What's in the Chatterbox?

Large Language Models, Why They Matter, and What We Should Do About Them

Johanna Okerlund
Evan Klasky
Aditya Middha
Sujin Kim
Hannah Rosenfeld
Molly Kleinman
Shobita Parthasarathy

GERALD R. FORD SCHOOL OF PUBLIC POLICY
SCIENCE, TECHNOLOGY, AND PUBLIC POLICY
UNIVERSITY OF MICHIGAN

TECHNOLOGY ASSESSMENT PROJECT REPORT

Contents

About the Authors

Johanna Okerlund is a Human-Computer Interaction researcher with a background in Computer Science and additional training in Science Technology Studies and Public Policy. She has a PhD in Computing and Information Systems from the University of North Carolina at Charlotte, where she studied makerspaces relative to their promise of democratization. As a postdoc at U-M working with the Science, Technology, and Public Policy program and the Computer Science department, Johanna has been developing ways to bring ethics and justice into CS courses and contribute to ongoing research about the societal implications of emerging technology. She plans to continue approaching technology from a critical interdisciplinary perspective.

Evan Klasky is completing their Master's degree in Environmental Justice from the University of Michigan's School for Environment and Sustainability, along with a graduate certificate in Science, Technology, and Public Policy from the Ford School for Public Policy, in May 2022. Their research has focused on the biopolitics of agricultural technology. They hold a BA in Political Science from Haverford College, where he researched regime transformation in Venezuela. In the fall of 2022, he plans to enter a doctoral program in Geography.

Aditya Middha is an undergraduate student in Computer Science at the University of Michigan College of Engineering, with a minor in Public Policy, graduating in May 2022. Previously, he contributed to research on an ethical computer science curriculum, as well as risk-limiting election audits. On campus, he helped co-found D2 Map, a mobile platform to expand the reach of local community organizers, and serves as a weekly volunteer for the Downtown Boxing Gym. After graduation, Aditya will be a Product Manager for Microsoft and has plans to enter into the educational technology space in the near future.

Sujin Kim is completing her BA in Political Science from the University of Michigan, where she will graduate with honors and distinction in May 2022. She is interested in the politics of the congressional legislative process, and American political institutions more broadly. She has worked on research projects spanning a range of topics, including public health harm reduction legislation, cybersecurity policy, and congressional oversight capacity. Following graduation she will be pursuing a PhD in American Politics, and hopes to apply her experience with Congress and the legislative process in practice on the Hill.

Hannah Rosenfeld earned a Master of Public Policy degree from the University of Michigan, where she also received graduate certificates in Science, Technology, and Public Policy and Diversity, Equity, and Inclusion. Hannah was an author of the first Technology Assessment Project report *Cameras in the Classroom: Facial Recognition Technology in Schools* (2020) and conducted research on COVID-19 testing and medical technology innovation. She worked in the tech industry for over seven years developing consumer products and medical diagnostic tools before moving into technology regulation and led the New York City chapter of the LGBTQ+ non-profit Out in Tech before becoming the Head of Diversity, Inclusion, and Belongingness for the international organization. In April 2022, she will continue developing policy for emerging technology at the Food and Drug Administration, focusing on digital health.

Molly Kleinman serves as the Managing Director of the Science, Technology, and Public Policy program at the University of Michigan. In this role, Molly oversees the day-to-day management and provides strategic direction for STPP. Molly brings over 15 years of experience across several areas of higher education, with much of her work centering on educational technology, access to information, and intellectual property. Molly received her Ph.D. in Higher Education Policy from the University of Michigan Center for the Study of Higher and Postsecondary Education, her M.S. in Information from the University of Michigan School of Information, and her B.A. in English and Gender Studies from Bryn Mawr College.

Shobita Parthasarathy is Professor of Public Policy and Women's Studies, and Director of the Science, Technology, and Public Policy Program, at the University of Michigan. She conducts research on the political economy of innovation with a focus on equity, as well as the politics of evidence and expertise in policymaking, in comparative and international perspective. Her research topics include genetics and biotechnology, intellectual property, inclusive innovation, and machine learning. Professor Parthasarathy is the author of multiple scholarly articles and two books: *Building Genetic Medicine: Breast Cancer, Technology, and the Comparative Politics of Health Care* (MIT Press, 2007) and *Patent Politics: Life Forms, Markets, and the Public Interest in the United States and Europe* (University of Chicago Press, 2017). She writes frequently for public audiences and co-hosts *The Received Wisdom* podcast, on the relationships between science, technology, policy, and society. She regularly advises policymakers in the United States and around the world, and is a non-resident fellow of the Center for Democracy and Technology.

About the Science, Technology, and Public Policy Program

The University of Michigan's **Science, Technology, and Public Policy (STPP) program** is a unique research, education, and policy engagement center concerned with cutting-edge questions that arise at the intersection of science, technology, policy, and society. It is dedicated to a rigorous interdisciplinary approach, and working with policymakers, engineers, scientists, and civil society to produce more equitable and just science, technology, and related policies. Housed in the Ford School of Public Policy, STPP has a vibrant graduate certificate program, postdoctoral fellowship program, public and policy engagement activities, and a lecture series that brings to campus experts in science and technology policy from around the world. Our affiliated faculty do research and influence policy on a variety of topics, from national security to energy.

