How Far Are We From AGI

ICLR 2024 Workshop Proposal

Jiaxuan You, Ge Liu, Yunzhu Li, Song Han, Dawn Song

Workshop Summary

Historical Context and Importance of AGI. Building artificial general intelligence (AGI) has been a defining challenge in the realm of AI research. Since the dawn of the computational era, the aspiration to achieve AGI has profoundly shaped the vision and mission of AI researchers [21]. Broadly defined, AGI refers to machines that can perform any intellectual task that a human being can do [11]. Over the decades, we have witnessed substantial progress in narrow AI, where systems excel in specific tasks. Still, the quintessential goal of creating an AGI, a system with broad and flexible cognitive abilities, remains elusive.

Recent Advances in AI: Towards AGI?. The recent advances in large language models (LLMs) like GPT-4 and LLama-2, present an intriguing twist in this narrative. These models have displayed remarkable capabilities, sometimes on par or even surpassing human abilities in specific domains, hinting at a form of the Turing test being passed [4, 6]. Furthermore, the scope of these models is no longer limited to just natural language processing (NLP). We're seeing promising signs of these models branching out, such as AI agents utilizing tools [23, 20], retrieval from external database [3], displaying reasoning abilities [28], competing complex tasks via writing code [14], multimodal learning with text and images [9], etc.

The Road Ahead: Challenges and Limitations. However, as promising as these advances are, the gap between current LLMs and true AGI is still significant. Notable limitations include diminishing returns and potential constraints to the scaling law [12], lacking robust reasoning capabilities [15], hallucination and factual inaccuracies [1], lack of commonsense reasoning capabilities [26], and many others. Furthermore, as we move closer to AGI, it becomes paramount to address the critical concerns of safety, ethics, and regulatory implications, for example, aligning AGI's values with humanity's diverse set of beliefs [10], navigating the moral dilemmas and ensuring AGI can make ethical decisions [27], and addressing security risks of generative AI [2].

Workshop Scope and Topics. This workshop aims to become a melting pot for ideas, discussions, and debates regarding our proximity to AGI. We invite submissions on a range of topics including, but not limited to:

- *Frontiers of AGI research*: examples include AI agents [19], embodied AI [8], retrieval-based [3] and tool-augmented LLMs, knowledge-enhanced AI, and multi-agent AI.
- *Classic AGI Attempts as Inspiration*: Delving into historical methods such as expert systems, symbolic AI [16], Type I and Type II reasoning [17] for insights that can guide LLM research further.
- *Interdisciplinary Insights for AGI*: Drawing parallels from fields like psychology [24], sociology [7], and neuroscience [18] to inspire and inform the development of LLMs towards AGI.
- *Fundamental Limitations of LLMs*: Analyzing the intrinsic capabilities or lack thereof in LLMs that might impede their progression to AGI. This includes discussions on reasoning, planning, and more [13].
- *Practical Limitations of LLMs and Foundation models*: Addressing external challenges like system constraints, computational costs [25], data acquisition barriers, and privacy concerns [29].
- *Safety, Ethics, and Regulation in AGI Development*: Exploring the complexity of moral, safety, and regulatory concerns that will shape AGI's evolution [22].
- *AGI's Economic and Societal Impacts*: Probing the potential changes AGI might initiate into our societies, economies, and daily lives [5].

This workshop hopes to provide a platform for AI researchers worldwide to collaboratively navigate the labyrinthine journey towards AGI. We await your insightful contributions and eager discussions.

Workshop Modality. We plan to host the workshop in hybrid mode in order to best benefit the AI and machine learning communities. We will follow the physical limitation of available rooms for workshops when arranging the in-person session. We plan to virtually live broadcast the event following the preferences and suggestions from the ICLR 2024 main conference and other workshops. The organizers have past experience with organizing using Zoom or YouTube

Live. We will create a dedicated website for the workshop. For example, an organizer, Jiaxuan You, has created a workshop website, `https://glfrontiers.github.io`, for the NeurIPS 2022&2023 GLFrontiers workshop which has been quite effective. Lastly, we are also flexible with the other modes of the workshop.

