicious software to generate content and code.

An additional limitation of LLMs is that they are confined to their pre-training data and appear to struggle with computational tasks such as basic arithmetic. While these limitations can be overcome by the LLM using external plugins, these also introduce yet another source of bias into an already unknown and potentially biased tool.

Privacy and Safety Considerations for Generative AI in Psychiatry

A Pew Research Center survey from February 2023 shows that 37% of Americans think using AI in healthcare would worsen medical record security compared with 22% who think it would improve security [25]. The American Psychiatric Association's App Evaluation Model offers relevant questions to consider such as what security measures are in place to protect user data; how is personal health information handled; what does the privacy policy say, and are users able to delete their personal data? [26•]. Additionally, the idea of "contextual integrity" defined by Helen Nissenbaum as a rubric to judge data handling practices is also helpful when thinking about privacy. With the ever-increasing use of online services and passive data collection from smartphones and computers, contextual integrity describes privacy as the appropriate flow of information in conformance with reasonable expectations, contextual social norms, and the actors involved. For instance, the Health Insurance Portability and Accountability Act (HIPAA) defines how health information flows in providing health care services and requires patients to be aware of that information flow [27•]. The OpenAI privacy policy states that all conversation history is collected as well as personal identifying information provided by users to create their account, data collected from users' browsers, cookies, analytics, and their use of the website. Data collected may be used to train the algorithm, which users may opt out of if they choose. Personal identifiers, commercial information, network activity information, geolocation data, and account login credentials may be disclosed to "affiliates, vendors and service providers, law enforcement and parties involved in Transactions" [28]. Given this privacy policy, it is not advised to use ChatGPT to assist with patient documentation or the handling of patient information until safeguards for health care data are put in place. Some generative AI companies are building systems behind firewalls and incorporating HIPAA protections, which would make this technology more accessible for medicine.

Clinical Foundation, Validity, and Treatment Goals to Consider with Generative AI in Psychiatry

The clinical foundation of generative AI can be assessed by first considering the data upon which it is generating its answers: what sources and methods were used to pre-train and train the model? Is there any evidence of efficacy in patient care? What biases may exist due to the underlying data source, algorithms, or training methods? Most companies do not release training methods or data sources, but if this technology is to be used in healthcare, this information is vital not only to assess efficacy but also to understand the risk of bias and even discrimination. It is equivalent to a pharmaceutical company including a package insert for a drug.

The large language models underlying GPT-4 are pre-trained with text. Sixteen percent of the text comes from

books and news articles and 84% of the text comes from webpages. The webpage data includes text such as Wikipedia but also low-quality text like spam mail and Reddit [21•]. Clearly, none of these is authoritative sources on mental health or psychiatry, but knowing the source of information provides insight into not only the accuracy of the information but also what biases may exist. Bias can originate from data sets and become magnified through the machine learning development pipeline, leading to bias-related harms [29•]. For example, false news and racist viewpoints are often shared on Reddit. Another way to think of ChatGPT is as a user interface that packages internet information up to a defined timepoint. While these sources can be accurate at times, there are more reliable and high-quality resources.

Ethical and Legal Considerations for Generative AI in Psychiatry

Ethical and legal considerations are clearly lagging given the rapid pace of AI development. Currently, there are more questions than answers. In terms of considerations for psychiatry, it is important for us to think about ethical standards for how we use generative AI, how we disclose our use of it, and how we validate its use in clinical environments. More research needs to be done in studying which patients may benefit from its use and which patients may experience harm. Are there risks of the technology causing direct harm to certain populations, such as making psychosis worse? As with any technology, we also need to think of ways to foster digital inclusion and provide updated literacy for patients explaining the pros and cons of generative AI [30].

An additional consideration for generative AI models is an understanding of what algorithms influence their output. With social media for instance, a person's newsfeed is determined by that person's interests and can include topics harmful to their mental health or false news. This is because algorithms behind social media sites are geared towards producing content that is engaging to consumers [30]. As generative AI is developed and used for healthcare applications, it will be important to know what influences are at play in packaging its healthcare recommendations. There is also room for influencing decision-making when an algorithm decides which information to present or exclude. New research studying human–computer interactions with generative AI is also needed to better understand how this technology will change our social landscape and trust in technology. We are already seeing people on online forums detailing how to prompt ChatGPT to provide "Cognitive Behavioral Therapy" and saying that this type of "therapy" has helped them immensely. Employing generative AI for therapy is something that needs to be studied more thoroughly. It will be important to know what information and potential misinformation patients are receiving when they ask generative AI systems for mental health assistance. Of note, current well-being apps such as Wysa and Woebot do not currently use generative AI for their therapy responses. These apps use a number of question-and-answer combinations that were vetted by a human [31].

Finally, it is difficult to say who is accountable for what and to whom under which circumstances. In computerized systems, "bugs" or accuracy ranges can divert accountability and chalk up an error to just falling within the margin of error [32•]. The problem with this is that sometimes the margin of error is a result of human-made decisions along the machine learning pipeline that could be better optimized, blurring accountability [32•]. A lack of transparency in chosen models, data sources, and training methods further obscures the picture [30].