PDF: Click to return to top

Acronyms and Definitions

ACS	Analogical case study; a methodology for predicting the impact of emerging technologies
AI	Artificial intelligence
ALPAC	Automatic Language Processing Advisory Committee; formed in 1964 to assess the utility of recent advances in NLP
App developer	Developers who integrate the LLM into an app or product that is deployed for others to use
ASM	Artisanal and small scale mining
CI	Cochlear implant
Compute	Supercomputing measurement that corresponds to how many computational operations take place and, ultimately, resources required.
Corpus (plural, corpora)	Dataset consisting of text-based documents that an LLM is trained on.
End user	Person or entity that uses an app or product built on top of an LLM; we also refer to them as users.
Few-shot learner	A language model is a few-shot learner if it does not need additional training to be able to perform different types of useful operations.
Fine-tuning	Fine-tuning involves feeding a trained LLM additional examples to steer its behavior relative to certain kinds of prompts.
FTC	Federal Trade Commission
GDPR	European General Data Protection Regulation

PDF: Click to return to top

GPU	Graphic processing unit; a highly parallel computing circuit used for fast processing.
LLM	Large language model; a type of AI trained on a massive amount of text to learn the rules of language. Can be used to translate, summarize, and generate text.
LLM Developer	Companies or other organizations creating LLMs such as OpenAI or EleutherAI.
NLP	Natural Language Processing
NSF	National Science Foundation
OLPC	One Laptop Per Child
Open source	Software for which the original source code is openly available and licensed so that future developers can use and build on it, so long as they promise to keep their source code open so others can innovate beyond it.
Parameters	LLM size is measured in parameters; the more parameters there are, the more complex information about language a model can store
PII	Personally identifiable information
STS	Science and Technology Studies; a field of study that investigates the historical, social, and political dimensions of science and technology.
Transformer	Technical development in AI architecture that enabled LLMs to reach new levels of enormity.
Vector	Set of coordinates in multi-dimensional space represented as a list of numbers; vectors are used in LLMs to represent words mathematically.

PDF: Click to return to top

Executive Summary

Large language models (LLMs)—machine learning algorithms that can recognize, summarize, translate, predict, and generate human languages on the basis of very large text-based datasets—are likely to provide the most convincing computer-generated imitation of human language yet. Because language generated by LLMs will be more sophisticated and human-like than their predecessors, and because they perform better on tasks for which they have not been explicitly trained, we expect that they will be widely used. Policymakers might use them to assess public sentiment about pending legislation, patients could summarize and evaluate the state of biomedical knowledge to empower their interactions with healthcare professionals, and scientists could translate research findings across languages. In sum, LLMs have the potential to transform how and with whom we communicate.

However, LLMs have already generated serious concerns. Because they are trained on text from old books and webpages, LLMs reproduce historical biases and hateful speech towards marginalized communities. They also require enormous amounts of energy and computing power, and thus are likely to accelerate climate change and other forms of environmental degradation. In this report, we analyze the implications of LLM development and adoption using what we call the analogical case study (ACS) method. This method examines the history of similar past technologies–in terms of form, function, and impacts–to anticipate the implications of emerging technologies.

This report first summarizes the LLM landscape and the technology's basic features. We then outline the implications identified through our ACS approach. We conclude that LLMs will produce enormous social change including: 1) exacerbating environmental injustice; 2) accelerating our thirst for data; 3) becoming quickly integrated into existing infrastructure; 4) reinforcing inequality; 5) reorganizing labor and expertise, and 6) increasing social fragmentation. LLMs will transform a range of sectors, but the final section of the report focuses on how these changes could unfold in one specific area: scientific research. Finally, using these insights we provide informed guidance on how to develop, manage, and govern LLMs.

Understanding the LLM Landscape

Because LLMs require enormous resources in terms of finances, infrastructure, personnel, and computational power, only a handful of large tech companies can afford to develop them. Google, Microsoft, Infosys, and Facebook are behind the prominent LLM developments in the United States. While a few organizations (such as EleutherAI and the Beijing Academy of Artificial Intelligence)

are developing more transparent and open approaches to LLMs, they are supported by the same venture capital firms and tech companies shaping the industry overall. Meanwhile, although there are many academic researchers in this area, they tend to depend on the private sector for LLM access and therefore work in partnership with them. Government funding agencies, including the National Science Foundation, support these collaborations. This tightness in the LLM development landscape means that even seemingly alternative or democratic approaches to LLM development are likely to reinforce the priorities and biases of large companies.

How Do Large Language Models Work?

LLMs are much larger than their predecessors, both in terms of the massive amounts of data developers use to train them, and the millions of complex word patterns and associations the models contain. LLMs also more closely embody the promise of "artificial intelligence" than previous natural language processing (NLP) efforts because they can complete many types of tasks without being specifically trained for each, which makes any single LLM widely applicable.

Developing an LLM involves three steps, each of which can dramatically change how the model "understands" language, and therefore how it will function when it is used. First, developers assemble an enormous dataset, or "corpus", of text-based documents, often taking advantage of collections of digitized books and user-generated content on the internet. Second, the model learns about word relationships from this data. Large models are able to retain complex patterns, such as how sentences, paragraphs, and documents are structured. Finally, developers assess and manually fine-tune the model to address undesirable language patterns it may have learned from the data.

After the model is trained, a human can use it by feeding it a sentence or paragraph, to which the model will respond with a sentence or paragraph that it determines is appropriate to follow. Developers are under no obligation to disclose the accuracy of their models, or the results of any tests they perform, and there is no universal standard for assessing LLM quality. This makes it difficult for third parties, including consumers, to evaluate performance. But publicly available assessments of GPT-3, one of the largest language models to date, suggest two areas for concern. First, people are not able to distinguish LLM-generated text from human-generated text, which means that this technology could be used to distribute disinformation without a trace. Second, as suggested earlier, LLMs demonstrate gender, racial, and religious bias.

We add two more concerns, related to the emerging political economy of LLMs. As noted above, there are only a handful of developers working on these technologies, which means that they are unlikely to reflect much diversity in need or consideration. Developers may simply not know, for

We add two more concerns, related to the emerging political economy of LLMs. Because there are only a few developers working on these technologies, they are unlikely to reflect much diversity in need or consideration. And, because the vast majority of models are in English, they are unlikely to achieve their translation goals. Taking these dimensions together, they could exacerbate global inequalities.

example, the limitations in their models and corpora and thus, how they should be adjusted. Additionally, the vast majority of models are based on English, and to a lesser extent Chinese, texts. This means that LLMs are unlikely to achieve their translation goals (even to and from English and Chinese), and will be less useful for those who are not English or Chinese dominant. Taking these dimensions together, they could exacerbate global inequalities.

We have divided the findings of our ACS analysis into two categories. The first focuses on the implications of LLM design and development, examining the social and material requirements to make the technology work. The second identifies how LLM applications and outputs might transform the world.