Submissions. We welcome both short research papers of up to 6 pages and full-length research papers of up to 9 pages (excluding references and supplementary materials). All accepted papers will be presented as posters. We will select around 10 papers for short oral presentations and 2 papers for outstanding paper awards with potential cash incentives.

Tentative Schedule and Organizational Details

Overview. We aim to showcase the latest in AGI research, delve into breakthroughs from LLMs and foundation models, and assess the current gaps and challenges towards AGI. With insights from renowned AI experts, the workshop seeks to spark new ideas, foster discussions, and compile a comprehensive summary to guide the broader AI and ML community. We have confirmed with many renowned researchers with interest in giving the keynote talks.

Tentative schedule. We propose to organize the workshop schedule with invited talks, poster sessions, a panel session, as well as short contributed talks on oral and outstanding paper submissions. The actual workshop schedule will be aligned with the NeurIPS 2023 official schedule. **All of the speakers have been contacted and 4 speakers have confirmed their interest in giving keynote talks and/or joining the panel discussion in the workshop.**

8:00-8:50AM	Poster setup
8:45-9:00AM	**Opening remarks**
9:00-10:00AM	**Keynote talk**: ***Yoshua Bengio***, MILA (confirmed)
10:00-11:00AM	**Keynote talk**: Yejin Choi, University of Washington (confirmed)
11:00-11:30AM	**Contributed talks from submissions**
11:30-12:00PM	Discussions and Coffee break
12:00AM-1:30PM	**Poster session 1 and Lunch break**
1:30-2:15PM	**Keynote talk**: Joshua Tenenbaum, MIT (confirmed)
2:15-3:00PM	**Keynote talk**: Andrej Karpathy, OpenAI (in contact)
3:00-3:30PM	**Contributed talks**
3:30-4:00PM	Discussions and Coffee break
4:00-5:00PM	**Poster session 2**
5:00-6:00PM	**Panel discussions: How far are we from AGI?** Michael Bernstein(Stanford, confirmed), Yoshua Bengio(tentative), Joshua Tenenbaum(tentative), Yejin Choi(tentative)

Expected Size. Given the wide spectrum of topics this workshop is covering and the growing public interest in the theme, we expect the AGI workshop to have a large size of audience (estimated 400-500 attendees). We expect that there will be broad interest from both academia and industry attendees with diverse backgrounds.

Details on the Discussion Session. We plan to encourage interactive discussions in the workshop. During the morning and afternoon breaks, we will organize a few breakout discussion groups each focusing on one topic regarding the path to AGI, based on the submissions and invited talks. All the attendees are welcome to join the discussions that they are interested in. We hope this design could give the attendees an opportunity to engage in casual networking, communicate opinions/ideas.

Details on the Panel Session. When authors are creating their submissions, we will ask them to optionally leave questions to the panel. We will create official accounts on social network platforms such as Twitter to publicize the workshop while collecting the questions and feedback. During the panel session, we will initiate the discussions from the highlighted questions we have collected, and transition to the questions from the live audience.

Differences with Previous Workshops. While there have been past workshops on related topics, none have specifically addressed our focus for today. For instance, there was a workshop on AI for Agent-Based Modelling (AI4ABM) at ICLR 2023, which concentrated on agent-based modeling techniques. However, our workshop aims to address a broader and more pressing question: based on recent advancements, how close are we to achieving AGI.

Details on the Invited Speakers

Yoshua Bengio (confirmed). Recognized worldwide as one of the leading experts in artificial intelligence, Yoshua Bengio is most known for his pioneering work in deep learning, earning him the 2018 A.M. Turing Award, "the Nobel Prize of Computing," with Geoffrey Hinton and Yann LeCun. He is Full Professor at Université de Montréal, and the Founder and Scientific Director of Mila – Quebec AI Institute. He co-directs the CIFAR Learning in Machines & Brains program as Senior Fellow and acts as Scientific Director of IVADO. In 2019, he was awarded the prestigious

Killam Prize and in 2022, became the most cited computer scientist in the world. He is a Fellow of both the Royal Society of London and Canada, Knight of the Legion of Honor of France, Officer of the Order of Canada, Member of the UN's Scientific Advisory Board for Independent Advice on Breakthroughs in Science and Technology since 2023 and a Canada CIFAR AI Chair. Concerned about the social impact of AI, he actively contributed to the Montreal Declaration for the Responsible Development of Artificial Intelligence.

