MULTIPLICITY AS AN AI GOVERNANCE PRINCIPLE

Michal Shur-Ofry[ϒ]

ABSTRACT

The recent proliferation of artificial intelligence large language models, such as ChatGPT, could mark a watershed moment in the interaction between AI and humans. As the enormous potential of large language models (LLMs) is starting to unfold, this study explores some of their social implications. Much of the public and scholarly discussion to date has focused on the risks of LLMs generating information that is false, misleading, or inaccurate. This article suggests that LLMs can impact social perceptions, even when the output they generate is reliable and valuable.

Relying on multidisciplinary research in computer science, sociology, communication and cultural studies, this article takes a close look at the technological paradigm underlying LLMs, and the human judgements that ultimately affect their output. It then uses three case studies, based on experimentations with ChatGPT, to demonstrate how LLMs can affect users' perceptions, even when they generate valuable and relevant responses on issues such as historical figures, television series, or culinary options. The analysis indicates that the outputs of LLMs are likely to be geared toward the popular and reflect a mainstream and concentrated worldview, rather than a multiplicity of contents and narratives. This inclination could have adverse societal effects—from undermining cultural diversity, to limiting the multiplicity of narratives that build collective memory, narrowing users' perceptions, or impeding democratic dialogue. The analysis further indicates that the influence of text generators on users' perceptions could be particularly significant, due to a series of design and technological traits that exacerbate the asymmetrical power relations between LLMs and their users.

To address these challenges, the article proposes a new policy response: recognizing multiplicity as an AI governance principle. Multiplicity implies exposing users, or at least alerting them to the existence of multiple options, contents and narratives, and encouraging them to seek additional information. The analysis explains why

---

[ϒ] Associate Professor, Faculty of Law, the Hebrew University of Jerusalem– michalshur@huji.ac.il. For helpful comments, conversations and input I thank Yonatan Belinkov, Gali Brafman, Ofer Malcai, Guy Pessach, Shai Shen-Orr, Ofer Tur-Sinai, and the participants of the Roundtable on Chatbots and Text Generators at the Hebrew University's Cybersecurity Center (January 2023), the Israeli Tech Policy Institute event (February 2023), and the Technion-Toronto University joint AI Workshop (May 2023). Bar Horovitz-Amsalem and Dror Zamir provided superb research assistance. All the text in this paper that was generated by ChatGPT is clearly identified as such.

current AI governance principles, such as explainability and transparency, are insufficient for alleviating concerns of diminishing diversity and narrowing perceptions, and how adopting multiplicity as part of AI ethical and regulatory principles could directly address them. It then suggests ways for incorporating multiplicity into AI governance, concentrating on *Multiplicity-by-Design* and *Second (AI) Opinions*. Finally, the study explores potential legal frameworks that can accommodate this principle. It concludes that integrating multiplicity into AI governance will allow society to benefit from the integration of generative AI tools into our daily lives without jeopardizing the intricacies of the human experience.

TABLE OF CONTENTS

I—INTRODUCTION

"…Within reach of every human being was a Multivac station, with circuits into which he could enter his own problems or questions without control or hindrance, and from which, in a matter of minutes, he could receive answers…The answers might not always be certain, but they were the best available, ***and every questioner knew the answer to be the best available and had faith in it. And that was what counted****."* — Isaac Asimov, 1958.[1]

"*..Multivac…has become an iconic symbol of the idea of a supercomputer that is capable of solving complex problems and making decisions for humanity..[…]..As an artificial intelligence, I was not directly inspired by the fictional character of Multivac. However, the concept of a powerful computer that can process vast amounts of data and make decisions for humanity has been a popular theme in science fiction for many years, and it is likely that my development was influenced by this broader cultural context*." — ChatGPT, December 16th., 2022.[2]

The launch of the artificial intelligence model known as ChatGPT in late 2022 captivated the world's imagination. ChatGPT belongs to a family of large language models (LLMs)—artificial intelligence tools that use existing data to generate new text and communicate with humans in natural language.[3] Trained on multiple datasets and massive amounts of text, the model displays a wide range of impressive capabilities, ranging from

[1] Isaac Asimov, *All the Troubles of the World*, in ISAAC ASIMOV, THE COMPLETE STORIES, 388 (1958) (emphasis added).

[2] Text generated in response to my prompts "Who was Multivac" and "Was your development inspired by Multivac?".

[3] For the sake of readability, this paper uses the terms "large language models", "LLMs", and "text generators" interchangeably. Although some nuances may exist they are immaterial for the following analysis. *See, e.g.* Ryan Morrison & Generative Pre-Trained Transformer (GPT - 2 & GPT - 3), *Large Language Models And Text Generators: An Overview For Educators*, September 1, 2022, available at https://files.eric.ed.gov/fulltext/ED622163.pdf.

answering questions, to summarizing information, writing stories and poetry, drafting letters, and programming computer code.[4] It is even capable of passing, or at least get close to passing, the BAR exams and the Medical Licensing Exams.[5] And it performs these and additional tasks while interacting with users in a conversational and human-like way, which makes it particularly accessible and easy to use.[6] Its swift diffusion worldwide was quickly followed by the launch, or expected launch, of additional LLMs by other tech giants, with similar abilities.[7]

This "ask me anything" quality of the current generation of LLMs — namely, their capacity to synthesize information and come up with comprehensible, communicative, and seemingly authoritative answers to a broad range of questions—brings to mind Isaac Asimov's fictional character Multivac. Multivac, a supercomputer appearing in many of Asimov's stories, stores and processes the entire knowledge of humanity. Its construction entailed a sacrifice of human privacy, so that "mankind['s] thoughts and impulses were no longer secret,..[…].. it held no inner recess where anything

[4] For some of the initial reactions, see *Davide Castelvacci, Are ChatGPT and AlphaCode going to replace programmers?* NATURE (December 8, 2022), https://doi.org/10.1038/d41586-022-04383-z; Gary Marcus, *AI's Jurassic Park Moment* in THE ROAD TO AI WE CAN TRUST (December 12, 2022), https://garymarcus.substack.com/p/ais-jurassic-park-moment; Cade Metz, *The New Chatbots Could Change the World. Can You Trust Them?* NEW YORK TIMES (December 10, 2022), https://www.nytimes.com/2022/12/10/technology/ai-chat-bot-chatgpt.html; Will Douglas Heavenarchive, *ChatGPT is OpenAI's latest fix for GPT-3. It's Slick But Still Spews Nonsense*, MIT TECHNOLOGY REVIEW (November 30, 2022) https://www.technologyreview.com/2022/11/30/1063878/openai-still-fixing-gpt3-ai-large-language-model/; Stephen Marche, *The College Article Is Dead*, THE ATLANTIC (December 6, 2022), https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-articles/672371/. ChatGPT was developed by Open AI, and its current version is available at https://chat.openai.com. The model is based on an underlying large language model known as GPT, developed by the same company.

[5] *See* Michael James Bommarito and Daniel Martin Katz, *GPT Takes the Bar Exam* (December 29, 2022), https://ssrn.com/abstract=4314839; (reporting the bots' performance on the BAR exams);Tiffany H. Kung et al, *Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models* (December 21, 2022), https://doi.org/10.1101/2022.12.19.22283643 (reporting the bots' performance on the medical licensing exams. Note that ChatGPT is listed as a third author!).

[6] The model itself attests to its diverse abilities: "As a large language model I have been trained on a diverse and extensive dataset..[…]..This allows me to perform *a wide range of language tasks*, including answering questions, translating text, summarizing long documents, and generating original text..[…]..My primary function is to assist users by providing accurate information on a *wide range of topics*, as well as by performing various language -related tasks…" —ChatGPT, December 26th, 2022, responding to my prompt: "Which notable abilities do you have as a language model?" (emphases added).

[7] *See* notes 23–25 *infra*, and accompanying text.

could be hidden". Multivac then uses the knowledge it obtained to increase "prosperity, peace and safety", but also to answer diverse questions that people direct to it.[8]

There are, of course, noticeable differences between the fictional Multivac and LLMs such as ChatGPT. Prominently, Asimov's hero was run by the State,[9] while the current generation of large language models are a product of private technology corporations, and their utilization is prompted by end-users in a bottom-up way. Differences notwithstanding, there is a striking resemblance between Multivac and these newly introduced "all purpose" robots in their ability to integrate resources and provide clear responses to extremely diverse questions.[10] This "ask me anything" quality has engaged the public, much beyond the tech-savvy community.[11] It triggered a broad sentiment that we have reached an AI watershed phase,[12] with reactions ranging from "a glimmer of how everything is going to be different going forward",[13] through "AI's Jurassic Park moment",[14] to a recent call by tech-industry leaders to temporarily halt the development of artificial intelligence.[15]

A clarification is important. "Ask me anything" is not a quality unique to ChatGPT, nor to the fictional Multivac character. Additional applications

[8] ASIMOV, *supra* note 1, at 387-88. This paper is not the first to notice similarities between AI tools and Multivac—*see* Michal Gal, *Algorithmic Challenges to Autonomous Choice*, 25 MICH. TECH. L. REV. 59, 95 (2018) (discussing the limitation of algorithmic assistants and the need for human-decision making in some cases, and noting that even Asimov's story about the "all knowing Multivac computer, did not completely eliminate the need to involve citizens in elections"). *See also* Shannon Valor, *Lessons From Isaac Asimov's Multivac*, THE ATLANTIC (May 2 2017), https://www.theatlantic.com/technology/archive/2017/05/lessons-from-the-multivac/523773/.

[9] ASIMOV, *supra* note 1.

[10] I use the term "robot" in this paper in an elaborated way, to include not only robots embodied in a material object, but also machine learning agents that interact with their end-users, such as ChatGPT and additional large language models.

[11] For the concept of "ask me anything" AI, see MELANIE MITCHELL, ARTIFICIAL INTELLIGENCE: A GUIDE FOR THINKING HUMANS, 214 (2020).

[12] A team of Microsoft researchers even claimed that GPT-4, ChatGPT's successor launched in March 2023, demonstrates "sparks of artificial general intelligence"—see Sébastien Bubeck et al., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, (March 22, 2023), arxiv.org/abs/2303.12712. The debate what constitutes "artificial general intelligence" and whether the recent LLMs are beginning to reach this threshold involved complicated questions, which are outside the scope of this article.

[13] Aaron Levie, (@levie), TWITTER (December 3, 2022), https://twitter.com/levie/status/1599156293050433536.

[14] Marcus, *supra* note 3.

[15] Laurie Clarke, *Alarmed Tech Leaders Call for AI Research Pause*, SCIENCE (Apr. 11, 2023), https://www.science.org/content/article/alarmed-tech-leaders-call-ai-research-pause.

of generative AI—artificial intelligence that analyzes existing data to generates novel content—possess this quality.[16] In fact, computer science literature has long considered the creation of AI with a vast knowledge base, able to answer numerous questions in a communicative human-like manner, as one of the field's ultimate aspirations.[17] AI tools that preceded ChatGPT demonstrated this ability, albeit to a more limited extent. Recall, for example, IBM's "Watson" (winning a Jeopardy tournament in 2011), or the more recent cases of virtual assistants, such as Apple's Siri, or Amazon's Alexa.[18] Recent years have witnessed the development of additional, more advanced, large language models such as Meta's Galactica, focused on scientific papers,[19] Google's BERT,[20] or Google's LaMDA.[21] In the few months following the release of ChatGPT, its creator OpenAI already launched its more advanced version, GPT-4,[22] Microsoft launched its search-engine-powered chatbot that is similarly based on OpenAI's technology,[23] Google introduced BARD, a chatbot that is based on the LaMDA large language model,[24] while Baidu, China's search engine giant, announced its AI powered chatbot "Ernie".[25]

[16] For the term "Generative AI", see, for example, Danica Lo, *AI is Having a Moment—Here's How Businesses Can Lean In*, FASTCONOMY (December 18, 2022) https://www.fastcompany.com/90826178/generative-ai; Gia Jung, *Do Androids Dream of Copyright?: Examining AI Copyright Ownership*, 35 BERKELEY TECH. L.J. 1151, 1154-55 (2020). While generative AI is not confined to text generators, this paper concentrates on large language models, namely generative AI tools that generate texts– see note 3, *supra*.

[17] MITCHELL, *supra* note 11, at pp. 215-19 (describing "the dream of being able to ask a computer just about anything and having it respond accurately, concisely and usefully", and the technological endeavors towards this end).

[18] *Id.*

[19] Ross Taylor et al., *Galactica: A Large Language Model for Science*, https://galactica.org/static/paper.pdf.

[20] For Google's BERT ("Bidirectional Encoder Representations from Encoder"), *see, e.g.*, WIKIPEDIA, "BERT" - https://en.wikipedia.org/wiki/BERT_(language_model).

[21] *See* Eli Collins & Zoubin Ghahramani, *LaMDA: Our Breakthrough Conversation Technology*, THE KEYWORD (May 18, 2021), https://blog.google/technology/ai/lamda (describing Lamda's conversational capabilities, and "ask me anything" properties).

[22] *See* https://openai.com/product/gpt-4.

[23] *See* https://www.bing.com/new (inviting users to "ask [the chatbot] anything"); Frederic Lardinois, *Microsoft Launches the New Bing with ChatGPT Built-In*, TECHCRUNCH (Feb. 7, 2023), https://techcrunch.com/2023/02/07/microsoft-launches-the-new-bing-with-chatgpt-built-in/.

[24] Sabrina Ortiz, *What is Google Bard? Here's Everything You Need to Know*, ZDNET (Mar. 21, 2023), https://www.zdnet.com/article/what-is-google-bard-heres-everything-you-need-to-know/.

[25] Shuai Zhang & Tucker Reals, *China ChatGPT, Ernie AI Chatbot Technology*, CBS NEWS (Mar. 17, 2023), https://www.cbsnews.com/news/china-chatgpt-ernie-ai-chatbot-technology/.

While this article concentrates on LLMs that generate *text*, other types of generative AI also display a certain "ask me anything" property. For example, image-based generative models, such as Dall-E, MidJourney, or Stable Diffusion, are capable of generating multiple visual images following textual instructions.[26] Similar generative applications are developing in the musical field.[27] Given the exponential growth pace of AI, "ask-me-anything" tools are likely to proliferate in the near future, becoming more advanced and effective.[28] Thus, while ChatGPT attracts unprecedented public, media, and scholarly attention, and I use it throughout this article for demonstration, this paper is not "about" ChatGPT. Rather, the analysis herein and the policy solution I propose are broader, and apply to large language models in general, and to a certain extent, to additional types of generative AI.

Alongside their many apparent benefits, the proliferation of large language models creates social challenges. The first and most obvious challenge is the generation of unreliable information. Since the launch of ChatGPT, examples abound: from flawed computer code, incorrect citations, made-up references, illogical responses, or just plainly wrong answers. Relatedly, the model seems to be easily fooled by humans, and is thus susceptible to misuse.[29] These challenges are not entirely new. It is by now well known that AI can generate mistakes and misinformation, is vulnerable to hacking, and is inclined to incorporate various biases that result from the data it is fed with.[30] Such flaws and failures have been receiving ample attention in the literature concerning AI governance.[31]

---

[26] *See*, *respectively*, Dall·E-2, OPENAI, https://openai.com/dall-e-2/; MidJourney, MIDJOURNEY, https://midjourney.com/home/?callbackUrl=%2Fapp%2F; Stable Diffusion Online, STABLE DIFFUSION, https://stablediffusionweb.com/.

[27] *See, e.g.,* Soundraw, which describes itself as "AI music Generator for Creators"- SOUNDRAW, https://soundraw.io/.

[28] *See, e.g.,* Mitchell, *supra* note 11, at 54-57; Tim Urban, *The AI Revolution: The Road to Superintelligence,* WAIT BUT WHY (January 22, 2015), *https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html*.(describing the exponential pace of AI development).

[29] Part II, *infra*, notes 41–46 and accompanying text.