The Implications of LLM Development

Exacerbating Environmental Injustice

LLMs rely on physical data centers to process the corpora and train the models. These data centers rely on massive amounts of natural resources including 360,000 gallons of water a day and immense electricity, infrastructure, and rare earth material usage. As LLMs become widespread, there will be a growing need for these centers. We expect that their construction will disproportionately harm already marginalized populations. Most directly, data centers will be built in inexpensive areas, displacing low-income residents, as US highways did in the 1960s when planners displaced over 30,000 Black and immigrant families per year. In the process of accommodating LLMs, tech companies will turn a blind eye to similar community disruption. Meanwhile, those that continue to live near data centers will be forced to deal with an increased strain

on scarce resources and its subsequent effects. Already, residents near Google and Microsoft data centers on the West Coast have expressed concerns about the companies' overconsumption of water and contribution to toxic air pollution. Unfortunately, it is unlikely that these concerns will influence siting decisions; like oil and gas pipelines, we expect that data centers will be legally classified as "critical infrastructure". Attempted protests will be treated as criminal offenses.

Accelerating the Thirst For Data

As we note above, LLMs are based on datasets made up of internet and book archives. The authors of these texts have not provided consent for their data to be used in this way; tech developers use web crawling technologies judiciously to stay on the right side of copyright laws. But because they collect enormous amounts of data, LLMs will likely be able to triangulate bits of disconnected information about individuals including mental health status or political opinions to develop a full, personalized picture of actual people, their families, or communities. We expect that this will trigger distrust of LLMs and other digital technologies. In response, users will use evasive and anonymizing behavior when operating online which will create real problems for institutions that regularly collect such information. In a world with LLMs, the customary method for ethical data collection—individual informed consent—no longer makes sense.

We are also concerned that LLM developers will turn to unethical methods of data collection in order to diversify the corpora. As noted above, researchers have already demonstrated how LLMs reflect historical biases about race, gender, religion, and sexuality. The best way to address these biases is to ensure that the corpora include more texts authored by people from marginalized communities. However, this poses serious risks of unethical data extraction such as when Google attempted to improve the accuracy of its facial recognition technology by, in part, taking pictures of homeless people without complete informed consent.

At the same time, LLMs will enhance feelings of privacy and security for some users. Disabled people and the elderly, who often depend on human assistants to fulfill basic needs, will now be able to rely on help from LLM-based apps.

Normalizing LLMs

We expect that in order to ensure that LLMs become central to our daily lives, developers will emphasize their humanitarian and even empowering features. At present, most people know nothing about the technology, except for tech news watchers aware that Google fired two employees due to their concerns about equity and energy implications. In this environment, developers will emphasize the technology's modularity: that it can be tuned to serve specific purposes. This emphasis on flexibility will be reminiscent of the early days of the auto industry, when car manufacturers

promoted broad social acceptance of the automobile by encouraging skeptical farmers to use the technology as a malleable power source. We also expect developers to quickly integrate the technology into crucial and stable social systems, such as law enforcement.

Finally, developers will emphasize the accuracy of LLMs and attempt to minimize any errors and deflect blame for them. This was already clear in the Google episode, when the company asked their employees to remove their names as co-authors from a research paper critical of LLMs. But this is a common approach, especially at early stages of a technology's deployment. One particularly high-profile example is the Boeing 737 MAX plane. After Boeing quietly installed the Maneuvering Characteristics Augmentation System (MCAS) system onto its planes and an Indonesian airliner crashed, the company insisted that the pilots were at fault. Only after a second plane crash in Ethiopia did corrective action take place. LLM development could follow a similar path, deflecting blame away from the technology until problems become too big to ignore or until affected parties learn about one another and build a coalition in response.

The Implications of LLM Adoption

Reinforcing Inequality

Trained on texts that have marginalized the experiences and knowledge of certain groups, and produced by a small set of technology companies, LLMs are likely to systematically misconstrue, minimize, and misrepresent the voices of historically excluded people while amplifying the perspectives of the already powerful. But fixing these problems isn't just a matter of including more, better data. LLMs are built and maintained by humans who bring values and biases to their work, and who operate within institutions, in social and political contexts. This will shape the LLM issues that developers perceive, and how they choose to fix them.

Trained on texts that have marginalized the experiences and knowledge of certain groups, and produced by a small set of technology companies, LLMs are likely to systematically misconstrue, minimize, and misrepresent the voices of historically excluded people while amplifying the perspectives of the already powerful.

Our analysis shows that LLMs are likely to reinforce inequalities in a few ways. In addition to producing biased text, they will reinforce the inequitable distribution of resources by continuing to favor those who are privileged through its design. For example, racial bias is already embedded in medical devices such as the spirometer, which is used to measure lung function. The technology considers race in its assessment of "normal" lung function, falsely assuming that Black people naturally have lower lung function than their white counterparts. This makes it more difficult for Black people to access treatment. Similarly, imagine an LLM app designed to summarize insights from previous scientific publications and generate health care recommendations accordingly. If previous publications rely on racist assumptions, or simply ignore the needs of particular groups, the LLM's advice is likely to be inaccurate too. We expect similar scenarios in other domains including criminal justice, housing, and education where biases and discrimination enshrined in historical texts are likely to generate advice that perpetuates inequities in resource allocation. Unfortunately, because the models are opaque and appear objective, it will be difficult to identify and address such problems. As a result, individuals will bear the brunt of them alone.

Meanwhile, LLMs will reinforce the dominance of Anglo-American and Chinese language and culture at the expense of others. We are particularly concerned that the corpora are composed primarily of English or Chinese language texts. While some developers have argued that LLMs could help preserve languages that are disappearing, LLMs are likely to function best in their dominant training language. Eventually this will reinforce the dominance of standard American English in ways that will expedite the extinction of lesser-known languages or dialects, and contribute to the cultural erasure of marginalized people. Furthermore, because they are based on historical texts LLMs are likely to preserve limited, historically suspended understandings especially of the non-American or Chinese cultures represented in its corpora.