Yejin Choi (confirmed). Yejin Choi is Wissner-Slivka Professor at the Paul G. Allen School of Computer Science & Engineering at the University of Washington and also a senior research director at AI2 overseeing the project Mosaic. Her research investigates a wide variety problems across NLP and AI including commonsense knowledge and reasoning, neural language (de-)generation, language grounding with vision and experience, and AI for social good. She is a MacArthur Fellow and a co-recipient of the NAACL Best Paper Award in 2022, the ICML Outstanding Paper Award in 2022, the ACL Test of Time award in 2021, the CVPR Longuet-Higgins Prize (test of time award) in 2021, the NeurIPS Outstanding Paper Award in 2021, the AAAI Outstanding Paper Award in 2020, the Borg Early Career Award (BECA) in 2018, the inaugural Alexa Prize Challenge in 2017, IEEE AI's 10 to Watch in 2016, and the ICCV Marr Prize (best paper award) in 2013.

Joshua Tenenbaum (confirmed). Josh Tenenbaum is Professor of Computational Cognitive Science in the Department of Brain and Cognitive Sciences at MIT, a principal investigator at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), and a thrust leader in the Center for Brains, Minds and Machines (CBMM). He is also a co-scientific director with the MIT Quest for Intelligence. His research centers on perception, learning, and common-sense reasoning in humans and machines, with the twin goals of better understanding human intelligence in computational terms and building more human-like intelligence in machines. The machine learning and artificial intelligence algorithms developed by his group are currently used by hundreds of other science and engineering groups around the world. He is a MacArthur Fellow and has received the National Academy of Sciences' Troland Research Award.

Michael Bernstein(confirmed). Michael Bernstein is an Associate Professor of Computer Science at Stanford University, where he is a Bass University Fellow and STMicroelectronics Faculty Scholar. His research in human-computer interaction focuses on the design of social computing systems. This research has won best paper awards at top conferences in human-computer interaction, including CHI, CSCW, ICWSM, and UIST, and has been reported in venues such as The New York Times, Science, Wired, and The Guardian. Michael has been recognized with an Alfred P. Sloan Fellowship, UIST Lasting Impact Award, and the Patrick J. McGovern Tech for Humanity Prize. He holds a bachelor's degree in Symbolic Systems from Stanford University, as well as a master's degree and a Ph.D. in Computer Science from MIT.

Andrej Karpathy (in contact). Andrej Karpathy is a researcher and a founding team member at OpenAI. From 2017 to 2022, he was the Director of AI at Tesla, leading the computer vision team for Tesla Autopilot. He received his PhD in Computer Science from Stanford, where he developed early multimodal deep neural networks that processed both images and text, for example, for image captioning applications.He was also the primary instructor for CS231n, the first deep learning class at Stanford, which he designed, and which went on to become one of the largest classes in the AI curriculum.

Diversity and Inclusion Statement

Diversity of Topics. Our workshop is dedicated to embracing a wide array of perspectives on AGI. We've selected topics that span from historical insights to future challenges, ensuring we cover the breadth of the field. Our call for papers, from the frontiers of AGI research to the economic and societal impacts, and from historical inspirations to forward-looking ethical considerations, reflects our commitment to this diversity. We welcome contributors from diverse academic and cultural backgrounds, believing that multiple viewpoints lead to richer discussions.