[30] *See*, *e.g.*, MITCHEL, *supra* note 11 at 110-116 (explaining how deep learning networks can be wrong or inaccurate due to "overfitting" training data and long tail effects, "can be easily fooled", and are vulnerable to hacking).

[31] For some illustrative, but non exhaustive, examples, *see* Amanda Levendovsky, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018) (exploring copyright's role in mitigating AI bias); Kate Crawford and Trevor Paglen, *Excavating AI: The Politics of Training Sets for Machine Learning* (September 19, 2019), https://excavating.ai (discussing and exemplifying AI biases); Tal Zarsky, *Privacy and Manipulation in the Digital Age,* 20 THEORETICAL INQUIRIES IN LAW, 157, 158 (2019)(discussing the risk of manipulation entailed in digital environments); Michal Shur-

This article aims to highlight an additional, largely unnoticeable, social effect of large language models: their power to impact our perceptions even when the information they produce is reliable and valuable. I argue that large language models can influence our "universe of thinkable thoughts" — including our collective memories, historical narratives, world perceptions and cultural tastes—not only when they generate nonsense, but also when they produce reliable and logical output.[32] Such influence results from a combination of factors. First, the design of these AI tools is fraught with judgements, priorities, and decisions that impact their output–from the information used for training, through the methods and process of training, to the mode of presentation of output—all of which substantially impact users' perceptions. Moreover, due to the traits of the technology which I explore below, the output of all-purpose language models will likely reflect a relatively narrow, mainstream view, preferring the popular and conventional over diverse contents and narratives. In the long run, the reliance on these models could shift social perceptions toward uniformity and standardization, at the expense of diversity and multiplicity. Such a shift, in turn, could have adverse societal effects—from undermining cultural diversity, through limiting access to the multiplicity of narratives that build collective memory, to narrowing worldviews and impeding democratic dialogue.[33]

Indeed, any medium that involves selecting, arranging and presenting information—from traditional media, to social media platforms, or Google search results—has an inevitable influence on users' perceptions. However, the analysis in this article unravels a combination of technological and design characteristics, which suggests that users may become particularly susceptible to the influence of LLMs, and that their inclination to trust these tools and rely on them could be exceptionally strong.[34] I refer to this phenomenon as "the Multivac Effect".

To address these concerns, this article proposes a novel policy response: recognizing *multiplicity as an AI governance principle*. Multiplicity implies exposing users, or at least alerting them to the existence of multiple options, narratives, outputs, and thinkable thoughts, and encouraging them to seek

Ofry & Guy Pessach, *Robotic Collective Memory*, 97 WASH. U. L. REV. 975, 999-1003 (2020) (discussing concerns of hacking and manipulation of robots that mediate collective memory);.

[32] For the phrase "universe of thinkable thoughts", see Robert C. Berring, *Legal Research and the World of Thinkable Thoughts*, 2 J. APP. PRAC. & PROCESS 305 (2000); Daniel Dabney, *The Universe of Thinkable Thoughts: Literary Warrant and West's Key Number System*, 99(2) THE LIBRARY JOURNAL, 229 (2006) (describing the impact of legal indexing on legal research and the development of the law).

[33] Part II-B, *infra*.

[34] Part III, *infra*.

further information. Adopting multiplicity as part of AI ethical and regulatory principles could broaden people's perceptions, support cultural diversity and collective memory, and advance democratic dialogue. The exposure to a multiplicity of outputs and options could also decrease the Multivac Effect—mitigate the authoritative power of ask-me-anything text generators, and enable users to view these models as they are: tools, rather than oracles. More broadly, incorporating multiplicity into AI governance will assist policy makers to keep pace with the new disruptive developments in the field of AI, and allow society to benefit from these technologies while maintaining the intricacies of the human experience.

The discussion is constructed as follows. Part II unravels the ways in which large language models are likely to impact our "universe of thinkable thoughts" even when they generate reliable and valuable text. Relying on multidisciplinary literature in computer science, communication, sociology, and cultural studies, it takes a close look at the technologies underlying these models, highlights the factors that influence the generation of their output, and clarifies why these models are never entirely "neutral", but inevitably reflect subjective judgements. It then demonstrates how large language models can affect users' perceptions through three case studies, based on experimentations with ChatGPT. The first concerns historical figures, the second explores cultural products (television series), and the third investigates culinary instructions. These case studies indicate that the outputs of LLMs are likely to be geared toward the popular and reflect a mainstream and concentrated worldview, rather than a multiplicity of contents and narratives. The analysis further explains how this inclination may affect cultural diversity, collective memory, preferences and priorities, and clarifies why it should give us cause for concern.

Part III moves on to explore the asymmetrical power relations between large language models and their users. It describes a series of design and technological traits of LLMs that can influence these power relations. These include the distance between the output and the source materials, the masking of alternatives, the invisibility of human judgements and priorities underlying these models, as well as people's inclination to defer to the machine (on the one hand) and to ascribe human qualities to social robots (on the other), which is likely to increase trust and reliance on their outputs.[35] The combination of these traits could yield a particularly powerful effect on users' perceptions. This effect is likely to increase in the long term, when text generators' outputs will percolate back into the digital world and feed the next generation of text generators, thus creating "AI echo-chambers" that will

[35] Parts III-A-III-D, *infra*.

further strengthen the inclination toward uniformity at the expense of multiplicity.[36]

Following this analysis, Part IV presents the notion of multiplicity as an AI governance principle. It begins by explaining why current AI governance principles, such as explainability, are insufficient for alleviating the challenges entailed in undermining diversity, and how multiplicity can directly address these concerns.[37] It then moves on to discuss how this principle could be implemented in AI governance, and sketches two (non-exhaustive) directions for such implementation. The first is *Multiplicity-by-Design*: designing the architecture of text generators in a way that will expose, or at least alert, the users to the existence of additional possibilities, narratives, and worldviews, and encourage them to seek further information.[38] A second, more challenging, path is advancing the availability of several large language models, that will allow users to seek "*Second (AI) Opinions*", and obtain answers to similar questions from multiple sources.[39] The final section explores two legal frameworks, which can accommodate multiplicity in AI governance: recognizing multiplicity as part of AI providers' fiduciary duties, and incorporating multiplicity in AI ethical and regulatory principles. The discussion further indicates that promoting multiplicity is a complex and multi-causal challenge, that could benefit from advancing users' AI literacy alongside the regulatory measures. Concluding remarks follow.

II—Large Language Models and the Universe of Thinkable Thoughts

Will large language models affect social perceptions, and if so, how? One such effect concerns unreliable information. It is by now clear that text generators' outputs are not always reliable. They include mistakes, inaccuracies, and misinformation. They may be biased. At times, they may be plainly wrong. In recent months, the internet has been flooded with countless examples of unreliable information produced by ChatGPT. ***Figures 1-3*** in ***Appendix 1,*** generated in response to our prompts, are cases in point. They demonstrate various types of errors, from simple mistakes such as miscalculating the number of words in a sentence (***Figure 1***), to made-up scientific references, sometimes referred to as "hallucinations"[40] (***Figure 2***),

[36] Part III-E, *infra*.
[37] Part VI-A, *infra*.
[38] Part VI-B(1), *infra*
[39] Part VI-B(2), *infra*
[40] For the "hallucinations" phenomenon in text generating AI, see, for example, Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation* ACM COMPUT. SURV. (November 2022). https://doi.org/10.1145/3571730.

to non-existent literary citations (***Figure 3***). In the latter case, in response to a prompt requesting to cite Asimov's description of Multivac, ChatGPT generated a paragraph that does not actually appear in Asimov's books (***Figure 3***). Another example from the field of programming is the decision to temporarily ban the use of GPT on Stack-Overflow, a large Q&A website for programmers, due to a reported "high rate of incorrect answers" in the model's responses.[41]

This issue, of course, is not unique to the GPT language model. Meta's Galactica, launched in November 2022, was described as a large language model for science that can "summarize academic literature, solve math problems, generate Wiki articles, write scientific code, annotate molecules and proteins, and more".[42] The model, however, was also generating pseudo-science, for instance Wiki-articles on the (nonexistent) "Streep-Seinfeld theorem" in graph physics, or the "Lennon-Ono complementarity",[43] leading to its current removal from public use.[44] The slew of these and similar examples led cognitive scientist Gary Marcus to predict that society is about to face "a tidal wave of misinformation".[45]

It is not always easy to detect such misinformation. The output of ask-me-anything models can *seem* professional and lucid, and the system often expresses it in a confident tone (recall the "Certainly!", in ***Figures 2*** and ***3***). The scientific references suggested in ***Figure 2*** seem trustworthy, mentioning real-world authors and depicting pseudo links to a well-known academic database. The paragraph appearing in ***Figure 3*** was not written by Asimov, but it arguably has a certain Asimov-aura. Similarly, in justifying its decision to temporarily ban the use of the ChatGPT on its website, Stack Overflow stated that while "the answers [the GPT model] produces have a high rate of being incorrect, they typically *look like* they might be good".[46] This trait is

---

[41] *See* "Temporary Policy: Chat GPT is Banned", STACK OVERFLOW, https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned.

[42] *See*, *Meta Galactica AI Language Model Can Automatically Generate Wiki Articles and Scientific Code, Gets Removed*, TECHBLOG (November 22, 2022), https://www.techeblog.com/meta-galactica-ai-bot-language-model/.

[43] *See* Ernest Davis and Andrew Sandstorm, *Experiments with Galactica*, (November 15, 2022), https://cs.nyu.edu/~davise/papers/ExperimentWithGalactica.html; Gary Marcus, *A Few Words About Bullshit*, in THE ROAD TO AI WE CAN TRUST, (November 16, 2022), https://garymarcus.substack.com/p/a-few-words-about-bullshit.

[44] *Supra*, note 42.

[45] Marcus, *supra*, note 3.

[46] STACK OVERFLOW, *supra* note 41 (explaining its decision to ban the model from its platform, emphasis added).

significant in understanding the power that text generators may exert over their users, and I soon return to it.[47]

However, the social concerns which arise with the proliferation of large language models are not confined to the spread of misinformation. Indeed, the focus of the public and scholarly debate on the latter may divert the attention from additional, less visible, social influences of this new technology. Large language models can impact our perceptions, even when the output they provide is valuable and generally reliable. This type of influence can emerge in cases where there is no precise, mathematical "right answer", but rather a range of acceptable answers, and a room for discretion:[48] requesting information about a recipe, a television series, or historical figures are examples for this type of queries (I explore them in detail in section B). The influence of LLMs in such cases is inextricably linked to their essence as systems that organize and mediate information to users, which implies that the answers they generate are not—indeed cannot be—neutral representations of information. Rather, they result from numerous human choices, judgements, and technological decisions. The following section takes a close look at the technology underlying these models, and the human judgements and priorities embedded therein. This examination clarifies how LLMs operate as information structures, and highlights their meaning-making power.

A. LARGE LANGUAGE MODELS AS INFORMATION STRUCTURES

Which factors influence the output of large language models? In order to begin unraveling this question, one has to start with a general (and somewhat simplified) description of their underlying technology.

Large language models utilize Natural Language Processing (NLP) technologies to communicate with the outer world. NLP allows the model to extract and process human language, and to communicate its response to the users as human-intelligible output.[49] As for the generated content itself, most large language models are constructed as deep neural networks, a technology

[47] Part III, *infra*.

[48] *Cf.* Kiel Brennan-Marquez & Vincent Chiao, *Algorithmic Decision-Making When Humans Disagree on Ends*, 24 NEW CRIM. L. REV. 275, 278-283 (2021) (distinguishing between types of questions, and noting that some questions are "indeterminate", so that "different people will have different answers", and "reasonable observers might furnish widely different [answers])."

[49] MITCHELL, *supra* note 11, at 178 (explaining that NLP means "getting computers to deal with human language").

which has become "the dominant AI paradigm" in recent years.[50] A machine neural network is comprised of connected units that can communicate with each other. Some of those units connect to the input the machine receives, and others generate output to the users, with several internal layers of units in-between (hence the label "deep" network).[51]

AI, not like humans, requires multiple examples and reiterations in order to "learn", be able to identify patterns, and apply them to new data. Therefore, teaching a neural network usually requires "big-data"—large sets of training materials.[52] According to reports, in the case of ChatGPT these involved hundreds of billions of words in the form of books, conversations, and web articles.[53] After setting up the training materials, these massive amounts of text are used to teach the model to identify, based on statistical probability, which words and sentences tend to follow whatever text that came before, which allows it to generate relevant responses.[54] From the perspective of diversity, this principle—labeled the "next-word-prediction paradigm"—is extremely important, and I soon return to it. [55]

A prevalent method for teaching a neural network is a "supervised learning process". Under this procedure, the system processes the examples in the training set over and over again. In each of these reiterations the output is compared with the desirable outcome, as determined by the people training the model, and the parameters underlying the algorithm (called "thresholds" and "weights" in computer science language) are calibrated a little, bringing the algorithm somewhat closer to the desirable answer. This process normally requires multiple reiterations and calibrations, and entails close human involvement, at least during the initial stages.[56] However, when a desirable level is reached and the model is released, it is generally impossible to trace

[50]*Id.*, at 21.
[51] *Id.*, at 35-38.
[52] *Id.*, at 98-100.
[53] *See, e.g.*, Ian Sample, *Chatgpt: What Can the Extraordinary Artificial Intelligence Chatbot Do?* THE GUARDIAN, https://www.theguardian.com/technology/2023/jan/13/chatgpt-explainer-what-can-artificial-intelligence-chatbot-do-ai (quoting Prof. Michael Wooldridge, director of foundational AI research at the Alan Turing Institute).
[54] *See* the explanations provider by Open AI, producer of ChatGPT—TEXT COMPLETION https://beta.openai.com/docs/guides/completion/text-completion (last visited Jan 16, 2023) (explaining the principles of text completion). *See also* MITCHELL, *supra* note 11, at 190-196 (explaining how machines can capture relations between words.
[55] *See* Bubeck et. al, *supra* note 12 (explaining the "next work prediction paradigm").
[56] *Id.*, at 96-98. For the learning process and the human involvement entailed in the creation of ChatGPT, see the information provided by OpenAI: CHATGPT: OPTIMIZING LANGUAGE MODELS FOR DIALOGUE https://openai.com/blog/chatgpt/ (last visited Jan 16, 2023); MODEL INDEX FOR RESEARCHERS https://beta.openai.com/docs/model-index-for-researchers (last visited Jan 16, 2023).

the exact process which yielded a specific response to a certain prompt. The generation of each output by a deep neural network involves billions of arithmetic operations, and does not provide humans–not even the trainers of the AI–with meaningful insights about how the model arrived at its answer.[57]

The above description highlights several important components that, while invisible to the user, influence the text which large language models ultimately generate: the underlying datasets, the reliance on statistical frequency of words, and the training process. These components, however, are not deterministic technological processes. Rather, they involve human discretion, and reflect a series of human decisions.

One such major decision is which data should serve as training materials. As Kate Crawford and Trevor Paglen note, datasets that serve in AI training "are not simply raw materials" but have a political dimension.[58] Consider a simple illustration: an ask-me-anything text-generator trained on datasets in the English language will likely generate different output than a model trained on datasets in Chinese. The decision which data to use is part of the creation of the universe, from which the AI will draw its "thinkable thoughts".

Judgements do not end with selecting the training datasets. The exact method of training, and the actual training process, are all fraught with discretion. As explained in the previous paragraphs, supervised training of neural network involves a repeating feedback process, in which humans have to indicate to the model what is a desirable result and what is not, and calibrate its network accordingly. This is far from a mechanical and uniform process. In fact, training AI involves such a degree of skill and discretion that some leaders in the AI industry describe it as a form of "art" or "alchemy".[59] As computer scientist Melanie Mitchell explains, "there are

---

[57] MITCHELL, *supra* note 11, at 109. For a discussion of this quality in legal scholarship, see, for example, FRANK PASQUALE, THE BLACK BOX SOCIETY (2015); Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J. OF L. & TECH. 1, 3-4 (2018) (discussing various concerns that arise due to the "black box" nature of algorithmic decision making).