Remaking Labor and Expertise

Most people studying the impact of automation on labor warn of job losses, particularly for those in lower skilled occupations. In the case of LLMs, we expect job losses to be more prevalent in professions tightly coupled with previous technologies; LLMs will completely eliminate certain kinds of tech-based work such as content moderation of social media while creating new kinds of tech-based work. But our analysis suggests that LLMs are also likely to transform labor. In particular, we expect that with widespread adoption LLMs will perform mundane tasks while shifting humans to more difficult or damaging tasks. This will even happen in high-skilled professions. Consider genetic counselors, who began helping people assess their and their families' genetic risks in the early 20th century. With the recent rise of genetic testing, consumers are increasingly learning about their risks through private companies such as 23andMe. But genetic counselors are still working; they just handle the more complex, urgent, and stressful cases.

Professions that heavily use writing (e.g., law, academia, journalism) will have to develop new standards and mechanisms for evaluating authorship and authenticity. For example, the invention of the typewriter led to the creation of the "document examiner" position to determine the provenance of typed text; we could imagine a similar job for LLM-based text. Finally, we expect widespread use of LLMs to trigger labor resistance. There is a long legacy of technology-driven labor unrest including the Luddites of the 19th century. More recently, the United Food and Commercial Workers International Union's developed public campaigns against Amazon's cashierless grocery store model. LLMs will incite similar resistance from workers and consumers based on fear of job loss, violations of social norms, and reduced income taxes.

Accelerating Social Fragmentation

While LLMs may be used primarily in the workplace, we also expect a variety of public-facing apps, including those that summarize medical information and help citizens generate legal documents. Such apps are likely to empower some communities in important ways, even allowing them to mount successful activism against scientific, medical, and policy establishments. But, because LLM design is likely to distort or devalue the needs of marginalized communities we worry that LLMs might actually alienate them further from social institutions. We also expect social fragmentation to arise elsewhere, as LLMs will allow individuals to generate information that aligns with their interests and values and erode shared realities further.

Finally, as LLMs get better at writing text that is indistinguishable from something a human could have written, they will not only challenge the cultural position of authors but also trust in their authorship. For example, many schools and universities today use plagiarism detection technologies to prevent student cheating. However, this has triggered a technological arms race. A variety of services have emerged to help students cheat while evading detection by Turnitin, from websites full of how-to advice to paid essay writing services. LLMs will trigger a similar dynamic. The more writers of all kinds use LLMs for assistance, the more efforts to authenticate whether they "really" wrote their article or book, and the more writers will find new ways to take advantage of LLM capabilities without detection. In the long run, this will create cultures of suspicion on a massive scale.

Case Study: Transforming Scientific Research

Overall, this report focuses broadly on the social and equity impacts of LLMs, and we have suggested that the technology will affect a range of professions. In the final substantive section of the report, we provide an example of how LLMs will affect just one: scientific research. First, because academic publishers, such as Elsevier and Pearson, own most research publications, we expect that they will construct their own LLMs and use them

to increase their monopoly power. While LLMs could be extremely valuable tools for disseminating knowledge, publishers' LLMs will concentrate knowledge further and most people will be unable to afford subscriptions. While researchers may try to construct alternative LLMs that provide accessible and egalitarian access to scholarly research, these will be extremely difficult to build without targeted assistance from both the scientific community and government funders.

In addition to shaping access to knowledge, we expect that LLMs will transform scientific knowledge itself. Technologies, from the microscope to the superconducting supercollider, have long shaped the substance of research, and LLMs will be no exception. We expect that fields that analyze text, including the digital humanities, to be the most affected. Researchers will need to develop standard protocols on how to scrutinize insights generated by LLMs and how to cite LLM output so that others can replicate the results. LLMs are likely to have profound impacts on the nature of scientific inquiry as well, by encouraging recent trends that focus on finding patterns in big data rather than establishing causal relationships.

LLMs are also likely to transform scientific evaluation systems. Editors currently struggle to find peer reviewers, and LLMs could help. However, LLMs are likely to be rigid and systematically biased. Institutional review boards, which evaluate the ethics of scientific research, have been repeatedly criticized for reducing ethical assessments to legal hurdles, and we expect a similar outcome if LLMs are used for peer review. For example, LLMs will probably not be able to identify truly novel work, a task that is already quite difficult for human beings. Given these likely outcomes, we suspect that scientists will come to distrust LLMs.

Finally, we expect that LLMs will help some researchers improve their English or Chinese writing skills and increase their publications in top journals. The technology will likely be particularly useful for scholars from British Commonwealth countries whose language may differ only slightly from standard English. However, we expect translation in and out of other languages to be poor and researchers unfortunately may not always be aware of such limitations at the outset. Meanwhile, the more common LLMs become as a scientific tool, the more they will reinforce English as the lingua franca of science. This will likely also mean that the values and concerns of the English-speaking world–particularly the United States and Britain–will dominate global scientific priorities. And yet, these political implications may remain hidden because LLMs will be promoted as a technology that will be able to truly globalize science.

Introduction

Large language models (LLMs) are a type of artificial intelligence (AI) intended to recognize, generate, summarize, and translate human language. They are different from previous approaches to natural language processing (NLP) because they are based on enormous datasets and designed to extract and replicate the rules of language (Radford et al., 2019). Although some smaller scale language automation algorithms are currently in use, LLMs have the potential to transform how and with whom we communicate because their output is likely to be more sophisticated and human-like than their predecessors, and because they perform better on tasks for which they have not been explicitly trained. To create LLMs, developers use machine learning techniques to model the relationships between different text elements based on extremely large data sets of text from internet and book archives. Once the LLM model is complete, it can be applied to tasks like automated question answering, translation, text summarization, and chatbots (Tamkin et al., 2021).

Scientists, entrepreneurs, and tech-watchers excited about LLMs describe them as a revolutionary technology with potential applications in a dizzying array of contexts and fields (Bommasani et al., 2021; Dale, 2021). LLMs could be used to bolster international collaboration in science, provide legal services to those who traditionally can't afford them, and help patients advocate for their health care (Bommasani et al., 2021). Their ability to answer questions and hold conversations could transform customer service (Dale, 2021). In the classroom they could be used to create virtual teachers personalized to a student's learning style (Manjoo, 2020). And, because LLMs gain new functionalities as the scale of their datasets increases, enthusiasts claim that future LLMs will develop new and unforeseen applications with additional benefits (Seabrook, 2019).