Diversity of Organization Committee. The organizing committee consists of researchers with a wide variety of demographic backgrounds and experiences, and we aim to promote diversity along several axes, including gender, seniority, experience, affiliation, and research areas. Among the 5 organizers, 2 of them identify themselves as female and 3 of them as male. There are 3 assistant professors, 1 associate professor, and 1 full professor. Organizers are affiliated with different institutions, including UIUC, MIT, and UC Berkeley. **All the organizers have recently served as organizers and/or invited speakers in AI and/or deep learning workshops.**

Diversity of Speakers. Our commitment to fostering diversity is evident in our choice of invited speakers. They come from varied institutions, such as Université de Montréal, University of Washington, MIT, and OpenAI. Their titles range from full professors to industry co-founders, showcasing a mix of academia and industry perspectives. Notably, our speaker lineup includes the distinguished AI researcher, Professor Yoshua Bengio. We're proud that our current list features speakers who have prominently contributed to the fields of AI ethics and safety. Moreover, we are conscious of

gender representation, with both female and male voices present. As we progress, we aim to further expand this roster, continually striving for a holistic representation of the global AI community.

Diversity of Participants. Our workshop is devoted to providing a welcoming atmosphere for participants with diverse backgrounds. With a fully diverse organization committee of invited speakers, we hope the attendees of our workshop can find common perspectives and ignite fruitful discussions.

Accessibility. We plan to hold the workshop in hybrid mode; we believe that face-to-face discussions are the best format for igniting research ideas. We plan to live-stream the full workshop to allow access from a broader community. We will ensure sufficient communication before, during, and after the workshop to disseminate important research ideas and attract a broader audience. We will publish the titles of invited talks and speakers, as well as the PDF versions of contributed works on the workshop website in advance. We will popularize the important findings, results and works through website updates, video publications, and online social networks such as Twitter.

Organizers

Jiaxuan You (jiaxuan@illinois.edu). Jiaxuan You is an Adjunct Assistant Professor at the University of Illinois Urbana-Champaign Computer Science Department (officially as an assistant professor in 2024), and a senior research scientist at NVIDIA. He obtained his CS PhD from Stanford University. His research focuses on empowering deep learning with graph-structured data. He has published more than 15 publications in NeurIPS, ICML, ICLR, etc, with more than 8000 citations. Jiaxuan is the creator of GraphGym and a main contributor to PyG, which are popular open-source libraries for graph machine learning. He has served as a program committee member of NeurIPS, ICML, ICLR, AAAI, KDD, WWW, IJCAI more than 20 times. Jiaxuan co-organized the Stanford Graph Learning Workshop in 2021, which attracted over 7000 virtual attendees. Jiaxuan has led the organization of NeurIPS 2022 and 2023 GLFrontiers workshops, which attracted around 300 physical attendees and was one of the most popular workshops in NeurIPS 2022.

Ge Liu (geliu@illinois.edu). Ge Liu is a senior applied scientist at AWS AI and adjunct professor at the University of Illinois Urbana-Champaign Computer Science Department (incoming assistant professor in 2024). She received her Ph.D. from MIT EECS department, advised by Professor David Gifford. Her research lies in the intersection of machine learning and biology. She develops reliable, interpretable, and efficient machine learning and computational techniques for solving high-stake problems in biomedicine and life sciences, with a special interest in therapeutics molecule design. The principled methods she developed can be widely used in high-impact applications such as antibody drug discovery, cancer immunotherapy, and vaccine development while also applicable to domains such as sequential decision making and recommender systems. She is the recipient of the David S. Y. Wong Fellowship at MIT, and her Ph.D. thesis won the MIT EECS George M. Sprowls Ph.D. Thesis Award in AI and Decision-Making.

Yunzhu Li (yunzhuli@illinois.edu). Yunzhu Li is an Assistant Professor of Computer Science at the University of Illinois Urbana-Champaign (UIUC). Before joining UIUC, he collaborated with Fei-Fei Li and Jiajun Wu during his Postdoc at Stanford. Yunzhu earned his PhD from MIT under the guidance of Antonio Torralba and Russ Tedrake. His work stands at the intersection of robotics, computer vision, and machine learning, with the goal of helping robots perceive and interact with the physical world as dexterously and effectively as humans do. Yunzhu received the Adobe Research Fellowship and was selected as the First Place Recipient of the Ernst A. Guillemin Master's Thesis Award in Artificial Intelligence and Decision Making at MIT. His research has been published in top journals and conferences, including Nature, NeurIPS, CVPR, and RSS, and featured by major media outlets, including CNN, BBC, The Wall Street Journal, Forbes, The Economist, and MIT Technology Review.