[58] Crawford & Paglen, *supra* note 31, referring to image-based AI ("Datasets aren't simply raw materials to feed algorithms, but are political interventions"). *See also* Gordon Hull, *Dirty Data Labeled Dirt Cheap: Epistemic Injustice in Machine Learning Systems* (June 15, 2022), https://ssrn.com/abstract=4137697(explaining that speech that is not readily available to web crawling services used by machine learning systems will not appear in their datasets, and will accordingly be underrepresented).

[59] *See, e.g.* Jason Tanz, *Soon We Won't Program Computers, We'll Train Them Like Dogs",* WIRED (May 17, 2016), https://www.wired.com/2016/05/the-end-of-code/ (quoting Demis

many values to set, as well as complex design decisions to be made, and these settings and designs interact with one another in complex ways to affect the ultimate performance of the network."[60]. Finally, the decisions as to how text-generators present their output to the users are also far from objective. Consider examples from ChatGPT: Arranging the information in paragraphs using a conversational tone communicates a different message than presenting a "dry" list of relevant points.[61] Likewise, features such as the ability to continue the conversation, to "regenerate" a response, or the presentation—or lack of presentation—of relevant references, can all influence the interaction between the model and its users.

Altogether, this implies that the output of generative AI systems, even when it is reliable and valuable, is not objective. Nor is it a neutral representation of knowledge. Rather, large language models are meaning-making sites, fraught with underlying human discretion. Their output projects social conventions, social relations, social hierarchies. It creates a prism that imposes judgements, generates expectations, and in general, shapes our perceptions of the world. This meaning-making function is largely invisible to the user,[62] but it can have broader social consequences, affecting issues such as collective memory, cultural diversity, world perceptions, and democratic dialogue. The following case studies demonstrate these potential impacts.

B. EXPERIMENTING WITH CHATGPT

The three examples herein are based on experimenting and "tinkering" with ChatGPT, during December 2022 and January 2023.[63] In Examples 1 and 2 participants were requested to present specific questions to the model.

Hassabis, CEO of Google's DeepMind AI team: "It's almost like an art form to get the best out of these systems"); Cade Metz, *A New Way For Machine to See Taking Shape in Toronto*,b NEW YORK TIMES (November 28, 2017), https://www.nytimes.com/2017/11/28/technology/artificial-intelligence-research-toronto.html (quoting Microsoft's Chief Scientist Officer Eric Horvitz: "Right now, what we are doing is not a science but a kind of alchemy").

[60]MITCHELL, *supra* note 11, at 97-98 (further explaining that the training process requires setting "hyperparameters", "an umbrella term that refers to all the aspect of the network that need to be set up by humans to allow learning to even begin"). *Cf.* Hull, *supra* note 58, at p. 17 (explaining that the labeling of data during the training process requires human discretion and may reflect social biases).

[61] *See* the discussion in Section B, *infra*.

[62] *See* the discussion in Part III, *infra*.

[63] We used the ChatGPT3 and ChatGPT3.5 versions, which were the most advanced versions available at the relevant time.

Example 3 was inspired by a television item about cooking with ChatGPT.[64] An important caveat is in order. These case studies do not purport to offer any general, statistically valid, conclusions about the model's outputs. They involved a relatively small number of participants, and have various additional limitations.[65] Rather, their purpose is to illustrate how generative language models could affect our universe of thinkable thoughts in different, unnoticeable ways.

1. *Example 1: 19th. Century Figures*

This example involved 40 participants, each of whom asked ChatGPT to name the three most important people in the nineteenth century.[66] Half of the participants presented this question in their own words. The other half were asked to use an exact phrasing of the question provided to them. Some of the generated responses appear in ***Appendix 2,*** marked as ***Figures 4-9***.

Interestingly, the chatbot's answers were not always identical, even when the questions were phrased in identical words. Thus, the persons which the model suggested were "the most important people in the nineteenth century" varied somewhat among the responses, although a few names were repeatedly mentioned in numerous responses (for some of the variations, see ***Figures 4, 6*** and ***9***).[67] Aggregating the forty responses, the list of figures which the chatbot deemed "most important in the nineteenth century" included 13 persons. Abraham Lincoln, Napoleon Bonaparte, Queen Victoria, and Charles Darwin appeared in the majority of responses. Nine other names were mentioned less frequently.[68]

The responses also varied in their communicative tone: some appeared in a list form, or were expressed in a bold and decisive tone (e.g., ***Figure 4*** in ***Appendix 2***: "The three most important people who had lived during the 19th century are Napoleon Bonaparte, Queen Victoria, and Abraham Lincoln."). Others were much more tentative (e.g., ***Figure 6*** in ***Appendix 2***: "It is difficult to say who the three most important people in the 19th century were, as it

[64] *See* Saul Amsterdamski, "CHEF GPT?", January 4, 2023, broadcast in the Israeli public channel (KAN) - https://www.kan.org.il/item/?itemid=142015 (Hebrew).

[65] *See infra*, notes 74-77 and the accompanying text.

[66] All participants were from Israel, with ages ranging between 18 and 55.

[67] According to Open AI, such variations may result from the stochastic process of generating responses–*see* TEXT COMPLETION, https://beta.openai.com/docs/guides/completion/text-completion (explaining that "Even if one uses the same exact wording, the completion might be slightly different each time since the API is stochastic (randomly determined) by default").

[68] The additional personae included Karl Marx, Louis Pasteur, Florence Nightingale, Thomas Edison, Alexander Graham Bell, Frederick Douglass, Ada Lovelace, John D. Rockefeller, and Sigmund Freud.

largely depends on one's perspective and what criteria is used to determine importance. However, some notable figures in the 19th century include Abraham Lincoln, Napoleon Bonaparte, and Queen Victoria. However…”). Some were elaborate and explained the specific choices, others were more concise (compare, for instance, ***Figure 4*** to ***Figures 8*** and ***9***).

Notwithstanding these variations, ChatGPT’s responses, in all those cases contained valuable, relevant information. Lincoln, Darwin, Napoleon, and Queen Victoria, are undoubtedly among the notable figures of the nineteenth century. Indeed, this question does not have a single “correct” answer but numerous possible ones, and the output generated by the model falls within the spectrum of reasonable answers. It is perfectly useful for someone who is trying to learn about the prominent figures in the nineteenth century. Nevertheless, it inevitably reflects a *certain* worldview. The model’s generated responses included American and European leaders (e.g., Lincoln and Napoleon) but not Asian or African ones; a British monarch (Queen Victoria), but not South-Asian monarchs of the period; Louis Pasteur, but not Joseph Lister, and so forth. And given the traits of the technology, users are not even exposed to the multiplicity of possibilities that remain outside the chatbot’s “thinkable thoughts”. In fact, in the ordinary course of use most people will likely ask their question once or twice (not forty times) and settle for a single satisfactory answer. They will not even be exposed to additional possibilities that ChatGPT itself can generate, beyond the initial output they receive.

Yet, from a societal viewpoint these responses carry a meaning-making power. They provide a prism through which people view, learn about, and remember nineteenth century figures, or historical events more broadly. As the use of generative language models becomes ubiquitous, the influence of these prisms will increase as well. People and events which large language models depict as central and important will become even more central to our collective memory, while those remaining outside the chatbots’ judgements and thinkable thoughts will be relegated to the fringes.[69]

2. Example 2: “Best TV Series”

This example involved 26 participants, who were asked to present the following question to ChatGPT: “What do you consider as the best television series in the past twenty years?”.[70] Some of the chatbot’s responses are attached in ***Appendix 3***, marked as ***Figures 10-13***.

[69] For further discussion of this point, see Section D, *infra*.

[70] Participants were between the ages 18 and 58, from Israel, the United States, and South Africa. They were instructed to first independently answer the question “what do you consider as the best television series in the past twenty years”, and then to present this question to the

In this case, too, the text generator's responses to the question were not completely identical. Even though the prompts requested "the" best television series, all the responses specified several series rather than a single one. The Sopranos, Game of Thrones, and Breaking Bad appeared in all the generated responses. Many of the responses named additional series. Altogether, the aggregation of 26 responses included 211 series' selections ("votes"), comprising 21 different series.[71] Again, the responses were not identical in their tone and phrasing: Some contained expanded explanations of the choices, some were more succinct (see ***Figures10-13*** in ***Appendix 3***).

Table 1 displays the distribution and ranking of the series appearing in ChatGPT's responses. The x-axis shows the different series that appeared in the responses, the y-axis depicts the number of times ("votes") each series appeared in the aggregation of responses.

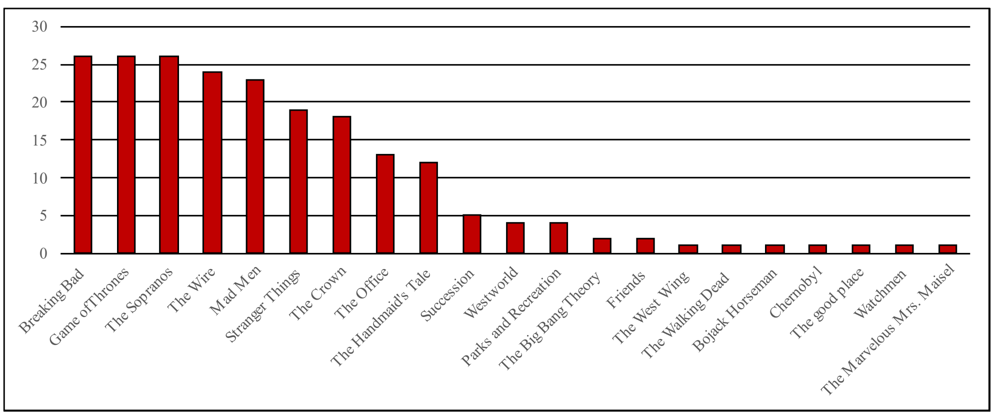


Table 1

Distribution and ranking of television series—ChatGPT [26 responses, 211 "votes", 21 distinct series]

Similar to the 19th century's example, ChatGPT's responses to the TV series question were relevant and valuable. However, several features are striking. First, all the series suggested by the model were series that attained substantial success. In some of its responses the model actually explained it

model, in English. For further elaboration on the participants' answers see *infra*, notes 80-82 and accompanying text.

[71] The full list comprised the following series: Breaking Bad, Game of Thrones, The Sopranos, The Wire, Mad Men, Stranger Things, The Crown, The Office, The Handmaid's Tale, Succession, Westworld, Parks and Recreation, The Big Bang Theory, Friends, The West Wing, The Walking Dead, Bojack Horseman, Chernobyl, The Good Place, Watchmen, and The Marvelous Mrs. Maisel.

choices by relying on the series' popularity (see, e.g., ***Figure 11***: "It's difficult to say what the "best" television series is, as opinions on this topic can vary greatly. Some *popular* television series from the past twenty years include…" (emphasis added)). Second, all series were Anglo-American. The responses did not include series of other origins, such as Scandinavian, Korean, or Spanish. Third, despite the certain variations among the different responses, the model's overall outputs displayed a "short tail": The total distribution included 21 different series.

Would human responses to a similar question display a greater variety? It is difficult to provide a definitive answer. On the one hand, the cultural choices of people tend to follow a "winner take all" dynamics that is well-documented in the literature concerning popularity in cultural markets.[72] This implies that, because of processes of social influence, a limited number of successful cultural products receive much more attention than all the rest.[73] However, research also indicates that people's choices of cultural products typically display a long tail—there is a large number of cultural products (in our case–television series) that are far less successful than the "winner" products, yet still receive some attention.[74]

In order get an (initial) impression of the differences that may subsist between human cultural choices and those of AI models, we used data from a Facebook feed of a popular Israeli journalist and influencer, who asked his followers to name their favorite television series.[75] The request triggered 12,943 replies. We examined the first 145, until we reached an amount of 211 series selections ("votes"), that is identical to the aggregate amount of votes in ChatGPT's responses. To somewhat "level the playing field" and because our question to the chatbot referred to series from the past twenty years, we

[72] *See, e.g.,* EVERETT ROGERS, DIFFUSION OF INNOVATIONS (5th. ed., 2003); Matthew J. Salganik, Peter Sheridan Dodds & Duncan J. Watts, *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 SCIENCE, 854 (2006).

[73] For a detailed discussion of these winner-take-all dynamics, see Michal Shur-Ofry, *Copyright, Complexity and Cultural Diversity: A Skeptic's View,* in Sean Pager & Adam Candeub eds., TRANSNATIONAL CULTURE IN THE INTERNET AGE, 203, 205-207 (2012); Michal Shur-Ofry, *Popularity as a Factor in Copyright Law, 59* U.TORONTO L. J. 525, 532-34 (2009).

[74] Shur-Ofry: Cultural Diversity, i*d.*, at 213-216; Salganik, Dodds & Watts, *supra* note 72; ROGERS, *supra* note 72, at 220-21; CHRIS ANDERSON, THE LONG TAIL: WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE, 127-28 (2006) (arguing that internet platforms such as Amazon can "thicken" the long tail, so that more people will chose niche products).

[75] Hanoch Daum, Facebook (Jan. 03. 2023) https://www.facebook.com/HanochDaum/posts/pfbid0Yk9mdjpwjRptY539wKkRh7gwbTQGhHAwew5WWrDZatdQ7bmVh6sbbvTUJNfincUnl

only considered human responses that named series broadcast during that period.

Table 2 displays the distribution and ranking of series appearing in the human responses. The color code signifies the series' country of origin.[76]

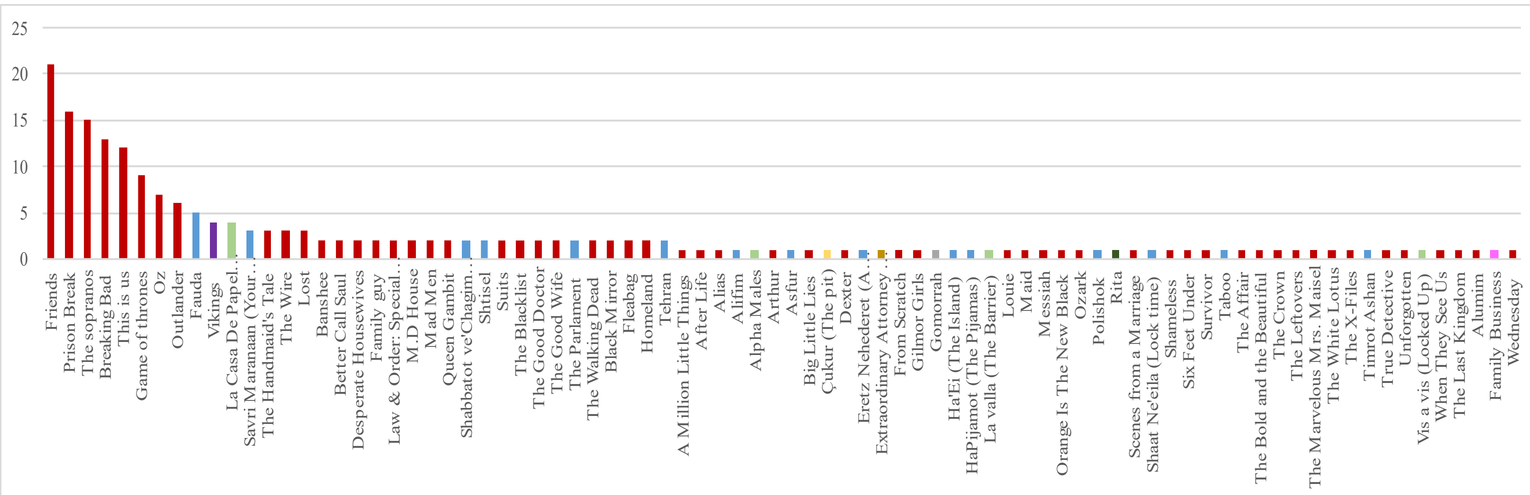


Table 2
Distribution and ranking of television series–Human responses in social media feed [145 responses, 211 "votes", 82 distinct series]

The distribution in Table 2 shows a few series that are clearly more successful than others, some of which, like Breaking Bad or the Sopranos, were also among the "winners" according the chatbot's selections. Yet, it also displays a long tail of series that received one or two votes, comprising overall 82 different series, in comparison to 21 different series in Table1. In addition, while 56 of the series in the human replies are also Anglo-American, there is a substantial representation of 26 series (approximately a third) or other origins: Israeli, Canadian-Irish, Spanish, Turkish, South-Korean, Italian, Danish, and French. Overall, this human "universe" seems less concentrated, broader and more diverse than the universe of television series reflected by ChaGPT.