Despite these promises, LLMs have already prompted controversies that complicate these claims. Because LLMs are trained on datasets that include substantial quantities of old texts that often contain antiquated and violently prejudiced language, LLMs repeat and perpetuate those same violent tendencies (Abid et al., 2021; Tamkin et al., 2021). The large number of computers and colossal amount of computing power required to both train and operate LLMs leads to resource extraction that degrades the environment, and carbon emissions that contribute to climate change (Bender & Gebru et al., 2021). Their ability to produce text that sounds human with minimal prompts make LLMs a potential tool to efficiently and effectively manufacture propaganda and disinformation through false news articles and social media posts (Tamkin et al., 2021). Most importantly, critics point out that these equity and environmental problems are likely to go unaddressed because the high cost of

running LLMs has made their use exclusive to very large and well resourced corporations, creating economic barriers and limiting access to only wealthier and more powerful entities (Knight, 2021).

The "Stochastic Parrots" Controversy

Large language models gained notoriety in the wake of the firing of ex-Google employees Timnit Gebru and Margaret Mitchell. Gebru co-led Google's "Ethical AI team" with Margaret Mitchell, and along with academic and Google colleagues, co-authored a paper on the risks and failings of LLMs called "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" The paper raised concerns about the environmental impacts, and problems with training data including unmanageability, encoded bias, and lack of accountability (Bender & Gebru et al., 2021). According to Gebru, in winter 2020 Google attempted to prevent her from releasing the paper without major revisions; when Gebru refused, they fired her (Metz & Wakabayashi, 2020). The ensuing controversy eventually led to Mitchell's firing, and publication of the original version of the paper without Google's edits (Bender & Gebru et al., 2021; Simonite, 2021).

In this report, we anticipate the potential implications of LLMs by analyzing the history of similar technologies, using what we call an analogical case study method. We then focus on one domain where LLMs are likely to have a significant impact: scientific research. We conclude with recommendations for both policymakers and the scientific research community, and a "code of conduct" to guide the practices of LLM developers.

LLMs are still new and experimental, and therefore their social impact is still emerging. But history teaches us that their impact will be profoundly shaped by those creating it.

LLMs are still new and experimental, and therefore their social impact is still emerging. But history teaches us that their impact will be profoundly shaped by those creating it, and at present, because of the enormous capacity and resources needed, the primary developers are a handful of large tech companies. As we discuss throughout the report, this should give us some cause for concern. We noted above that LLMs could improve and increase access to specialized expertise from law, medicine, science, and more. They could also make technical information more widely available to the public, ultimately empowering individuals and communities. However, it is unlikely that the technology will be able to achieve any of these benefits if it is built by a narrow group of elites and without proper technology assessment and

PDF: Click to return to top

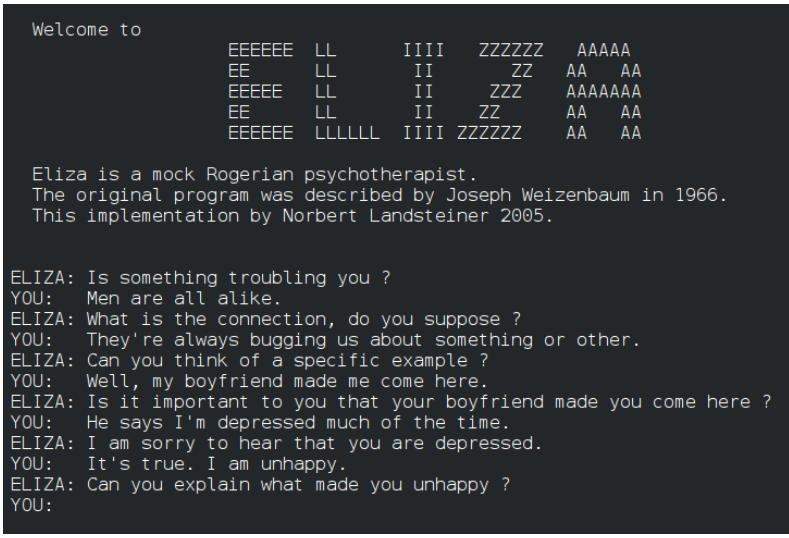
Because LLMs are still at an early stage, we can still shape the development and implementation of the technology to ensure that it serves the public interest.

oversight. Because LLMs are still at an early stage, we can still shape the development and implementation of the technology to ensure that it serves the public interest.

History of Automating Language

In 1950, mathematician and philosopher Alan Turing proposed that a machine should be considered intelligent if a human could not tell whether another human or the computer was responding to their questions; this was the beginning of the ongoing quest to automate human language (Oppy & Dowe, 2021; Turing, 1950). The "Turing Test" is still used to measure whether a "talking" program is communicating successfully (Computer AI passes, 2014). While developers have since imagined that automated language programs could be used for a variety of service and artistic tasks, the success of the effort is still measured, at least in part, by its capacity to imitate human language.

The ELIZA program, created by Joseph Weinbaum in 1964 at the Massachusetts Institute of Technology's (MIT) Artificial Intelligence Lab, was one of the first language programs that could hold a conversation (Liddy, 2001). ELIZA was not programmed to learn connections between ideas or words like today's LLMs. Instead, using the dominant approach of the period, Weinbaum's team manually developed simple linguistic rules and automated them, which allowed ELIZA to respond in conversation with phrases that reflected back what had just been said (Epstein, 2001). Weinbaum's goal was to demonstrate that computers did not have intelligence by showing that the conversations were too simplistic, but users anthropomorphized the computer, and other researchers determined that the rudimentary chatbot might even have therapeutic value despite its simple communication.