Song Han(songhan@mit.edu). Song Han is an associate professor at MIT EECS. He received his PhD degree from Stanford University. He proposed the "Deep Compression" technique including pruning and quantization that is widely used for efficient AI computing, and "Efficient Inference Engine" that first brought weight sparsity to modern AI chips, which influenced NVIDIA's Ampere GPU Architecture with Sparse Tensor Core. He pioneered the TinyML research that brings deep learning to IoT devices, enabling learning on the edge (appeared on MIT home page). His team's work on hardware-aware neural architecture search (once-for-all network) enables users to design, optimize, shrink and deploy AI models to resource-constrained hardware devices, receiving the first place in many low-power computer vision contests in flagship AI conferences. Song received best paper awards at ICLR and FPGA, faculty awards from Amazon, Facebook, NVIDIA, Samsung and SONY. Song was named "35 Innovators Under 35" by MIT Technology Review for his contribution on "deep compression" technique that "lets powerful artificial intelligence (AI) programs

run more efficiently on low-power mobile devices." Song received the NSF CAREER Award for "efficient algorithms and hardware for accelerated machine learning", IEEE "AIs 10 to Watch: The Future of AI" award, and Sloan Research Fellowship.

Dawn Song (dawnsong@cs.berkeley.edu). Dawn Song is a Professor in the Department of Electrical Engineering and Computer Science at UC Berkeley. Her research interest lies in deep learning and security. She has studied diverse security and privacy issues in computer systems and networks, including areas ranging from software security, networking security, database security, distributed systems security, applied cryptography, to the intersection of machine learning and security. She is the recipient of various awards including the MacArthur Fellowship, the Guggenheim Fellowship, the NSF CAREER Award, the Alfred P. Sloan Research Fellowship, the MIT Technology Review TR-35 Award, the George Tallman Ladd Research Award, the Okawa Foundation Research Award, the Li Ka Shing Foundation Women in Science Distinguished Lecture Series Award, the Faculty Research Award from IBM, Google and other major tech companies, and Best Paper Awards from top conferences. She obtained her Ph.D. degree from UC Berkeley. Prior to joining UC Berkeley as a faculty, she was an Assistant Professor at Carnegie Mellon University from 2002 to 2007.

Program Committee members

Based on our past experience in organizing workshops, we anticipate 80-100 paper submissions. We will ensure that each paper submission will receive 3 reviews, and each reviewer will review at most 3 reviews. We will reach out to a list of program committee (PC) members based on their expertise in NLP, CV, RL, Robotics, AI Ethics, ML systems, etc., and experience with previous workshops. We will confirm their availability to join the program committee again after the workshop submission site has been set up.

- Aaron Gokaslan, Cornell
- Aniruddha Kembhavi, AI2, UW
- Anthony Francis, Google
- Dan Fu, Stanford University, Together
- Julien Launay, Hugging Face, ENS
- Daniel Hesslow, Hugging Face
- Tri Dao, Princeton University, Together
- Azalia, Mirhoseini, Anthropic
- Beidi Chen, Meta, CMU
- Bruno Ribeiro, Purdue University
- Bowen Liu, Insitro
- Bowen Jing, MIT
- Bryan Perozzi, Google Research
- Christopher Morris, Polytechnique Montréal
- Dylan Bourgeois, EPFL
- Guangtao Wang, JD.com
- Hongyu Ren, Stanford University
- Ines Chami, Stanford University
- Kaidi Cao, Stanford University
- Kexin Huang, Stanford University
- Matthias Fey, TU Dortmund
- Thomas Kipf, Google Brain
- Ting Chen, Google Brain
- Chengshu Li, Stanford
- Chuang Gan, IBM, MIT
- Dhruv Batra, GaTech, Meta AI
- Fei Xia, Google
- Josh Tenenbaum, MIT
- Matt Deitke, AI2, UW
- Xiang Ren, University of Southern California
- Xiaowen Dong, University of Oxford
- Yujia Li, DeepMind
- Petros Maniatis, Google Brain
- Shuxin Zheng, Microsoft Research
- Tinglin Huang, Yale University
- William Hua, MILA
- Weikang Qiu, Yale University
- Yingheng Wang, Cornell University
- Shirley Wu, Stanford University
- Yukuo Cen, Tsinghua University
- Zhengdao Chen, Google
- Zijie Huang, UCLA
- Ziniu Hu, UCLA
- Vikas Verma, Aalto University