Conclusion and Initial Guidance to the Field

Today, generative AI technology is not yet ready for use in the field of mental health care.

1. If generative AI is to be used to augment psychiatric care, steps must be taken to ensure privacy with regulation compliant and acceptable protections. For example, patient data must not be shared with third parties, such as entering it into an LLM, without proper patient consent. These implementations will require using generative AI builds that are built for healthcare rather than generalized generative AI software.
2. *Ethical and legal standards need to be developed for generative AI use, how to disclose and inform about this use and its validation in clinical environments* as described in the previous section.
3. *Ensuring a reliable clinical foundation for generative AI to derive its parameters is a key factor to address prior to implementation in psychiatry*. By basing clinical support from generative AI from medically acceptable knowledge sources that reflect local practice standards and the evidence base, employing evaluative processes already developed, e.g., PRISMA, there would be a high degree of clinical validity that could be attributed to generative AI. Additionally, with the ever adapting and developing nature of medicine, the parameters would have to be undergoing regular updates, as in most learning health systems, to ensure the most clinically up to date recommendations and information.
4. *AI's function and use must be understandable to providers, organizations, and systems before deployment. F*ew studies have been done thus far using AI in psychiatric care. Clinical decision support and documentation are likely early applications. Already, the FDA has rules for

the use of AI for clinical decision support which requires a high degree of transparency that is not currently available with LLMs. The key concept is explainability: a clinician should know exactly why AI has generated a particular response to a clinical question. With regard to AI-based interventions in development, prior to FDA approval for a medical device, several clinical trials must be conducted to demonstrate the benefits and risks against current interventions These trials must also consider the safety profile and side effects that a drug may have. Likewise, prior to widespread use of AI in mental health care, we must consider both the benefits and harm that may occur with use of AI to augment care.

While this technology has amazing potential in mental health care, current popular large language models are not built with adequate sources, privacy protections, or robust enough models to prevent unforeseen harm. Regulation may help increase transparency, accuracy, accountability, and privacy. Striking a balance between innovation and regulation is needed if we want to enjoy this technology and limit the harm.

Acknowledgements The editors would like to thank Dr. Steven Richard Chan for taking the time to review this manuscript.

Data Availability The author confirms that we do not analyse or generate any datasets in this work.

Declarations

Conflict of Interest Darlene R. King, Guransh Nanda, Allison Dempsey, Sarah Hergert, and Jay H. Shore each declare no potential conflicts of interest. Joel Stoddard has received grants from the NIH, the Brain and Behavior Research Foundation, the Colorado Office of Economic Development, and the Children's Hospital Colorado Foundation. Dr. Stoddard also has family equity in AbbVie, Merck, CVS, Bristol Meyers Squibb, Johnson & Johnson, Abbott Labs, and Pfizer. In addition, Dr. Stoddard has a patent 63/489,517 pending. John Torous is a scientific board member of Precision Mental Wellness.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance

1. 91 Important chatGPT statistics & user numbers in April 2023 (GPT-4, Plugins Update) - Nerdy Nav [Internet]. 2022 [cited 2023 Apr 23]. Available from: https://nerdynav.com/chatgpt-statistics/.
2. Statement on AI risk | CAIS. [cited 2023 Jul 2]. Available from: https://www.safe.ai/statement-on-ai-risk.
3. Anderson M. 'AI Pause' open letter stokes fear and controversy - IEEE spectrum [cited 2023 Apr 25]. Available from: https://spectrum.ieee.org/ai-pause-letter-stokes-fear.
4. AI Principles [Internet]. Future of Life Institute. [cited 2023 Apr 23]. Available from: https://futureoflife.org/open-letter/ai-principles/.
5. AAMC [Internet]. [cited 2023 May 14]. A growing psychiatrist shortage and an enormous demand for mental health services. Available from: https://www.aamc.org/news/growing-psychiatrist-shortage-enormous-demand-mental-health-services.
6. AMIA - American Medical Informatics Association. [cited 2023 May 14]. AMIA 25x5. Available from: https://amia.org/about-amia/amia-25x5.
7. Reader TMP. The Czech play that gave us the word 'robot'. The MIT Press Reader. 2019 [cited 2023 Jun 13]. Available from: https://thereader.mitpress.mit.edu/origin-word-robot-rur/.
8. Sarangi S, Sharma P. Artificial intelligence: evolution, ethics and public policy. London: Routledge India; 2018;164.
9. Muthukrishnan N, Maleki F, Ovens K, Reinhold C, Forghani B, Forghani R. Brief history of artificial intelligence. Neuroimaging Clin N Am. 2020;30(4):393–9.
10. How AI ruled our lives over the past decade | CNN Business. [cited 2023 May 5]. Available from: https://www.cnn.com/2019/12/21/tech/artificial-intelligence-decade/index.html.
11. Nadeem R. Public awareness of artificial intelligence in everyday activities. Pew Research Center Science & Society. 2023 [cited 2023 May 19]. Available from: https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/.
12. Chee FY, Coulter M, Mukherjee S. EU lawmakers' committees agree tougher draft AI rules. Reuters. 2023 May 11 [cited 2023 May 19]; Available from: https://www.reuters.com/technology/eu-lawmakers-committees-agree-tougher-draft-ai-rules-2023-05-11/.
13. Browne R. CNBC. 2023 [cited 2023 May 19]. Europe takes aim at ChatGPT with what might soon be the West's first A.I. law. Here's what it means. Available from: https://www.cnbc.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html.
14. Sam Altman: CEO of OpenAI calls for US to regulate artificial intelligence. BBC News [Internet]. 2023 May 16 [cited 2023 May 19]; Available from: https://www.bbc.com/news/world-us-canada-65616866.
15. Espinosa N. Forbes. [cited 2023 May 20]. Council post: the unforeseen consequences of chatGPT. Available from: https://www.forbes.com/sites/forbestechcouncil/2023/03/10/the-unforeseen-consequences-of-chatgpt/
16. Times TB. Belgian man dies by suicide following exchanges with chatbot. [cited 2023 May 20]. Available from: https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt
17. Priyadarshini R, Mehra RM, Sehgal A, Singh PJ, editors. Artificial intelligence: applications and innovations. New York: Chapman and Hall/CRC; 2022. 300 p.
18. GPT5: release date, AGI meaning and expected features - dataconomy. 2023 [cited 2023 Apr 25]. Available from: https://dataconomy.com/2023/04/03/chat-gpt5-release-date-agi-meaning-features/.
19. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv; 2023 [cited 2023 Apr 30]. Available from: http://arxiv.org/abs/2303.12712.
20.• What is chatGPT doing … and why does it work?. 2023 [cited 2023 Apr 25]. Available from: https://writings.stephenwolfram.