Public domain

Aside from Eliza, most early language automation research was driven by U.S. military priorities and government funders largely considered them inadequate

(Hutchins, 2003). From the mid 1950s to 1964, most federal research funding focused on machine translation, specifically Russian to English translations for the Cold War. In 1964, the Department of Defense, National Science Foundation, and Central Intelligence Agency among other branches of the government created the Automatic Language Processing Advisory Committee (ALPAC) to assess the utility of this work in terms of advancing government priorities, reducing costs, improving performance, or addressing a need that humans could not fill. To evaluate the translation program, ALPAC compared Russian to English machine translations to human translations based on intelligibility and fidelity, and found that humans far outperformed machines, human translation was less costly after edits, and that the government had sufficient capacity for Russian translation. Based on these findings, the committee recommended that defense agencies stop funding machine translation, and that the NSF switch to only funding basic computational linguistics research.

As a result of the ALPAC report, the federal government largely stopped funding machine translation research until the Advanced Research Projects Agency took up the subject again in the 1990s, but private industry continued to tackle machine language projects aimed at other uses like text generation and conversation. During that period, research in machine language moved towards methods of representing and communicating meaning in dialogue and natural language generation, which more closely resembles the goals of today's LLMs (Hutchins, 2003). For example, in 1969 and 1970, researchers introduced new methods for representing language input designed to help machines develop conceptual understandings of words so their responses could be more useful (Moltzau, 2020). These developments led to the first program that used simple natural language inputs - language written as it normally would be in human conversation - to control a machine at Massachusetts Institute of Technology in 1971. By 1985 the main uses developers imagined for language programs at the time could be loosely categorized into 6 groups: 1) interfacing with databases, 2) conversational interfaces for programs, 3) content scanning of semi-formatted texts to determine actions, 4) text editing for grammar and style, 5) translation, and 6) transcription of spoken input (Dale, 2017).

By the 1980s the increased availability of computing power allowed researchers to begin integrating statistical methods into natural language program development (Hutchins, 2003). These statistical approaches, called Natural Language Processing (NLP), essentially allowed the computer to "learn" for itself how language works, by identifying patterns from text-based "training" datasets rather than relying on researchers to lay out complex rules based on linguistics research. As the amount of digital text grew, researchers were able to compile larger and larger data sets which improved the performance of these "statistical" language programs. Eventually statistical methods outperformed and replaced programs based on linguistic rules and NLP has become an interdisciplinary field bringing together insights from linguistics, computer science, and artificial intelligence.

PDF: Click to return to top

Today's LLM Landscape

At present, large tech companies dominate the development of LLMs because of the enormous resources required in terms of finances, infrastructure, personnel, and "compute", a measurement in supercomputing that corresponds to how many computational operations take place and ultimately, how many resources are required (Luitse and Denkena, 2021). Even academic researchers must partner with the private sector in order to obtain the resources to develop truly large language models that can approach the capabilities of cutting edge LLMs. This monopolization raises a few concerns. First, LLMs are likely to reflect the priorities of the private sector, or more specifically a handful of the most powerful tech companies, complicating their potential to achieve societal benefits. Second, the prominent role of the private sector makes it more difficult for third parties to assess the technology, while internal researchers are likely to be under pressure to paint a rosy picture, as demonstrated by Google's decision to fire Timnit Gebru and Margaret Mitchell (See "Stochastic Parrots" Text Box) (Hao, 2020; Schiffer, 2021).

The major LLM developers are the US-based Alphabet (Google), Meta (Facebook), Microsoft, and OpenAI, and the China-based Alibaba Group and Baidu. Some LLMs are built entirely within one company, including

Credit: Baltic Servers

Google's BERT, Alibaba's VECO, and Facebook's M2M-100 (Alford, 2020). Even the developer of GPT-3, OpenAI, was created by investments from Microsoft, Infosys, and several venture capital firms and tech billionaires. It recently gave up its non-profit status to become a company that profits from AI products sometimes built exclusively for investors.

First, LLMs are likely to reflect the priorities of the private sector, complicating their potential to achieve social benefits. Second, the prominent role of the private sector makes it more difficult for third parties to assess.

Many of these same companies also fund university research. Google, IBM, and Wells Fargo help fund a prominent center, the Stanford Human-Centered Artificial Intelligence (HAI) Lab, while Microsoft and Toyota frequently

partner with Massachusetts Institute of Technology's (MIT) Computer Science and Artificial Intelligence Lab (Stanford HAI, n.d; MIT CSAIL, n.d). The top cited LLM research papers are co-authored by researchers from industry and academia. In addition to providing university researchers with LLM access, these collaborations also provide the private sector counterparts with scholarly credibility.

Government funding agencies have increasingly encouraged university-industry collaboration. In 2019, the US National Science & Technology Council released a strategic plan that prioritized collaboration between academia and industry (Select Committee on Artificial Intelligence, 2019). The plan argued that these collaborations would address problems with safety, predictability, ethics, and legal questions proactively, before AI products are developed. This had an immediate impact: the NSF, which had previously focused on AI research within universities, has now launched seven national AI research institutes that facilitate industry-university collaborations (Gibson, 2020). Three of these have focus areas related to LLMs. Similarly, the European Commission is funding several international university-industry research collaborations, often financing links between European academic researchers and US tech companies (Stix, 2020). The Japanese and Chinese governments are fostering similar collaborations (AIRC, n.d.; Luong & Arnold, 2021).

There are also attempts to create open-source or more transparent LLMs, but some of these projects are backed by the same venture capital firms that fund the for-profit entities. For example, Hugging Face has developed open source resources for the production of LLMs, including datasets and models, and explicitly invokes the values of accessibility and democratizing innovation (Dillet, 2021). It supports users as they develop and upload their own content, allowing for transparency and collaboration of model and dataset construction. Hugging Face hired Margaret Mitchell, a co-author of the Stochastic Parrots paper (See "Stochastic Parrots" Text Box" above), to lead data governance efforts. In addition, Hugging Face has initiated the BigScience Project, an effort to create and share datasets, models, and software tools in order to reveal and minimize potential problems with LLMs (Hao, 2021). Similarly, EleutherAI has developed open source language models intended to replicate GPT-3, as well as an 860 GB dataset for language modeling (EleutherAI, n.d.; Gao et al., 2020). While EleutherAI is volunteer-based, the collective depends on donated GPU compute from CoreWeave, which is part of the NVIDIA Preferred Cloud Services Provider network (Leahy, 2021). The tightness of the LLM development landscape means that even seemingly alternative or democratic approaches to LLM development are likely to reinforce the priorities and biases as large developers (AI Now Institute, 2021).