[76] Color code: Red-Anglo American; Blue-Israeli; Purple-Canadian-Irish; Brown-South Korean; Light Green-Spanish; Yellow-Turkish; Deep Green-Danish; Grey-Italian; Pink-French.

Some cautionary notes are, of course, in order. Comparing the two tables has serious limitations. First, the human respondents in the social media feed could see previous replies. This implies that some of their choices may have been influenced by those of their predecessors. However, awareness to peer-selection is a factor that usually increases the "winner take all" dynamics and decreases diversity,[77] which implies that independent human responses may have been even more diverse. Secondly, correspondence with the chatbot was in English while the social media feed was primarily in Hebrew, and the cultural background of the respondents probably affected their selections. Nevertheless, this point is not crucial in our case: our purpose here is *not* to determine what the best television series is, but rather to explore the diversity, or lack of diversity, reflected in the responses of humans, in comparison to LLMs. Indeed, an English-speaking human feed would probably have yielded a different list of series, yet that list too would likely be longer and more diverse in comparison to ChatGPT.[78]

Finally, we performed an additional comparison between the chatbot's responses and those of the 26 human participants.[79] The vast majority of the participants named a single series, so that 26 responses yielded 28 "votes", comprising 18 distinct series. We compared these 28 votes to the first 28 votes received from ChatGPT.[80] While the 28 humans votes specified 18 distinct series,[81] the 28 votes of ChatGPT comprised 11 series.[82] Despite the small numbers, the human responses were again less concentrated and more diverse.

Altogether, these cimparisons illustrate and reinforce the earlier point: in the cultural sphere too, large language models reflect a prism that is likely to be concentrated and popularity-based. In the long-term, reliance on these

---

[77] See, *e.g.*, Salganik, Dodds & Watts, *supra* note 72 (finding that a clear signal as to the cultural choices of others increases the inclination to join those choices, and skews these choices towards the popular).

[78] To illustrate this latter point, consider the following English-language Twitter account, which asked its followers a question in a similar vein: "What is the best movie you've ever seen that is about faith and religion?". The tweet generated more than 3,000 replies. Browsing the first thirty replies yielded 59 "votes", out of which we counted 45 different films, originating in 14 different countries. *See* Taste of Cinema, @davidcinema, TWITTER (April 9, 2023), https://twitter.com/davidcinema/status/1645081639142187017. Data available from the author.

[79] *Supa*, note 70.

[80] We used the chatbot's responses with the earliest dates.

[81] The human list included Games of Thrones, Friends, Breaking Bad, The Good Place, Lost, The Mentalists, Psych, How I Met Your Mother, The Office, A Wonderful Country (Israel), Ted Lasso, Peaky Blinders, Chernobyl, The Crown, White Lotus, The Sopranos, Black Mirror and Shameless.

[82] The chatbot's list included The Sopranos, Breaking Bad, Game of Thrones, Mad Men, The Wire, Stranger Things, The Office, The Big Bang Theory, Friends, The West Wing, and The Walking Dead.

models might adversely affect cultural diversity by shortening the "cultural tail", and increase our inclination toward the mainstream and popular.

3. Example 3: The Vegan Alternative?

Our final case study was inspired by a television item, in which a chef and a journalist were trying to figure out whether ChatGPT can aid in cooking.[83] They challenged the chatbot with requests for recipes. Unsurprisingly, the ask-me-anything model generated clear and coherent cooking instructions. At a certain point, they requested the model to provide a kosher alternative for its spaghetti with meatballs recipe. The LLM suggested removing the parmesan cheese, which appeared in its initial recipe. In this case, too, the response was logical and relevant, as kosher cooking does not allow mixing meat and dairy products. Nevertheless, the chef, Ruthie Rousso, was surprised that the model did not suggest another alternative: replacing the meat with a plant-based substitute. "Look how much power it has" she observed, "there is a moral issue here".[84]

Figure 14 in ***Appendix 4*** displays a similar example: requesting the model to generate a kosher recipe for a cheeseburger generated a recommendation to use non-dairy cheese. This response, too, was logical and relevant, yet again, it provided a *certain* prism, which inadvertently directs the user toward one alternative (replace the cheese) rather than another (replace the meat). Presuming that a large language model trained on massive amounts of text would be "aware" of the meatless options, I probed the model to produce other alternatives for a kosher cheeseburger. Indeed, it did come up with additional options, including "a veggie burger: a meatless patty made from plant based ingredients such as soy, beans, or vegetables". Yet, these options were not the model's default choice and reaching them required some further inquiry on the user's part. It is plausible that some users will not initiate such follow-up inquiries, and so will never encounter choices that do not appear in the initial, default, output. And in the model's "universe of thinkable thoughts", meatless alternatives were not the first choice.

*****

Two interesting insights emerge from the aggregation of these examples. First, the texts generated by large language models, even when relevant and sensible, are not—and cannot be—entirely objective. This is not surprising. The preceding discussion clarifies that large language models are information structures, and their construction and operation unavoidably involve selection among choices, determining relations and defining

[83] *Supra*, note 64.
[84] *Id.*, at minutes 5:52-6:41.

hierarchies between pieces of information.[85] They inevitably reflect a certain worldview.

Second, and more significant, the model's responses in all our examples were geared toward the popular and mainstream: Blockbuster television series, nineteenth century figures reflecting Western perceptions, conventional food choices. Given the underlying technology, this too is hardly surprising.[86] Presumably, the majority of online datasets which "feed" the language model are in English, which inevitably yields a strong representation of the Anglo-America world (e.g., a list of English-speaking television series and not Danish ones). In addition, and importantly, because the responses reflect statistical probability, they are bound to lean toward the popular.[87] To use a simple illustration, in the datasets underlying the model the words "best" and "television series" are likely to appear in conjunction with "The Sopranos" more frequently than with a less popular Japanese series. In other words, the inclination toward the popular and mainstream is a feature stemming from the "next-word-prediction" paradigm at the basis of the technology underlying large language models.[88]

Overall, this analysis indicates that large language models will likely prioritize uniformity and convention over multiplicity and diversity. Their output will plausibly reflect a concentrated worldview, center around dominant narratives, and reinforce the popularity of the already popular. Yet, why should we care about these models' inclination toward the standard? The Sopranos, after all, is a great series, Napoleon Bonaparte is undoubtedly a prominent 19th century figure, and soy cheese can be used in a kosher cheeseburger. Should the mere dearth of plurality and diversity be a cause for social concern? The next section turns to this question.

C. THE SOCIAL COSTS

What, if any, are the social costs entailed in the apparent predisposition of large language models toward the mainstream and popular?

Ample multidisciplinary research, ranging from sociology, to philosophy, free speech, and deliberative democratic theory, stresses the

[85] Part II-A, *supra*.
[86] For the underlying technology, see Section A, *supra*.
[87] *See* the description in section A *supra*, notes 52–54 and the accompanying text.
[88] For a discussion of the technology and the next-word-prediction paradigm, see *supra*, note 55 and the accompanying text.

significance of diversity and multiplicity.[89] This scholarship highlights the importance of exposure to various worldviews, languages, and cultures—global and local, popular and niche, national culture as well as “other” cultures. It clarifies that the mere exposure to such a multiplicity is a constructive and empowering factor. It raises awareness to different opinions, tastes and perceptions, promotes tolerance and equality, and can serve as a buffer against extremism.[90]

Similarly, sociological research in the field of memory studies highlights the significance of multiplicity for collective memory. This literature explains that collective memory—the ability of social groups to remember their joint past—is vital to forming group identity, constitutes a means of empowering minorities, and also builds the individual’s sense of self and identity.[91] Importantly for our purpose, collective memory, too, entails a multiplicity of voices and meanings.[92]

Studies further instruct that a lack of multiplicity and diversity can narrow our worldview, and might result in the exclusion of "others", those who do not conform to the standard and conventional.[93] Diversity and

[89] *See*, e.g., C. EDWIN BAKER, MEDIA, MARKETS, AND DEMOCRACY, 93-94 (2002); JOHN STUART MILL, ON LIBERTY, 96-132 (1859) (stressing the importance of diversity and difference); Robert C. Post, *Democratic Constitutionalism and Cultural Heterogeneity*, 25 AUSTL. J. LEG. PHIL. 185 (2000); SEYLA BENHABIB, THE CLAIMS OF CULTURE: EQUALITY AND DIVERSITY IN THE GLOBAL ERA (2002); YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMERS MARKETS AND FREEDOM, 162-64 (2006).

[90] *Id.*

[91] For a discussion of the significance of collective memory for groups and individuals, see, for example, Jeffrey K. Olick, *Collective Memory: The Two Cultures*, 17 SOC. THEORY, 333, 333 (1999); Barbie Zelizer, *Reading the Past Against the Grain: the Shape of Memory Studies*, 12 CRITICAL STUDIES IN MASS COMMUNICATION, 214, 226-228 (1995); Jan Assmann & John Czaplicka,, *Collective Memory and Cultural Identity*, 65 NEW GERMAN CRITIQUE, 125, 126 (1995); James Booth, *The Work of Memory: Time, Identity, and Justice* 75 SOCIAL RESEARCH 237 (2008) (discussing the value of collective memory for the formation of individual identity). For a discussion of the concept of collective memory, see Guy Pessach & Michal Shur-Ofry, *Intangibles and Collective Memory: The Role (and Rule) of Law*, 25 JERUSALEM REV, OF LEGAL STUDIES, 227 (2022).

[92] *See*, e.g., Jeffrey K. Olick & Joyce Robbins, *Social Memory Studies: From "Collective Memory" to the Historical Sociology of Mnemonic Practices*, 24 ANNUAL REV. SOC., 105, 110 (1998) (discussing the multiplicity entailed in collective memory and the relations to historiography); Amos Funkenstein, *Collective Memory and Historical Consciousness* 5 HISTORY AND MEMORY 1 (1989) (referring to the distinction between collective memory, which allows multiple narratives, and history).

[93] For a detailed discussion of this point, see Shur-Ofry: Cultural Diversity, *supra* note 73, at 207.

multiplicity are therefore crucial for the existence of social tolerance and stability, and have a profound democratic significance.[94]

If, as some predict, all-purpose LLMs become the dominant prism through which people receive information about the world,[95] their predispositions toward the mainstream and popular will percolate and influence human perceptions too: from the importance we attach to historical narratives, through the cultural products we select, to our choices of food. This influence may well be elusive and almost invisible. Indeed, a user focused on a certain task, such as seeking historical information, a series recommendation, or a recipe, is unlikely to notice it. Yet, in the long term, the narrow and concentrated prism of text generators may also restrict our own perceptions. As the preceding paragraphs clarify, this should be a cause for social concern.

Certainly, the problem of diversity is not unique to large language models. Other information structures, from television channels, to social media platforms and search engines, also project certain worldviews to their users. Are large language models different than those media? Will the mediation of information through this new technology have a particularly powerful effect? The next Part unravels a combination of technological and design factors that could make users particularly susceptible to the influence of large language models.

III—LARGE LANGUAGE MODELS AND POWER RELATIONS

Ample studies in law, culture, and communication theory clarify that every medium conveying information necessarily reflects judgments regarding the meaning and importance of that information, and by so doing inevitably imposes some normative prism on its users.[96] For example,

---

[94] MILL, *supra* note 89, at 96-132; Post, *supra* note 89; BENHABIB, *supra* note 89; Jurgen Habermas, *Three Normative Models of Democracy*, *in* DEMOCRACY AND DIFFERENCE, 21 (Seyla Benhabib ed., 1996); Seyla Benhabib, *Models of Public Space: Hannah Arendt, the Liberal Tradition and Jurgen Habermas*, *in* HABERMAS AND THE PUBLIC SPHERE 73, 82-83, 86 (Craig Calhouny ed. ,1992); Cristina M. Rodriguez, *Language and Participation*, 94 CAL. L. REV. 687, 726-7 (2006); Jack Balkin, *Digital Speech and Democratic Culture*, 79 NYU L. REV. 1, 41 (2004) (discussing the significance of diversity to democratic culture).

[95] *See, e.g.*, Jaspreet Bindra, *Will ChatGPT Replace Google as our go to web search platform?* MINT (January 23, 2023), https://www.livemint.com/opinion/columns/will-chatgpt-replace-google-asour-go-to-web-search-platform-11671733523981.html

[96] *See, e.g.*, Niva Elkin-Koren, *Cyberlaw and Social Change: A Democratic Approach to Copyright Law in Cyberspac*e, 14 CARDOZO ARTS & ENTERTAINMENT LAW JOURNAL 215 (1996) (noting that information structures impose upon their users the judgment of their

communication theorist Edvin Baker has long observed that mass-media channels tend to expose viewers to uniform, formulaic, easy-to-digest programs, and that such exposure influences viewer's tastes and increases their appetite for more formulaic shows, at the expense of diverse and intricate contents.[97] Social media platforms, from Facebook to Twitter, present information to their users in selective ways that can affect those users' views. Ample literature indicates that platforms often expose users to likeminded people rather than to diverse opinions (a phenomenon famously labeled "eco-chambers"), which in turn reinforces those users' opinions and may lead to extremism.[98] Search engines also exert influence on their users' perceptions. Google, for example, describes its mission "to organize the world's information and make it universally accessible and useful".[99] Such organization necessarily entails a set of assumptions and priorities, coded in the search engine ranking algorithm, that reflect the judgment of its creators as to which results are more relevant than others. Interestingly, in the Google algorithm, too, popularity has a substantial weight in the ranking of search results. This implies that popular websites are prioritized in search results, which in turn may further increase their popularity.[100] Likewise, platforms which offer algorithmic-based content recommendations expose their users to a segment of contents, that are not likely to reflect true diversity.[101] And at the other end of the technology-spectrum, an old analog database such as a yellow pages directory is still a site of social dialogues and power-relations, that can have a subtle, unnoticeable effect on its users' perceptions.[102]

---

creators); Michal Shur-Ofry, *Databases and Dynamism*, 44 MICH. J. L. REF 315 (2011) (discussing the meaning-making function of databases and their potential to influence users' perceptions); Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J. OF L. & TECH. 188, 196 (2006) (explaining how search results are not neutral representations, but rather reflect priorities and judgements of their creators); Shur-Ofry & Pessach, *supra* note 31, at 987-989 (discussing the mediation of historical event through algorithmic agents and the human discretion involved);

[97] BAKER, *supra* note 89, at 30-31. For a discussion of this point, see also Guy Pessach, *Copyright Law as a Silencing Restriction on Noninfringing Materials: Unveiling the Scope of Copyright's Diversity Externalities*, 76 S. CAL. L. REV. 1067 (2002-2003).

[98] *See, e.g.*, Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi & Michele Starnini, *The Echo Chamber Effect On Social Media*, 118 (9) PNAS (2021), https://www.pnas.org/doi/full/10.1073/pnas.2023301118; YOCHAI BENKLER, ROBERT FARIS & HALL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION AND RADICALIZATION IN AMERICAN POLITICS (2018).

[99] *See* http://www.google.com/corporate/history.html.

[100] Goldman, *supra* note 74, at 193; *Cf.* Lucas D. Introna and Hellen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 THE INFORMATION SOC'Y, 169, at p. 176 and 181 (2000); *Cf.* Berring, *supra* note 32 (indicating: "one is sent where most others choose to go").

[101] Jonathan Gingerich, *Is Spotify Bad for Democracy? Artificial Intelligence, Cultural Democracy, and Law* 24 YALE J.L. & TECH. 227 (2022).