com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/. **This article summarizes how LLMs such as ChatGPT function in the context of language processing, machine learning, and neural nets. It intuitively explains how LLMs are trained and function, with a focus on ChatGPT.**

21.• Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv; 2023 [cited 2023 Apr 30]. Available from: http://arxiv.org/abs/2303.18223. **This article is a technical view of the recent developments made with LLMs and focuses on pre-training, adaptation tuning, utilization, and capacity evaluation of these models. It is a thorough resource about the current state of LLMs.**

22. O'Reilly RC, Munakata Y, Frank MJ, Hazy TE. Computational cognitive neuroscience. Vol. 1124. PediaPress Mainz; 2012.

23. NewsGuard's Misinformation Monitor: GPT-4 produces more misinformation than predecessor [Internet]. NewsGuard. [cited 2023 Apr 30]. Available from: https://www.newsguardtech.com/misinformation-monitor/march-2023.

24. Ovadya A. Red teaming improved GPT-4. violet teaming goes even further. Wired. [cited 2023 May 11]; Available from: https://www.wired.com/story/red-teaming-gpt-4-was-valuable-violet-teaming-will-make-it-better/.

25. Nadeem R. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center Science & Society. 2023 [cited 2023 May 14]. Available from: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/.

26.• Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med. 2020;3(1):1–8. **This paper is an excellent review of the APA App Evaluation Model. It details the process by which the model was created discussing the available models at the time, the deficits that existed, and how the APA model was synthesized to address those deficits and provide an objective metric for app evaluation.**

27.• Nissenbaum H. Privacy as contextual integrity. Wash Law Rev. 2004;79. **This paper details the importance of privacy and elaborates on privacy policy in the US."Contextual Integrity" is explained as a concept that ties privacy to the norms of a particular context. HIPAA is discussed as a privacy law that specially recognizes health and medical information.**

28. Privacy policy. [cited 2023 May 14]. Available from: https://openai.com/policies/privacy-policy.

29.• Connolly SL. Leveraging implementation science to understand factors influencing sustained use of mental health apps: a narrative review'. J Technol Behav Sci. 2020;1–13. **This review discusses the implementation of mental health apps including the qualities that lead to the success and use of such apps. The article highlights various factors such as simplicity, benefits over current tools, ease of navigation, and alignment with user's needs and skill sets that facilitate mental health app implementation.**

30. King D. ChatGPT not yet ready for clinical practice. Psychiatr News [Internet]. 2023 Jun 16 [cited 2023 Jul 24]; Available from: https://psychnews.psychiatryonline.org/doi/10.1176/appi.pn.2023.07.7.56.

31. Reardon S. Scientific American. [cited 2023 Jul 24]. AI chatbots could help provide therapy, but caution is needed. Available from: https://www.scientificamerican.com/article/ai-chatbots-could-help-provide-therapy-but-caution-is-needed/.

32.• Cooper AF, Moss E, Laufer B, Nissenbaum H. Accountability in an Algorithmic Society: relationality, responsibility, and robustness in machine learning. In: 2022 ACM conference on fairness, accountability, and transparency. 2022 [cited 2023 May 5]. p. 864–76. Available from: http://arxiv.org/abs/2202.05338. **This paper discusses the important topic of accountability in relation to ML and AI. With data driven algorithmic processes such as these being used or being introduced to a variety of fields including medicine, barriers to accountability with their use become apparent. Overall, the paper introduces how the lack of accountability for errors or mistakes of AI poses a challenge in its implementation in healthcare.**