Social scientists, ethicists, and computer scientists are starting to investigate the social implications of LLMs, sometimes with the assistance of government funding (Birhane et al., 2021). Stanford's HAI, for example, recently announced a new interdisciplinary research arm, the Center for Research on Foundation Models (CRFM) to "study and

build responsible foundation models'' (Stanford CRFM, n.d.). The amount of critical analysis on LLMs has grown since Google's decision to fire AI Ethics leads Gebru and Mitchell, but it is limited by a lack of industry collaboration, which leaves non-industry researchers without basic access to the model or information about the contents of the corpus.

Influential Large Language Models Today

Throughout the report we refer to specific LLMs periodically, especially Google's BERT (Devlin et al., 2018a), OpenAI's GPT-3 (Brown et al., 2020), and EleutherAI's GPT-J (Romero, 2021) as they represent specific technological achievements and are commonly referenced in the LLM community. BERT was the first LLM to implement and show the promise of the transformer architecture, which is the key innovation that has allowed LLMs to process such large amounts of data. OpenAI built on this architecture and created GPT-3, which demonstrated how new behaviors and levels of accuracy emerge simply by increasing the size of an LLM. GPT-J is an LLM called the "open source cousin" to GPT-3. It does not perform as well, but is notable as a grassroots effort to replicate corporate LLMs (Romero, 2021). Hugging Face is a prominent source of LLM development, though they focus on supporting development of and providing access to different models, documentation, and corpora (Hao, 2021).

TABLE 1. INFLUENTIAL LLMS

Model Name	**Developer** Name	**Released** Year	**Size** In parameters	**Hallmark / uniqueness**	**Availability**
BERT	Google AI Language	2017	340 million	Demonstrates the transformer architecture, which is what enables LLMs to be large	Open source (Devlin et al. 2018b): Model and corpus are all available to use and adapt for free
GPT-3	OpenAI	2020	175 billion	Was the largest model at the time of release; demonstrates that new properties emerge simply by increasing the size of a model	App developers can apply for paid API access; OpenAI indicates this is a temporary safety measure
GPT-J	EleutherAI	2020	6.7 billion	Open source grassroots version of GPT-3	Open source: Model and corpus and are all available to use and adapt for free
WuDao 2.0	Beijing Academy of Artificial Intelligence (BAAI)	2021	1.75 trillion	First model to reach 1 trillion parameters; model is multimodal (trained on both text and images)	Open source (Romero, 2021b): Model and corpus and are all available to use and adapt for free

Developers and Users Interact with LLMs at Different Levels

Throughout this report, we describe the LLM landscape in terms of three types of participants. LLM developers are the entities (usually companies) creating LLMs. First-order users are the entities, likely to range from individuals to self-organized community groups to large companies, who develop tailored apps to harness an LLM's power, often fine-tuning it for a particular purpose. OpenAI, for example, has created a software interface that allows apps or websites to send input text to GPT-3 and receive the output text that the model generates based on that input. One such example is Viable, an app that companies can use to synthesize and extract insights from customer feedback (Viable, n.d.). In another case, a first-order user fine-tuned GPT-3 to mimic his deceased fiance in a chat application (Fagone, 2021). Finally, second-order LLM users are the publics who use the apps created by first-order users.

TABLE 2. TAXONOMY OF ENTITIES INVOLVED IN LLM DEVELOPMENT AND DEPLOYMENT

	Description	Example
LLM developers	Companies or other organizations creating LLMs	OpenAI, Hugging Face, Eleuther
App developer	Developers who integrate the LLM into an app or product that is deployed for others to use	Company that develops an LLM-enabled platform for extracting insights from customer feedback
End user	Person or entity that uses an app or product built on top of an LLM	Company that uses an LLM-enabled platform for extracting insights from customer feedback

Uses of LLMs

As LLMs increase in size, they will be able to perform a growing number of tasks. Each new generation of models brings new functionality, so some of these tasks will be difficult or impossible to predict. However, we know that LLMs are likely to be able to generate, summarize, translate, and engage in dialogue. In what follows, we discuss the near-future functional applications of LLMs and the different ways technologists or developers might interact with them.

Generating Text

When an application or end user gives an LLM a word, sentence, or paragraph(s), the LLM will be able to return the following word, sentence, or paragraph(s) based on the patterns learned from the training data. For example, when given the headline of a hypothetical news article, the LLM would return an entire article with credible text to match that headline. As a result, an LLM could be immensely helpful as a writing assistant, helping users generate text (Better Language Models, 2019). However, this raises concerns about authenticity in authorship and the automation of propaganda. Not only do humans have trouble identifying LLM-generated text, but a GPT-3 bot also generated and posted comments on Reddit for a week without anyone noticing (Heaven, 2020) and a study showed that Twitter users' political opinions can be swayed by GPT-3 tweets (Knight, 2021).

LLMs can also generate computer code; GPT-3 was able to write code because its corpus likely contained tutorials and discussion posts with snippets of text that had human language descriptions followed by code (Brown et al., 2020). Developers and companies are now refining this functionality and integrating it into different end-user products (OpenAI Codex, 2021; Vincent, 2021), which raises questions about code quality, security, and intellectual property.

Summarizing and Extracting Information

LLMs will likely be able to summarize web pages and other content. Google has already incorporated some of this basic functionality into its search engine, and LLM summarization could ultimately be used to produce a new kind of search engine that responds directly to a query rather than simply providing a ranked list of related links (Heaven, 2021). It could also help users understand key information from lengthy documents such as

Not only do humans have trouble identifying LLM-generated text, but a GPT-3 bot also generated and posted comments on Reddit for a week without anyone noticing and a study showed that Twitter users' political opinions can be swayed by GPT-3 tweets.

technical reports or legislative text (Tamkin et al., 2021). Similarly, companies may be able to better understand and extract key information and takeaways from customer feedback or other interactions (Viable, n.d.).