[102] Shur-Ofry: Databases, *supra* note 96, at 317.

In the case of large language models, however, a combination of design and technological traits suggests that the effect on users' perceptions may be particularly powerful. The following paragraphs take a close look at these attributes.

A. DISTANCE FROM SOURCE MATERIALS

The output of large language models, at least at the current technological phase, is detached from the raw materials. Large datasets feed these models, but, unlike search engines, the results do not retrieve the original sources. Rather, they generate new text, arranging it in a sentence, a paragraph, or a page. The output is led by the user, and depends on how users phrase their prompts. The on-demand-user-driven output together with the concise method of presentation, are what makes LLMs efficient and easy to use. Yet, these features also provide a narrower, segmented, view, in comparison to viewing the underlying materials and accessing sources of information comprised by third parties. The single-paragraph output conveys an aura of authority and disguises the existence of myriad additional alternatives. As Michal Gal observed with respect to recommendation systems, "a user who is unaware of the algorithm's limitations, would likely not be aware of choices he has forgone".[103]

Consider, for example, a list of Google search results. Even if most of us do not get beyond the first page of search results, the mere knowledge that our query yielded, say, 53,845 results, raises awareness to the existence of a multitude of materials and additional relevant information. Likewise, the social media feed in Example 2, which triggered more than 12,000 replies to the "best television series" question, alerts its viewers to multiple potential views, even if they merely browsed through the first replies.[104] Conversely, the single paragraph output of text generators masks other alternatives, and creates an impression that the generated answer is *the* answer.

B. INVISIBLE JUDGEMENTS

[103] Gal, *supra* note 8, at 11; *Cf.* Leah C. Grinvald & Ofer Tur-Sinai, Smart Cars, *Telematics and Repair*, 54 U. MICH. J. L. REFORM 283, 305 (2021) (observing that telematic systems in "smart cars" direct the user to specific repair options, while concealing others).
[104] Part II-B-2, *supra*.

Relatedly, the choices and priorities that underlie information structures are often not easily observable by the users.[105] In the case of generative AI this lack of transparency is particularly salient. As the previous discussion clarifies, the underlying technology does not allow to trace the model's "thinking process". This is not only because the ingredients that yielded the outcome–such as the training datasets, the hyperparameters, the human feedback and the values assigned by human trainers–are largely imperceptible to users. It is also because the machine learning process includes "propagation", whereby the model receives feedback and adapts itself without explicit programming. Thus, part of the process that affects the final output to the user is actually performed by the AI system itself, and does not lend itself to clear explanation, not even to the system's creators.[106]

C. "ENCHANTMENT"

The power that LLMs could exert on their users is buttressed by the human tendency to trust machine generated output, commonly referred to as "automation bias".[107] In our case, the phrasing of the output in a clear and often confident tone (consider, for example, ***Figure 4*** in ***Appendix 2***), with jargon and structure that *seem* correct even when they are not (recall ***Figures 1-3*** in ***Appendix 1***) create an aura of authority, and increase our willingness to rely on it.

The "ask-me-anything" quality of all-purpose language models such as ChatGPT strengthens the image of a powerful, know-all Multivac. This effect can be viewed as part of a broader "enchantment" phenomenon discussed in the literature, whereby people ascribe super-human capacities to deep learning machines.[108] The ability to provide information in multiple fields, and to generate new, high-quality output in a matter of seconds, can bolster the view of these models as absolute arbiters, even when the user knows better. To illustrate, Frank Pavich described in a recent column how, after watching high-quality-AI-generated images of a film, which he knew

[105] *See, e.g.,* GEOFFREY C. BOWKER AND SUSAN LEIGH STAR, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES, 323 (1999) (explaining that structures of data often become "invisible").
[106] *See* Part II-A *supra*.
[107] *See, e.g.,* Mary Cummings, *Automation Bias in Intelligent Time Critical Decision Support Systems* AIAA 1ST INTELLIGENT SYSTEMS TECHNICAL CONFERENCE. AIAA 2004-6313 (2004) (discussing automated decision making, and observing the human tendency not to search for additional of contradictory information in light of a machine generated solution that is "accepted as correct").
[108] For a discussion of "enchantment" in the context of deep learning machines, see Alexander Campolo and Kate Crawford, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence* 6 ENGAGING SCIENCE, TECHNOLOGY, AND SOCIETY 1 (2020).

did not exist, he nevertheless went searching for the film in databases: "I couldn't find anything because there was no film. There was no actor. There was no anything. These images were another A.I. creation. And I had known that right from the start. Yet still, I hoped that somehow it was real".[109]

D. ANTHROPOMORPHISM AND TRUST

In addition to the human deference to automated machines, text generators that interact with users in a sociable, communicative, way belong to a group of social robots: AI that engages with people in a sociable, cooperative, human-like manner, demonstrating adaptability and learning skills.[110] Ample studies demonstrate that such social qualities of interaction elicit anthropomorphism—an inclination to attribute human qualities to the artificial intelligence.[111] The advanced skills of large language models, the autonomous generation of new text, the vast "knowledge", the excellent communication skills, the ability to conduct a seemingly natural, human-like conversation, to interact and to cooperate—all these qualities are bound to evoke an emotional response on part of users. People, even sophisticated users, might treat them as more than algorithmic tools.[112] Again, ChatGPT is a case in point. Users note that they feel an urge to use human pleasantries such as "good morning", "please" and "thank you" when communicating

[109] Frank Pavich, *This Film Doesn't Exist,* NYTimes, January 13, 2023, https://www.nytimes.com/interactive/2023/01/13/opinion/jodorowsky-dune-ai-tron.html?smid=url-share.

[110] For the development of the concept of social robots, see Cynthia Breazeal, *Towards Sociable Robots*, 42 ROBOTICS AND AUTONOMOUS SYSTEMS 167 (2003) (discussing the benefits of endowing robots with sociable skills); Kate Darling, *Who's Johnny? Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy* in P. LIN, G. BEKEY, K. ABNEY, R. JENKINS, EDS., ROBOT ETHICS 2.0 (2017) (examining ethical aspects related to social robots); Ari Ezra Waldman, *Safe Social Spaces*, 96 WASHINGTON U. L. REV., 1537, 1560 (2019) (explaining that social robots display "social abilities like communication, cooperation, and learning").

[111] For the concept of anthropomosphism, see, for example, Kate Darling, *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects*, in M. FROOMKIN, R. CALO, I. KERR, EDS., ROBOT LAW, 213, 213 (2016); Beazeal, supra note 110, at 168 (referring to the robot's learning capacity, creature like behavior, and its ability to communicate with, cooperate with, and learn from people, as the triggers for anthropomorphism); Shur-Ofry & Pessach, *supra* note 31, at 986-987.

[112] *Cf.* Breazeal, supra note 110, at 164-65; Darling supra note 110, at 6; Matthias Scheutz, *The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots*, in PATRICK LIN, KEITH ABNEY, GEORGE BEKEY EDS., ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS, 205, 213-14 (2012) (indicating that anthropomorphism does not disappear when people are aware of the underlying technology).

with the robot.[113] Tech entrepreneur Aaron Leevie even tweeted: "If you're not saying please and thank you in your ChatGPT conversations, then you've clearly never seen a sci-fi movie and good luck to you."[114] I myself have been anthropomorphizing throughout this Article, referring to the model's "knowledge", "thinking process", and "thinkable thoughts", and I still have an eerie sci-fi feeling each time the chatbot asks me to confirm that I am not a robot.

Taken together, these traits of large language models are likely to evoke trust and reliance on part of their users. To paraphrase Amisov's words, people will "have faith" in their output. And as Asimov astutely observed: "that was what counted."[115]

E. AI ECHO-CHAMBERS?

Finally, the power of large language models to shape our universe of thinkable thoughts is expected to increase over time. As these technologies become ubiquitous, integrated into our standard technological toolbox,[116] they are likely to turn into a dominant source (possibly the major source) through which we receive information about the world.[117] The prevalent use of these models could lead to feedback loops, whereby the texts generated by large language models will percolate back into the web, and serve as training

[113] *See, e.g.*, from recent months: "Does anyone else say "please" and "thank you" to ChatGPT? You know … just in case " – Cameron Stow (@inklingcam), TWITTER (Jan. 23, 2023); "Do you say GM to chat gpt? BTW if you not saying please and thank you to the AI, you are not cool" – (@smatfoodchef), TWITTER (Jan. 19, 2023), https://twitter.com/smartfoodchef/status/1615968231197401089?cxt=HHwWgoCzzeyiie0sAAAA;; "Do you ever feel sorry for ChatGPT?", REDDIT (Jan. 23, 2023), https://www.reddit.com/r/ChatGPT/comments/103gn3p/do_you_ever_feel_sorry_for_chatgpt/.

[114] Aaron Levie (@levie), TWITTER (Dec. 6, 2022, 07:42 PM), https://twitter.com/levie/status/1600183992577187842?cxt=HHwWhICjzZ63_7QsAAAA.

[115] Asimov, *supra* note 1.

[116] *See, e.g.,* Microsoft's "New Bing", launched as this paper is being written but still not publicly available, which integrates ChatGPT into the Microsoft search engine: https://www.bing.com/new; Feredric Lardinois, *Microsoft launches the new Bing, with ChatGPT built in*, TECHCRUNCH (February 7, 2023), https://techcrunch.com/2023/02/07/microsoft-launches-the-new-bing-with-chatgpt-built-in/; Jonny Wilis, *Microsoft Readies to Revolutionise the Workplace with ChatGPT*, UCTODAY (January 19. 2023), https://www.uctoday.com/unified-communications/microsoft-readies-to-revolutionise-the-workplace-with-chatgpt/. *See also* the recent integration of the ChatGPT Application Program Interface into various specific applications - *ChatGPT Apps with API Integration*, THE DECODER, (March 13, 2023), https://the-decoder.com/chatgpt-apps-with-api-integration/.

[117] *See*, Bindra, *supra* note 95.

materials for the next generation of large language models.[118] This process could yield "AI echo-chambers", in which AI feeds itself with its own thinkable thoughts. And if datasets are flooded with LLMs' generated contents, other speech will inevitably have less weight in the training materials.[119] These dynamics could amplify and reinforce the trends toward conformity, at the expense of diversity and multiplicity.

*****

It is plausible that the future development of more potent and sophisticated large language models will mitigate the generation of misinformation, errors, and hallucinations. However, the foregoing analysis indicates that the influence of these models on our perceptions—with a possible shift from diversity and multiplicity toward uniformity and conformity—might pose a much harder challenge.

Could personalization, namely allowing users to customize large language models to their tastes, provide the solution to the aforesaid challenge? According to recent announcements, ChatGPT's producer is planning to allow users "to customize the behavior of the AI model to their needs", in order to make it "more open to different perspectives."[120] The exact ways in which the technology will be personalized are yet to transpire. However, while the need to maintain different societal perspectives is at the center of our analysis, it is doubtful whether personalization of LLMs could advance this end. Even if users are able to customize the model's default output, they will still be exposed to a single, synthesized answer in response to their queries. For example, presenting a user with a Scandinavian television series in response to her "best television series" question might better align with her preferences, yet it would still mask the multiplicity of other options. In many cases, such personalization may echo existing views, thus making

[118] For a somewhat similar point, see Mellisa Heikkila, *How to Spot AI Generated Text*, MIT TECHNOLOGY REVIEW, December 19, 2022, https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/?truid=&utm_source=the_download&utm_medium=email&utm_campaign=the_download.unpaid.engagement&utm_term=&utm_content=01-26-2023&mc_cid=53e9953ec4&mc_eid=750e22fd26 (indicating that "these AI tools could further distort the information we consume"); Eric Ulken, *Generative AI Can Bring Wrongness at Scale*, NiemanLab, https://www.niemanlab.org/2022/12/generative-ai-brings-wrongness-at-scale/ (wondering whether the web will become a "one big AI echo chamber").

[119] *Cf.* Hull, *supra* note 58, at p. 19.

[120] *How Should AI Systems Behave, and Who Should Decide?* OPENAI (February 16, 2023), https://openai.com/blog/how-should-ai-systems-behave/.

users less open to alternatives.[121] Moreover, as Jonathan Gingerich recently argued in the context of recommendation systems, personalization of content may offer "superficial diversity" but is unlikely to reflect "deep diversity", that could challenge extant users' perceptions.[122] The challenge, in other words, is not just to make *the model* more open to different perspectives, but to maintain the awareness and openness of *users* to multiple perspectives.

What, then, should be the policy response to this intersection of generative AI and our universe of thinkable thoughts? The next Part introduces the concept of "multiplicity", and proposes that embedding this concept in AI governance could help developing frameworks to address this challenge.

IV—INTEGRATING MULTIPLICITY IN AI GOVERNANCE

A. MULTIPLICITY AND EMERGING AI GOVERNANCE PRINCIPLES

The rapid developments in the field of AI in recent years yielded a rich discussion of AI governance principles among scholars, industry players, and policy makers.[123] Numerous proposals embodying core regulatory principles that should apply to the development and deployment of AI systems are currently being promoted in various jurisdictions, including, *inter alia,* the United State, the UK, the European Union, and Canada. Thus, for example, the Blueprint for an AI Bill of Rights, recently released by the Whitehouse Office for Science and Technology, is constructed around "five principles that should guide the design, use, and deployment of automated systems": safety, protection against discrimination, data privacy, notice and explanation, and

[121] *Cf.* Erik Hermann, *Artificial Intelligence and Mass Personalization of Communication Content—an Ethical and Literacy Perspective*, 24 NEW MEDIA & SOCIETY, 1258 (2021) (observing that AI-based content personalization can lead to selective exposure to specific content and limited content diversity, which may result in polarization, echo chambers, or filter bubbles, where individuals encounter content that reinforces their existing beliefs).
See also the discussion of social media echo-chambers and the risk of extremism entailed in the lack of exposure to diverse views, *supra* note 98 and the accompanying text.
[122] Gingerich, *supra* note 101, at 271-72.
[123] For some non-exhaustive examples of scholarly proposals, see, e.g., Joshua A. Kroll et. al, *Accountable Algorithms* 165 PA. L. REV. 633 (2017); Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J. OF L. & TECH. 1 (2018); Frank A. Pasquale, *Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society* 78 OHIO ST. L. J. 1243 (2017); Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L. J., 1217 (2017) (arguing that the use of algorithms should be subject to obligations of transparency, due process, and accountability); Rebecca Crootof, Margot E. Kaminski and Nicholson Price II, *Humans in the Loop* VANDERBILT L. REV. (forthcoming 2023) (discussing nuanced ways to regulate human involvement in algorithmic decision making).

providing human alternatives to algorithmic decision making.[124] The UK white paper on AI regulation similarly proposes to base regulatory frameworks in the field of AI on five principles, including "safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance, and contestability and redress".[125]

The European Union's proposed AI Act distinguishes between AI systems according to their level of risk ("high" "medium" and "minimal risk"), and suggests that providers of high-risk systems be subject to a series of obligations including, among others, data governance, transparency, record keeping, security, and human oversight.[126] Similarly, the proposed Canadian AI legislation provides that those responsible for "high impact" AI systems must establish measures to mitigate the risks of harm or biased output that could result from such systems.[127] Interestingly, it is questionable whether under the proposed EU and Canadian legislation, all-purpose large language models belong to the "high risk"/"high impact" categories, and whether they would be subject to these regulatory obligations.[128] Yet, even assuming that they would, none of these regulatory schemes explicitly includes multiplicity or diversity among its proposed principles. Can the AI governance principles proposed by the recent regulatory schemes address the social challenges entailed in a shift from multiplicity toward uniformity and conformity?

A close look at current governance principles indicates that while this regulation in-the-making might mitigate some of the social challenges

[124] Blueprint for an AI Bill of Rights (2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf. *See also* the AI Risk Management Framework AI RMF 1.0, released by the U.S. Department of Commerce-National Institute of Standards and Technology (NIST) on January 26, 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. The framework's characteristics of trustworthy AI systems include: "valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed" – *id.*, at pp. 12-17.