References

[1] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, et al. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*, 2023.

[2] Clark Barrett, Brad Boyd, Ellie Burzstein, Nicholas Carlini, Brad Chen, Jihye Choi, Amrita Roy Chowdhury, Mihai Christodorescu, Anupam Datta, Soheil Feizi, et al. Identifying and mitigating the security risks of generative ai. *arXiv preprint arXiv:2308.14840*, 2023.

[3] Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*, pages 2206–2240. PMLR, 2022.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[5] Erik Brynjolfsson and Andrew McAfee. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

[6] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[7] Manuel Castells. Toward a sociology of the network society. *Contemporary sociology*, 29(5):693–699, 2000.

[8] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(2):230–244, 2022.

[9] Nanyi Fei, Zhiwu Lu, Yizhao Gao, Guoxing Yang, Yuqi Huo, Jingyuan Wen, Haoyu Lu, Ruihua Song, Xin Gao, Tao Xiang, et al. Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, 13(1):3094, 2022.

[10] Iason Gabriel. Artificial intelligence, values, and alignment. *Minds and machines*, 30(3):411–437, 2020.

[11] Ben Goertzel. Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to ray kurzweil's the singularity is near, and mcdermott's critique of kurzweil. *Artificial Intelligence*, 171(18):1161–1173, 2007.

[12] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

[13] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253, 2017.

[14] Yujia Li, David Choi, Junyoung Chung, Nate Kushman, Julian Schrittwieser, Rémi Leblond, Tom Eccles, James Keeling, Felix Gimeno, Agustin Dal Lago, et al. Competition-level code generation with alphacode. *Science*, 378(6624):1092–1097, 2022.

[15] Gary Marcus. The next decade in ai: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*, 2020.

[16] Allen Newell and Herbert A Simon. Computer science as empirical inquiry: Symbols and search. In *ACM Turing award lectures*, page 1975. 2007.

[17] Allen Newell, Herbert Alexander Simon, et al. *Human problem solving*, volume 104. Prentice-hall Englewood Cliffs, NJ, 1972.

[18] Randall C O'reilly and Yuko Munakata. *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT press, 2000.

[19] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*, 2023.

[20] Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, et al. Tool learning with foundation models. *arXiv preprint arXiv:2304.08354*, 2023.

[21] Stuart Russell. *Artificial Intelligence: A Modern Approach, eBook, Global Edition*. Pearson Education, Limited, 2016.

[22] Stuart Russell. *Human compatible: Artificial intelligence and the problem of control*. Penguin, 2019.

[23] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*, 2023.

[24] Robert L Solso, M Kimberly MacLin, and Otto H MacLin. *Cognitive psychology*. Pearson Education New Zealand, 2005.

[25] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.

[26] Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. Large language models still can't plan (a benchmark for llms on planning and reasoning about change). *arXiv preprint arXiv:2206.10498*, 2022.

[27] Wendell Wallach and Colin Allen. *Moral machines: Teaching robots right from wrong*. Oxford University Press, 2008.

[28] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. Multimodal chain-of-thought reasoning in language models. *arXiv preprint arXiv:2302.00923*, 2023.

[29] Shoshana Zuboff. The age of surveillance capitalism. In *Social Theory Re-Wired*, pages 203–213. Routledge, 2023.