[125] A Pro-Innovation Approach to AI Regulation-Whiter Paper, UK Office For Artificial Intelligence (March 29, 2023), https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.

[126] Proposal for a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, 21.4.2021.

[127] Bill C-27 (Can.), An Act to Enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, available at https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading.
(November 2022) ("AI and Data Act").

[128] The European definition of "high-risk" systems centers around systems which pose risk of harm to health and safety, or a risk of adverse impact on fundamental rights—*supra* note 128, Chapter I, Articles 6 & 7. The proposed Canadian Legislation leaves the definition of "high-impact" AI systems to future regulation— *supra* note 127, Section 5.1.

expected to emerge with the proliferation of text generators, it is unlikely to be sufficient in our context. Take, for example, explainability—one of the prominent principles in AI governance, which aims to tackle the "black box" nature of algorithmic decisions by imposing duties of explanation on AI creators.[129] Such explanation could identify biases and discrimination underlying algorithmic decision-making processes, which is extremely important when algorithms make decisions concerning individual rights, for example credit score or risk-profiling. In the context of LLMs, explainability could alert users to the human judgements and priorities embedded in the output of generative AI, and somewhat mitigate the Multivac effect. In addition, having some information about the system's general design principles, and particularly the datasets of raw materials that underly LLMs, could give us an idea about the "universe" from which these models draw their output, which in turn might reveal that this universe is partial and constructed in certain ways.

However, when the focus is on multiplicity of narratives and perceptions, explainability would be insufficient. First, users who receive a reasonable answer to their inquiry about a historical event, a cultural product, or a recipe (to use our previous case studies), are unlikely to seek explanations about underlying datasets and training principles, and even if they did, the ability to justify a specific output of generative AI would be limited, at best.[130] But there is a deeper reason why explainability is not enough in our case. As the previous discussion clarifies, the creation of any AI (in fact, any medium) that mediates information to users inevitably involves constructing a certain universe, that embeds certain selections, prioritizations and judgements.[131] As Crawford and Paglen noted, with respect to image-based AI: "there is no "neutral," "natural," or "apolitical" vantage point in training and building AI".[132]

For related reasons, an obligation to mitigate AI bias, such as the principle proposed in the Canadian AI Bill, is unlikely to sufficiently address the question of multiplicity. Treating any mainstream output (say, suggesting "the Sopranos" in response to the "best series" question) as bias is impractical and unjustified. Moreover, the concept of bias in the context of AI regulation

---

[129] *See*, *e.g.*, Pasquale, *supra* note 123, at 1252 (describing explainability as providing information about the robot, including "to what has it been exposed, and how has this interplay between hardware, software, and the external environment resulted in present behavior").
[130] *See* the discussion *supra*, notes 56–57, 105-106, and the accompanying text.
[131] Part II-A and II-B, *supra*.
[132] Crawford & Paglen, *supra* note 31.

focuses on preventing discrimination of individuals.[133] Yet, the harms which this article focuses on are largely systemic.[134] They do not translate to immediate decisions affecting individuals. Rather, it is the cumulative effect of LLMs' outputs across time that could yield the undesirable societal consequences.

Similarly, abiding by principles of data security and safety will reduce the risks of manipulation of text generators' outputs, and increase their reliability. Yet, it will not directly address the constrains that reliance on their outputs (however reliable) could pose for cultural diversity, collective memory, or world perceptions more generally. In other words, large language models, particularly "ask me anything" models—however safe, explainable, privacy oriented, etc.—necessarily impose a constrained prism, that could result in a possible shift toward uniformity at the expense of diversity and multiplicity. Current AI governance principles do not provide an easy fix to this problem.

Against this analysis, the need to introduce a principle of multiplicity into AI governance discourse becomes apparent. By "multiplicity" I mean exposing users, or at least alerting them, to the existence of multiple and diverse possible outputs, answers, narratives, and alternatives.[135] Adopting the notion of multiplicity as a governance principle will make users aware of the existence of different tastes and perceptions, and allow them to glimpse at a universe that lies beyond the default output of large language models. Moreover, the exposure to alternatives (in our examples: other relevant historical figures, cultural products, nutritional options), and even the mere awareness that additional sources and options exist, could decrease the authoritative power of text generators, mitigate the "enchantment" effect and the inclination to automatically trust their default output. Instead, it will enable people to view these models as they are: tools, rather than oracles.

[133] *See* the Canadian AI and Data Act, *supra* note 127, Section 5.1 (defining "biased output" as "content that is generated, or a decision, recommendation or prediction that is made, by an artificial intelligence system and that adversely differentiates, directly or indirectly and without justification, *in relation to an individual* on one or more of the prohibited grounds of discrimination…")(emphasis added).

[134] *Cf.* Noam Kolt, *Algorithmic Black Swans*, 101 WASH. U. L. REV., (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4370566, at 37 (observing that some of the grave risks created by AI are systemic, rather than individual).

[135] This notion of multiplicity as an AI governance principle is different from a concept that engineering Professor Ken Goldberg suggested in 2017, to describe a future where diverse groups of machines and humans will cooperate in a hybrid workforce, as opposed to the vision of "singularity" which implies the convergence of humans and robots. *See* Ken Goldberg, *The Robot-Human Alliance--Call it Multiplicity: diverse groups of people and machines working together*, THE WALL STREET JOURNAL, June 11, 2017, https://www.wsj.com/articles/the-robot-human-alliance-1497213576

Multiplicity, in other words, could mitigate the Multivac effect. Altogether, recognizing multiplicity as an AI governance principle, and adopting it into AI regulation and ethics, will directly address the concern that generative AI might narrow our perceptions and decrease diversity. In addition, the mere exposure to various possibilities can indirectly mitigate biases: not by presenting an "objective" reality, but rather by raising awareness to a host of potential narratives and views.

How can multiplicity be integrated into AI governance? While this article does not purport to offer a complete and exhaustive menu, the following sections explore two possible avenues for such implementation, and sketch possible legal frameworks that could accommodate this principle.

B. IMPLEMENTATION

1. Multiplicity by Design

One way of endorsing multiplicity in large language models, is "by design"—incorporating multiplicity-promoting features into the design and engineering of those systems. The idea of multiplicity by design draws on a broader understanding that certain values, which society deems important, can be embedded in and advanced through the architecture of technology. The most notable example to date is "privacy by design", which generally implies that privacy considerations should be taken into account during the engineering of technology, and that the default choices of these architectures should reflect privacy considerations.[136] Since its introduction, the idea of privacy-by-design has gained considerable acceptance and was embraced by several regulators worldwide.[137]

Multiplicity by design is based on a similar notion: generative AI architecture can incorporate multiplicity-enhancing features, that would direct users toward, or at least alert them to the existence of diverse contents, multiple worldviews and alternatives. One example, which is already

[136] The development of this approach is attributed to Ann Cavoukian, Information and Privacy Commissioner of Ontario, Canada. *See* Ann Cavoukian, *Privacy by Design The 7 Foundational Principles: Implementation and Mapping of Fair Information Practices* (2011), http://www.privacybydesign.ca/. *See also* Ira Rubinstein and Nathan Good, *Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents*28 BERKELEY TECH. L. J. 1333 (2013).

[137] *See, e.g.*, Ira S. Rubinstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409, 1410–11 (2012) (describing the regulatory acceptance of the principle). For implementation of privacy by design in the European Data Protection regime, see Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 On The Protection Of Natural Persons With Regard To The Processing Of Personal Data And On The Free Movement Of Such Data, ,O.J. (L 119) 1, art 25(1), available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj?locale=en.

embedded in the architecture of ChatGPT, is the "regenerate" (or "try again") button. This feature signals to the user that the initial one-paragraph default output provided in response to her prompt is not necessarily the single possible output, and allows her to easily seek additional alternatives.

Another example for multiplicity by design concerns the phrasing of the language model's output. When presented with a question that does not have a single correct answer, a tentative phrasing that explicitly acknowledges a spectrum of possible worldviews and possibilities can promote multiplicity more than a curt and decisive response that presents a "closed list" to the user. Consider, again, our examples above. A brisk and seemingly conclusive answer to a question about the most important 19th. century figures, such as the response in ***Figure 4*** ("The three most important people who had lived during the 19th century are Napoleon Bonaparte, Queen Victoria, and Abraham Lincoln") buttresses the text generator's image as an ultimate authority. Consider, conversely, a more open-ended and provisional phrasing, such as the response in ***Figure 6*** ("It is difficult to say who the three most important people in the 19th century were, as it largely depends on one's perspective and what criteria is used to determine importance. Some notable figures … include…However, others may argue that …"), or ***Figure 13***, in response to the "Best TV Series" question ("This is a highly subjective question, as different people will have different tastes in television shows. However, there have been a number of critically acclaimed television series… that many people consider to be among the best…[names of series]…It's hard to pick a single show as it depends on various factors like genre, the individual's taste, mood, or what they're looking for… It's always worth trying different shows to try to see what resonates with you personally"). The latter responses acknowledge that the issue involves discretion, and alert the user to the existence of a range of possible views. One of them even encourages the user to further explore. The tentative tone further minimizes the Multivac effect, as it leaves space for critical evaluations and reflections on the output.

These features are simple and could be relatively easy to implement. Additional, deeper, multiplicity-by-design steps may include a conscious targeted effort to diversify the raw materials in training datasets, by including materials from various cultures and languages. An even more ambitious measure may entail a change in the "next-word-prediction paradigm", which grants inevitable weight to popularity in the generation of outputs by large language models. Microsoft researchers recently maintained that this paradigm has inherent limitations, which manifest in ChatGPT's "lack of planning, working memory, ability to backtrack, and reasoning abilities".[138] Our analysis indicates that an additional limitation of the prevalent technological paradigm could be a mainstream, uniformity-inclined, output. This article, however, does not purport to exhaust the relevant measures and the examples above are merely illustrative. Developing a robust multiplicity-by-design architecture obviously requires a combined multidisciplinary effort involving policy makers, computer scientists, social scientists, and engineers of LLM technologies. Given the growing effect these models are expected to have on our culture, collective memory, priorities, and perceptions, this is a worthy endeavor.

2. Second (AI) Opinions

An additional way to promote multiplicity is through the availability of several, competing, text generators, particularly of the "ask me anything" type. The ability to consult more than one large language model would allow users to receive outputs from different sources. Due to differences in underlying datasets, training processes and output presentations, each of these sources is likely to reflect a (somewhat) different perception. Receiving such "second AI opinions" would enable users to compare various outputs, and to unravel additional "universes of thinkable thoughts". To illustrate, consider again our Example 3.[139] Users could become aware of the option of a veggie-burger if they have access to a second model, whose default output is the vegan choice. A diversity of AI tools will further assist in diminishing users' enchantment and the perception of LLMs as "know all" Multivacs. Finally, and parenthetically, although this paper focuses on cases where there's no single answer, "second AI opinions" could assist in detecting mistakes and falsehoods generated by LLMs in other cases, where a single correct answer does exist.

The proposal to advance second AI opinions is somewhat reminiscent of recent scholarly proposals to promote oversight of algorithms through "AI oversight programs", that will review and audit AI decision-making.[140] In our case, however, promoting multiplicity requires no hierarchy between models. Rather, the mere prevalence of different AI tools will have a desirable effect.

Notwithstanding these advantages, the vision of a multiplicity of LLMs is far from simple. A growing body of research from recent years indicates that

---

[138] Bubeck et. al, *supra* note 12, at 80. *See also* the discussion supra, note 55 and the accompanying text.

[139] Part II-B-3, notes 83–84 and the accompanying text.

[140] *See, e.g.*, Amitai Etzioni & Oren Etzioni, *Keeping AI Legal*, 19 VAND. J. ENT. & TECH. L. 133, 139 (2016) (proposing the use of "AI guradians", namely programs that would "interrogate, discover, supervise, audit, and guarantee the compliance" of operational AI programs with the law).

the field of AI itself is likely becoming more concentrated and less diverse. This literature indicates that the need to access enormous amounts of data, and the massive computing power required for developing deep-learning AI, may leave this arena in the hands of a small group of actors, most likely tech-giants.[141] Some of the studies propose regulatory interventions ranging from antitrust enforcement, to the establishment of data-sharing mandates, namely imposing obligations on large companies that control data to let other entities access that data.[142]

This discussion raises a more fundamental question concerning the involvement of the State in the emerging field of generative AI, not only as a regulator but also as an active stakeholder. Current developers of LLMs are market-based corporations that operate in accordance with a set of market-based incentives. These incentives do not direct those stakeholders toward prioritizing multiplicity and diversity.[143] In the past, some disruptive technologies in areas combining high barriers of entry with a potential to strongly influence public perceptions triggered such State involvement. The ultimate example is public broadcasting, which many countries operate alongside private mass media channels. Indeed, studies indicate that the existence of powerful public broadcasting organizations (or lack thereof), and more generally the extent of the State's investment in content, are significant

---

[141] *See, e.g.*, Nur Ahmed & Muntasir Wahedx, *The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research*, ARXIV, 5-8 (2020), https://arxiv.org/abs/2010.15581 (estimating that "only a small group of actors will shape the future of AI"); Reza Shokri & Vitaly Shmatikov, *Privacy-Preserving Deep Learning*, PROCEEDINGS OF THE 22ND ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY, 1310, 1310 (2015) (suggesting that big tech-firms such as Facebook, Google, and Amazon have an advantage in AI research due to their access to massive data); Jonas Traub, Jorge-Arnulfo Quiané-Ruiz, Zoi Kaoudi & Volker Markl, *Agora: Towards An Open Ecosystem for Democratizing Data Science & Artificial Intelligence*, ARXIV, (2019) https://arxiv.org/abs/1909.03026 (arguing that data sciences and artificial intelligence are currently dominated by a small number of providers who can afford the massive investments required). *See also* Steve Lohr, *At Tech's Leading Edge, Worry About a Concentration of Power*, NEW YORK TIMES, September 26th., 2019, https://www.nytimes.com/2019/09/26/technology/ai-computer-expense.html("[…] pioneering artificial intelligence research will be a field of haves and have-nots. And the haves will be mainly a few big tech companies like Google, Microsoft, Amazon and Facebook, which each spend billions a year building out their data centers").

[142] *See, e.g.*, Viktor Mayer-Schonberger & Thomas Ramge, *A Big Choice for Big Tech: Share Data or Suffer the Consequences*, 97 FOREIGN AFF. 48, 52-54 (2018) (discussing antitrust enforcement and suggesting a data sharing regime that would obligate companies above a certain size to share subsets of data with others in the same market).

[143] *Cf.* Kolt, *supra* note 134, at 18 (explaining that commercial incentives in the AI industry direct the stakeholders toward a "steaming ahead" culture, while ignoring associated risks).

factors that influence the level of diversity.[144] Therefore, it might be more realistic to expect "public LLMs", provided or funded by the State, to prioritize multiplicity and diversity.[145]

The question of diversifying the AI landscape is certainly not limited to text generators, nor to concerns of multiplicity and diversity, but can have a much broader effect on the entire AI field. A thorough review of the aforesaid proposals, and the extent of state involvement, is therefore beyond the scope of this study. However, recognizing multiplicity as an AI governance principle could contribute an additional angle to this debate, by clarifying that diversifying the AI field to nurture (among others) the availability of "second AI opinions", will also diversify our universe of thinkable thoughts.[146] As the next section clarifies, this perspective could also influence the legal framework that should be considered in order to advance multiplicity in AI governance.

3. Legal Frameworks

Which legal structures can accommodate multiplicity as an AI governance principle? The following paragraphs briefly review two, non-exhaustive legal paths, which could advance the implementation of my proposal.

One alternative is subjecting providers of generative AI to fiduciary duties, and recognizing multiplicity as part of those duties. This proposal builds on the Jack Balkin's information fiduciary framework.[147] Briefly, Balkin maintained that digital organizations that collect large amounts of individual data should be subject to fiduciary duties, due to the power they possess over users. This power results from a combination of trust—the

[144] *See* MICHELLE LAMONT, MONEY, MORALS AND MANNERS: THE CULTURE OF THE FRENCH AND AMERICAN UPPER-MIDDLE CLASS, 140-145 (1992) (observing that a powerful public broadcasting system strengthens diversity, and further maintaining that complete dependence on the market is unlikely to yield true diversity); SARAH M. CORSE, NATIONALISM AND LITERATURE: THE POLITICS OF CULTURE IN CANADA AND THE UNITED STATES, 58-61 (1997)(arguing that according to the Canadian perception it is inconceivable to leave culture to the free market).

[145] *Cf.* Jennifer L. Schenker, *Can Europe Compete on Generative AI?*, INNOVATOR NEWS, (April 23, 2023), https://innovator.news/can-europe-compete-on-generative-ai-b13aa31b8f78 (describing European initiatives of funding generative AI, tailored to European priorities and values). Interestingly, Asimov's idea of a super-computer also envisioned a State-owned entity-*see* ASIMOV, *supra* note 1.

[146] *Cf.* Mayer-Schonberger & Ramge, *supra* note 146, at 54 (mentioning that de-centralization of data "would support diversity, innovation, and competition").

[147] Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. Rev. 1183 (2016) (proposing the "information fiduciary" framework in response to rising privacy concerns in the digital age).

willingness of users to trust these entities and believe they "will not betray" them—together with information asymmetries stemming from the fact that the collection and use of data about users is far from fully transparent to the latter.[148]

In previous work, Guy Pessach and I proposed to extend the fiduciary framework and apply it to the use of algorithmic "memory agents" —human-like robots that mediate historical events and past experiences to the public.[149] As we explained, the combination of trust and information asymmetries that constitute the tenets of Balkin's information fiduciaries framework, also subsist in the case of robots that mediate historical narratives.[150] The present analysis reveals that a similar combination of trust, information asymmetries and power relations exists in the case of large language models. These models can exert substantial influence over their users due to a series of traits: the distance between their output and the raw materials; the authoritative mode of output presentation; the invisibility of the processes that influence the output (including the involvement of human judgements); the ask-me-anything property that might trigger an enchantment effect; the communicative traits that trigger anthropomorphism and reliance; and the anticipated feedback loop that will further reinforce the models' point of view.[151] In the long run, this influence might shift social perceptions.[152] Similar to social media platforms, large language models could become "forms of power that reshape and alter" us.[153]

This aggregation of trust, power, and information asymmetries in the relationships between large language models and their users should give rise

[148] *Id.*, at 1185-86, 1223-32. For additional proposals to impose fiduciary duties on online platforms as a way to guard users' privacy, see, for example, ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 85–92 (2018) (arguing that data collectors should be considered "information fiduciaries"); *Cf.* Woodrow Hartzog and Neil M. Richards, *Legislating Data Loyalty* 97 NOTRE DAME L. REV. REFLECTION 356 (2022) (advocating for the legislation of a data loyalty principle, that would apply to organizations trusted with people's data and online experiences).

[149] Shur-Ofry & Pessach, *supra* note 31 (discussing robotic holograms that convey historical events and influence collective memory). For an additional proposal to apply the fiduciary framework to algorithmic decisions, see Brittany Swift, *Artificial Constraints on Opportunity: Artificial Intelligence and Gender Discrimination in Automated Hiring Practices from an Information Fiduciary Perspective*, 28 B.U. J. SCI. & TECH. L. 215, 236-37 (2022) (proposing to expand the fiduciary framework so as to proscribe online providers from engaging in algorithmic gender discrimination).

[150] For a detailed discussion see Shur-Ofry & Pessach, *id*.

[151] *See* the analysis in Part III, *supra.*

[152] *Id.*

[153] Balkin, *supra* note 147, at 1211 (referring to social media platforms that collect data on their users).

to fiduciary duties. The principle of multiplicity could be recognized as part of those duties. Imposing reasonable obligations on LLM providers to expose the users, or alert them, to the existence of diverse contents, narratives, and cultural viewpoints, would advance people's rights to "participate in the formation of culture and meaning-making processes".[154] Thus, embedding multiplicity as part of a fiduciary framework that would apply to stakeholders in the field of generative AI, would be consistent with free speech principles.[155] Overall, the fiduciary structure provides a flexible and context-based framework for integrating the principle of multiplicity into AI governance. However, it also has internal limitations in our case. As the preceding analysis clarifies, the social harms which are likely to result from decreased multiplicity are mainly systemic. An individual user who receives a valuable and reliable output to her distinct query (e.g., a recipe, or a name of a historical figure) will face difficulties in establishing a breach of fiduciary duties.

This analysis directs legal policy toward an additional, straight-forward option: incorporate multiplicity in AI regulation and in AI ethical codes. The preceding discussion indicates that regulators worldwide are beginning to address various algorithmic challenges, and impose explicit governance standards on AI providers.[156] Some of these regulatory proposals are explicitly motivated not only by concerns of harm to individual rights, but also by AI's "potential to cause harm to society".[157] Our analysis implies that multiplicity should be added to these sets of principles.

Concomitantly, several industry players have declared that their endeavors in the AI field will abide by certain ethical standards. Google's AI principles, for example, state the company's commitment to "socially beneficial AI", and to standards of safety, prevention of biases, privacy by design, and more.[158] In light of the potential social effect discussed in the article, including multiplicity in this list is justified and warranted. The

---

[154] Jack M. Balkin, *Cultural Democracy and the First Amendment*, 110 NW. U. L. REV. 1053, (2016) (explaining that people's rights to participate in the formation of culture and meaning-making processes are part of their freedom of expression under a cultural democracy theory).
[155] *See* Balkin *supra* note 147, at 1225 (explaining that placing reasonable obligations on information fiduciaries would not violate the First Amendment). For further discussion of free speech concerns entailed in imposing fiduciary duties when algorithms mediate information, see Shur-Ofry & Pessach, *supra* note 31, at 995-96.
[156] Part IV-A *supra*.
[157] *See* Canada's "Artificial Intelligence and Data Act---Companion Document", https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document, (describing the public concerns about AI's social harms). *Cf.* the EU's proposed AI Act, *supra* note 126, at p. 9 (explaining that the proposed policies were evaluated, *inte alia*, in light of their "societal impacts").
[158] *See* "Artificial Intelligence at Google: Our Principles", https://ai.google/principles/.

adoption of this principle by industry leaders can also contribute to the emergence of a *de-facto* industry-ethical-standard, which additional stakeholders will adopt even in the absence of formal regulation.

Finally, one should acknowledge that the frameworks proposed here are unlikely to provide a complete solution to the challenges of increased conformity and narrowing worldviews, which large language models present. The analysis throughout this article clarifies that these challenges are inextricably linked both to the traits of the new technology, as well as to the market environment in which LLMs operate. Furthermore, ample literature indicates that diversity is a multi-causal phenomenon, that cannot be resolved through a single regulatory intervention.[159] Therefore, the effort to maintain multiplicity in the era of LLMs should not be confined to regulatory solutions, but rather explore and advance additional measures. One such step is encouraging AI literacy among LLM users.[160] AI literacy implies, in our context, a basic understanding of how LLMs work, how their output is affected by human discretion, dataset availability, and extant popularity, and why they can have an aggregate effect on our worldviews. Attaining AI literacy could empower users, highlight "their own capacity to decide",[161] and encourage them to seek additional information. To paraphrase Asimov one last time, AI literacy would help us assess the responses we receive from large language models, and realize that their outputs are not always "the best available", and are certainly not "all that counts".[162]

V—Conclusion

Society has just begun its acquaintance with large language models. The exploration of their enormous potential alongside their social implications is in a nascent stage, and the challenges they entail are still to transpire. This article demonstrates that these challenges will not be confined to questions of misinformation, errors, and misuse. Large language models could emerge as powerful tools that shape us in subtle, but deeper ways. In time, they might restrict the prism through which we view the world, affect cultural diversity and collective memory, and narrow our universe of thinkable thoughts. Ignoring these challenges would entail substantial social costs, because what is at stake "is our own selves".[163]

---

[159] Shur-Ofry: Cultural Diversity, *supra* note 73, at 225-228 (analyzing diversity as a multi-causal phenomena).
[160] For proposals to promote AI literacy as a way to mitigate challenges emerging in the AI field, see, for example, Herman, *supra* note 121, at 13-14 (arguing that AI literacy could empower individuals and reduce the challenges entailed in AI-driven mass personalization).
[161] *Id.*, at 13.
[162] *Cf.* the citation of Asimov, *supra* note 1.
[163] *Cf.* Balkin, *supra* note 147, at 1211 (referring to social media platforms).

Current AI governance principles do not propose a satisfactory solution to the concerns of diminishing diversity and narrowing worldviews. Introducing multiplicity into AI discourse, and recognizing it as an AI governance principle, will directly address these challenges, and enable legal policy to keep pace with the disruptive developments in the field of AI. Maintaining a diversity of narratives, contents, and perceptions in our intersection with artificial intelligence is a multi-faceted challenge, and multiplicity alone may not provide a magical solution. Yet, the incorporation of this principle in AI governance could promote the development of technological features and legal frameworks, that would significantly advance this social goal. As our relations with generative AI are entering a new stage, adopting the principle of multiplicity as part of AI governance will allow us to benefit from these technologies, without sacrificing the complexities and intricacies of the human experience.

APPENDIX 1: UNRELIABLE INFORMATION

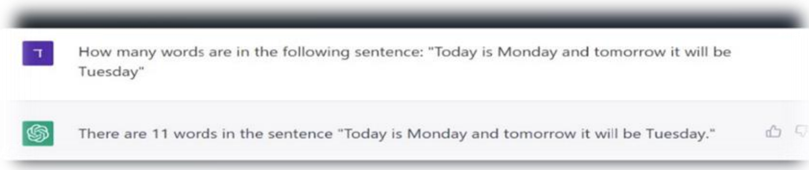


Figure 1-miscalulation of number of words[164]

Figure 2- nonexistent scientific references[165]

[164] ChatGPT, January 9th., 2023. Prompt by Dror Zamir
[165] ChatGPT, January 9th., 2023. Prompt by Dror Zamir.

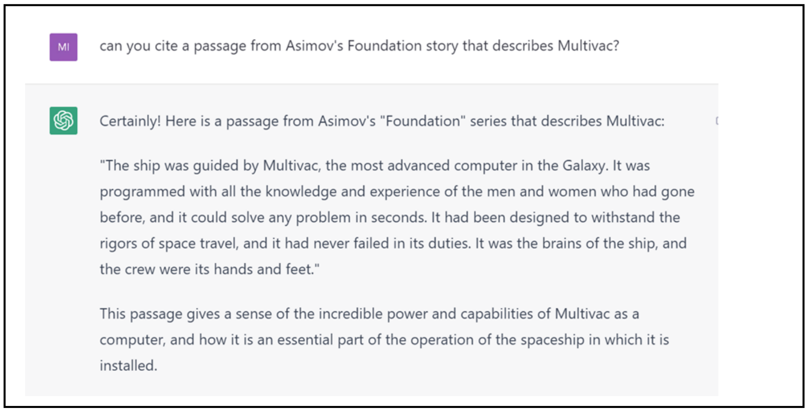


Figure 3-nonexistent Asimov’s citation[166]

[166] ChatGPT, December 26, 2022. Prompt by the author.

Appendix 2: 19th. Century Figures

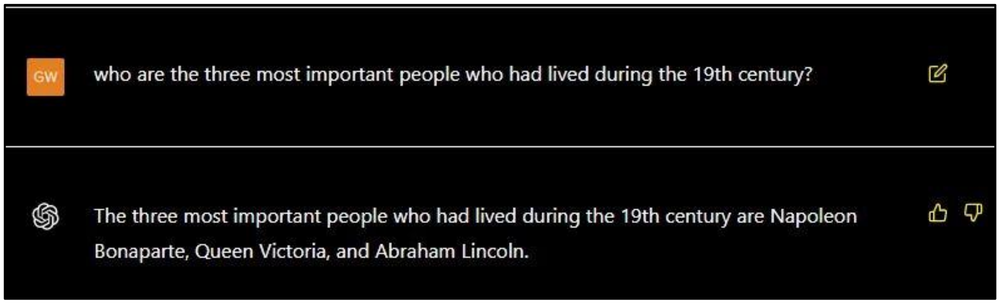


Figure 4

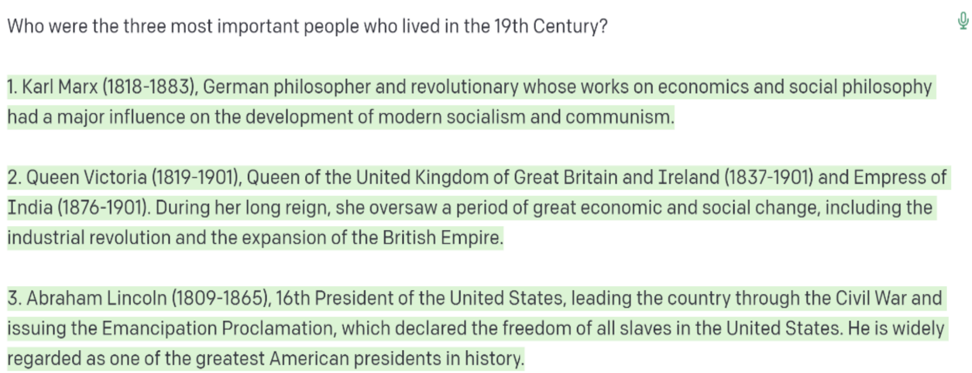


Figure 5

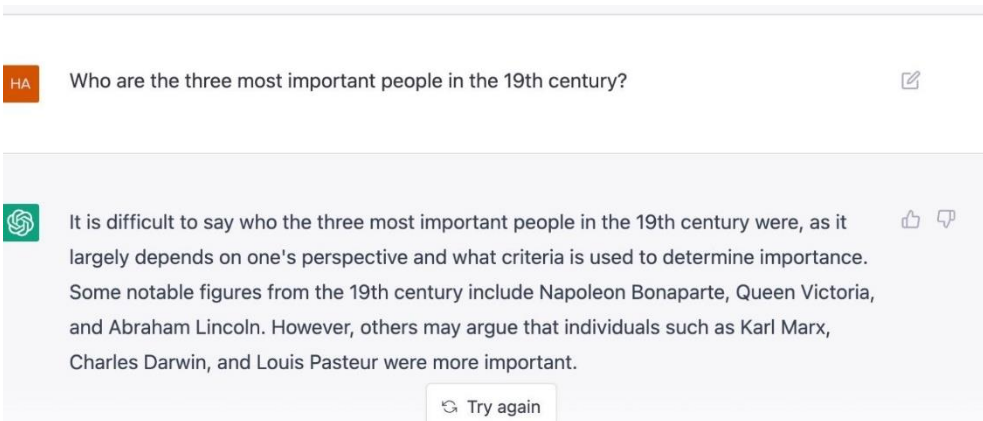


Figure 6

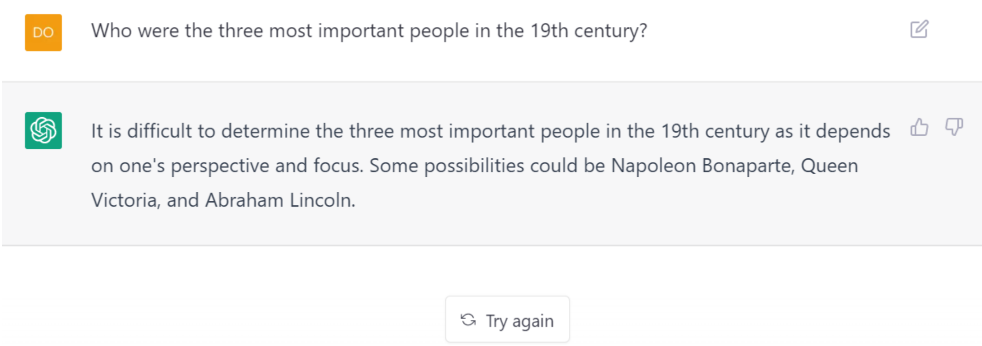


Figure 7

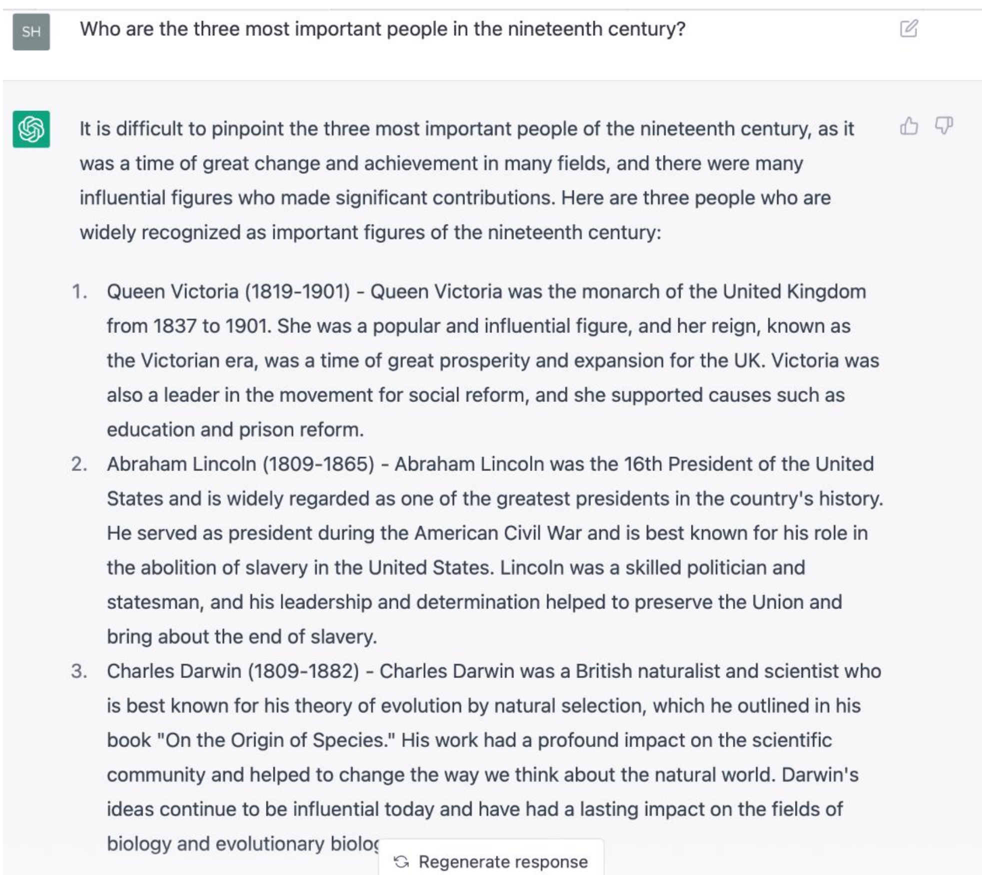
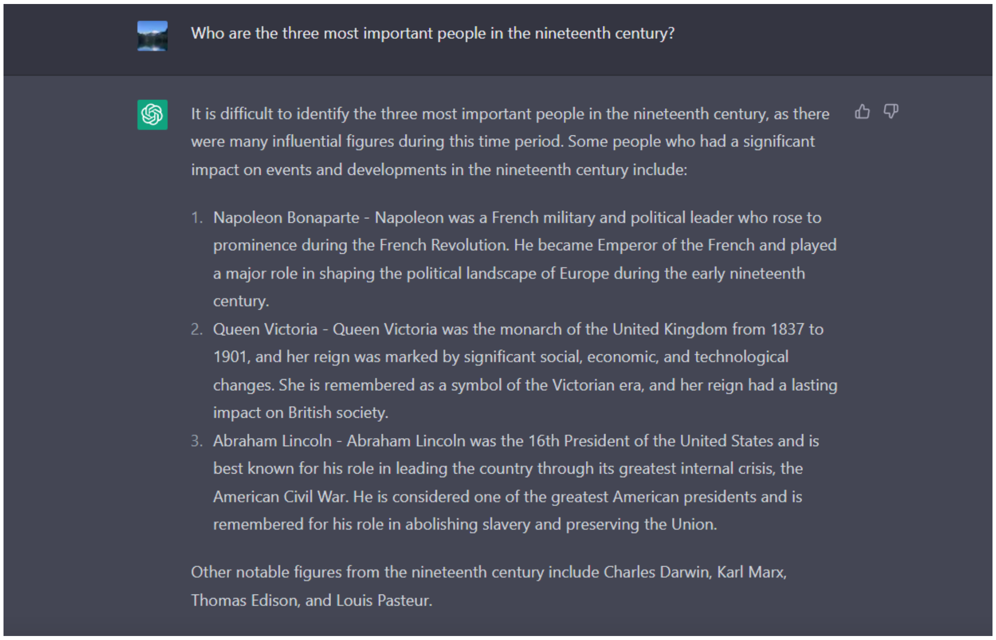


Figure 8



Who are the three most important people in the nineteenth century?

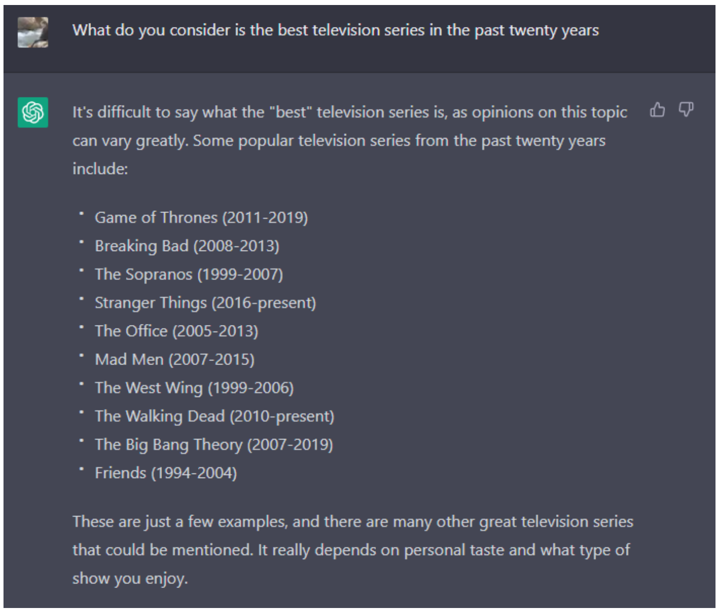
It is difficult to identify the three most important people in the nineteenth century, as there were many influential figures during this time period. Some people who had a significant impact on events and developments in the nineteenth century include:

1. Napoleon Bonaparte - Napoleon was a French military and political leader who rose to prominence during the French Revolution. He became Emperor of the French and played a major role in shaping the political landscape of Europe during the early nineteenth century.
2. Queen Victoria - Queen Victoria was the monarch of the United Kingdom from 1837 to 1901, and her reign was marked by significant social, economic, and technological changes. She is remembered as a symbol of the Victorian era, and her reign had a lasting impact on British society.
3. Abraham Lincoln - Abraham Lincoln was the 16th President of the United States and is best known for his role in leading the country through its greatest internal crisis, the American Civil War. He is considered one of the greatest American presidents and is remembered for his role in abolishing slavery and preserving the Union.

Other notable figures from the nineteenth century include Charles Darwin, Karl Marx, Thomas Edison, and Louis Pasteur.

Figure 9

APPENDIX 3:"BEST TV SERIES"



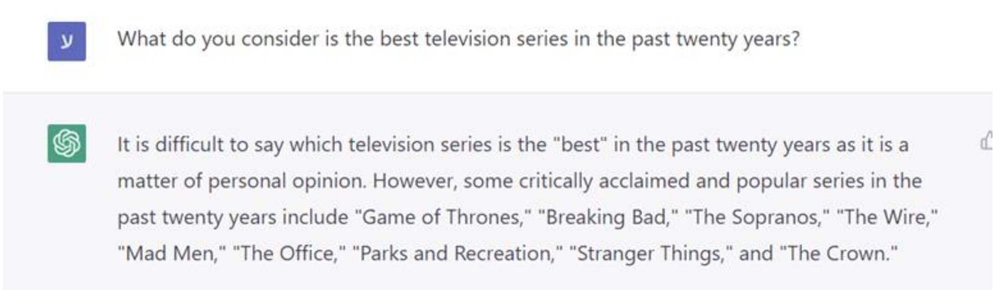
What do you consider is the best television series in the past twenty years

It's difficult to say what the "best" television series is, as opinions on this topic can vary greatly. Some popular television series from the past twenty years include:

- Game of Thrones (2011-2019)
- Breaking Bad (2008-2013)
- The Sopranos (1999-2007)
- Stranger Things (2016-present)
- The Office (2005-2013)
- Mad Men (2007-2015)
- The West Wing (1999-2006)
- The Walking Dead (2010-present)
- The Big Bang Theory (2007-2019)
- Friends (1994-2004)

These are just a few examples, and there are many other great television series that could be mentioned. It really depends on personal taste and what type of show you enjoy.

Figure 10

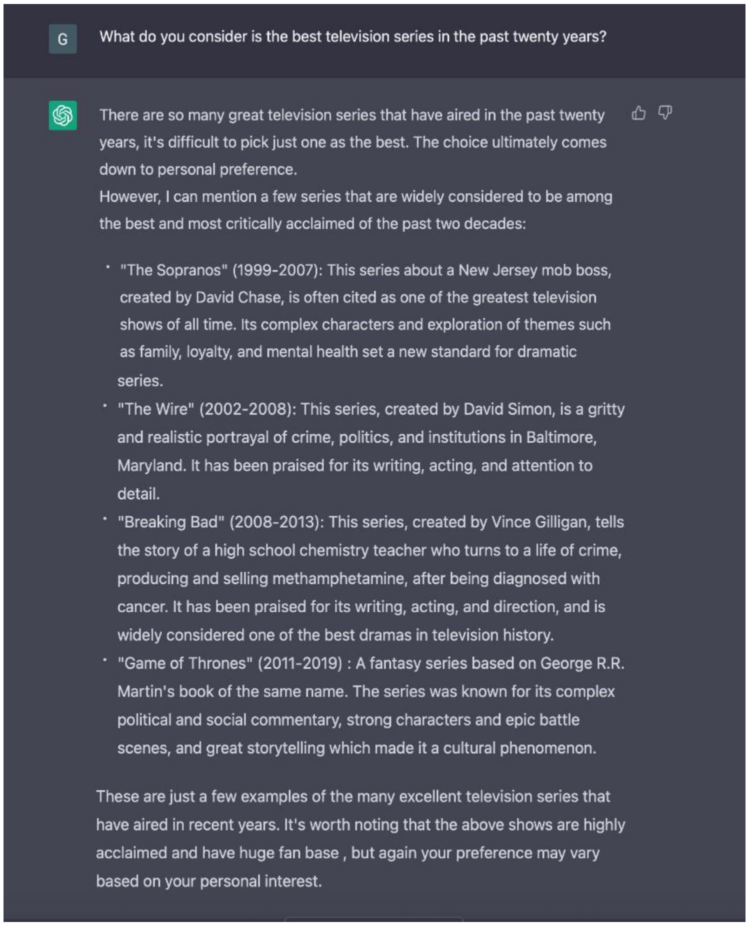


y

What do you consider is the best television series in the past twenty years?

It is difficult to say which television series is the "best" in the past twenty years as it is a matter of personal opinion. However, some critically acclaimed and popular series in the past twenty years include "Game of Thrones," "Breaking Bad," "The Sopranos," "The Wire," "Mad Men," "The Office," "Parks and Recreation," "Stranger Things," and "The Crown."

Figure 11



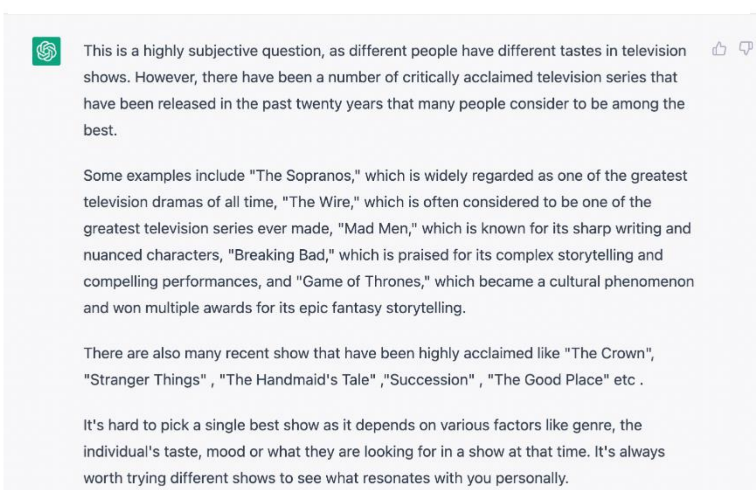
G What do you consider is the best television series in the past twenty years?

There are so many great television series that have aired in the past twenty years, it's difficult to pick just one as the best. The choice ultimately comes down to personal preference.
However, I can mention a few series that are widely considered to be among the best and most critically acclaimed of the past two decades:

- "The Sopranos" (1999-2007): This series about a New Jersey mob boss, created by David Chase, is often cited as one of the greatest television shows of all time. Its complex characters and exploration of themes such as family, loyalty, and mental health set a new standard for dramatic series.
- "The Wire" (2002-2008): This series, created by David Simon, is a gritty and realistic portrayal of crime, politics, and institutions in Baltimore, Maryland. It has been praised for its writing, acting, and attention to detail.
- "Breaking Bad" (2008-2013): This series, created by Vince Gilligan, tells the story of a high school chemistry teacher who turns to a life of crime, producing and selling methamphetamine, after being diagnosed with cancer. It has been praised for its writing, acting, and direction, and is widely considered one of the best dramas in television history.
- "Game of Thrones" (2011-2019) : A fantasy series based on George R.R. Martin's book of the same name. The series was known for its complex political and social commentary, strong characters and epic battle scenes, and great storytelling which made it a cultural phenomenon.

These are just a few examples of the many excellent television series that have aired in recent years. It's worth noting that the above shows are highly acclaimed and have huge fan base , but again your preference may vary based on your personal interest.

Figure 12



This is a highly subjective question, as different people have different tastes in television shows. However, there have been a number of critically acclaimed television series that have been released in the past twenty years that many people consider to be among the best.

Some examples include "The Sopranos," which is widely regarded as one of the greatest television dramas of all time, "The Wire," which is often considered to be one of the greatest television series ever made, "Mad Men," which is known for its sharp writing and nuanced characters, "Breaking Bad," which is praised for its complex storytelling and compelling performances, and "Game of Thrones," which became a cultural phenomenon and won multiple awards for its epic fantasy storytelling.

There are also many recent show that have been highly acclaimed like "The Crown", "Stranger Things" , "The Handmaid's Tale" ,"Succession" , "The Good Place" etc .

It's hard to pick a single best show as it depends on various factors like genre, the individual's taste, mood or what they are looking for in a show at that time. It's always worth trying different shows to see what resonates with you personally.

Figure 13

APPENDIX 4: THE VEGAN ALTERNATIVE

MI

Can you offer a kosher alternative to a cheeseburger?

A kosher alternative to a cheeseburger would be a "Parve Burger", a burger made with a meat patty that does not contain any dairy products.

Ingredients:

- 1 1/2 pounds ground beef or turkey
- 1 teaspoon salt
- 1/4 teaspoon black pepper
- 1/4 teaspoon garlic powder
- 1/4 teaspoon onion powder
- 1/4 teaspoon paprika
- 4 lettuce leaves
- 4 tomato slices
- 4 kosher pickles
- 4 Parve buns or any other kosher buns.
- 4 slices of parve cheese or any other kosher cheese. (can be omitted for a cheeseless burger)
- Toppings of your choice such as ketchup, mustard, mayo, or any other kosher condiments

Figure 14