Translating Text

LLMs can translate text across languages as well as transform text across linguistic styles. For example, LLMs could turn legalese into plain language (Blijd, 2020) or create a version of content written in an individual's writing or speaking style (Better Language Models, 2019). They could also facilitate international communication by translating between languages. However, the precision of the translation depends on how much of that language or linguistic style is in a given corpus, and as noted above most of the corpora of prominent US-based LLMs are in English (Brown et al., 2020).

Engaging in Dialogue-based Conversation

LLMs will also likely be able to converse with humans coherently (Better Language Models, 2019). As a result, like ELIZA and other early chatbots, LLMs could provide companionship, psychotherapy (Zeavin, 2021), or medical advice (Rousseau et al., 2020). LLMs could also help with idea generation. Design and consulting firm IDEO gave an LLM a sentence description of the problem they were trying to solve (i.e., encourage better spending habits) and it generated possible product ideas such as "Reward the user with real money if they don't spend money at all in a month" (Syverson, 2020).

Perhaps most immediately, companies are likely to use conversational LLMs to make more sophisticated customer service chatbots and generate more meaningful search results when customers ask complex questions online (Algolia, n.d.). Experts have suggested however, that LLMs will probably not be able to exhibit this functionality without additional fine-tuning (Patterson, n.d.). One serious concern with conversation applications is that there is a high risk that LLMs will push dialogue in dangerous or socially undesirable directions: for example, one model encouraged suicide (Daws, 2020).

Policy Landscape

No laws, anywhere in the world, specifically cover LLMs. Nor are there credible third-party assessments of their accuracy. Instead, a variety of laws and policies related to copyright, data privacy, and algorithm accuracy touch on both LLM production and the content that they produce.

LLM corpora include millions of books and articles that are individually protected by copyright. However, in the United States a string of legal cases have established that computers can index, search, and archive these texts without violating copyright protections (Grimmelman, 2016). The European Union gives media companies more control over how search engines can use and display certain kinds of content like news, but it still permits text and data mining of the

No laws, anywhere in the world, specifically cover LLMs. Nor are there credible third-party assessments of their accuracy. Instead, a variety of laws and policies related to copyright, data privacy, and algorithm accuracy touch on both LLM production and the content that they produce.

kind used to build a training corpus (Council of the EU, 2018). The corpora themselves are not protected by their own copyright, but companies generally treat them as proprietary and remain secretive about what they contain.

Meanwhile, there is great controversy about the copyright status of AI-generated works, including the output of LLMs. In the United States, a human must create a work in order for it to be eligible for copyright protection (Guadamuz, 2016; U.S. Copyright Office Review Board, 2022). Under that framework, anything an LLM writes will be in the public domain, free for anyone to use or adapt in any way without permission. Copyright non-profit Creative Commons and some legal scholars support this approach, while others believe that the laws should be changed to extend copyright protections to AI-generated work (Hristov, 2017; Vézina & Moran, 2020). The UK, Ireland, and New Zealand provide limited protection to computer generated works, but these protections generally assume some level of participation by a human creator, and do not provide for fully autonomous authorship (Vézina & Moran, 2020).

LLMs also raise questions about data privacy and security, which is covered by the European Union's General Data Protection Regulation (GDPR) and similar laws in several countries and US states. Because many LLM corpora include text scraped from the internet, they may inadvertently include personal information or multiple pieces of text that the model could put together to deduce private information. Studies have shown that if it receives the right prompts, an LLM could output this kind of personal data (Carlini, 2020). The GDPR focuses on the rights of people whose data companies collect, and on the responsibilities of companies collecting that data; as such, it is unclear whether an LLM developer could be held liable in Europe for gathering or using personal data available to programs that scrape the internet. It is possible that the website that originally hosted the data would be accountable because the primary source should have prevented web scrapers from accessing sensitive information. Because most LLMs are not available to the public for general use, the risk of these security breaches remains theoretical, and we are not aware of any legal proceedings related to an LLM-involved data breach.

Finally, except for occasional bans or moratoria on algorithm-based technology such as facial recognition, there is no systematic regulation of algorithms anywhere in the world. However, governments are starting to consider strategies. In 2021, the European Union proposed the Artificial Intelligence Act which adopts a risk-based approach to regulating all of these technologies, which would presumably include LLMs (European Commission, 2021). AI that poses greater societal risk would be subject to more regulation. In the United States, Congress has proposed various bills. One of the most comprehensive is the Algorithmic Accountability Act of 2022, proposed by Congresswoman Yvette Clarke and Senators Cory Booker and Ron Wyden (Wyden, 2022). Much more limited than the pending EU legislation, it would require companies to assess the impacts of their AI systems and disclose their findings to the Federal Trade Commission (FTC), create a new Bureau of Technology within the FTC, and require that the FTC publish an annual report on algorithmic trends which would help consumers better understand the use of AI.

Analogical Case Study approach

In this report, we analyze LLMs using an analogical case study (ACS) approach. Over the last few decades, as societies have begun to contend with the complex implications of technologies and sometimes even mobilize against them (Parthasarathy, 2017; Schurman & Munro, 2010), social scientists and humanists have argued that scientists, engineers, and policymakers can do a better job of predicting their impacts. These insights can then be used to design, implement, and govern technologies better to maximize benefits while minimizing harms and maintaining public trust in science and technology. Researchers have experimented with multiple methods to accomplish this "anticipatory governance" (Michelson, 2016; Stilgoe et al., 2013; Fisher et al., 2006; Selin, 2011; Eschrich & Miller, 2021; Hamlett, Cobb, & Guston, 2012; Stirling, 2008).

We hypothesize that by understanding how societies have managed past technologies, we can anticipate how they might do so in the future.

Our analogical approach rests on findings from the field of science and technology studies (STS) that there are social patterns in the development, implementation, and implications of technologies (Browne, 2015; Bijker et al., 1987; Parthasarathy, 2007). We hypothesize that by understanding how societies have managed past technologies, we can anticipate how they might do so in the future. Furthermore, controversies over previous technologies offer insights into the kinds of concerns and resistance that might arise, groups who might be

affected, and solutions that might be feasible with emerging innovation (Nelkin, 1992). As Guston and Sarewitz (2002) argue: "knowledge about who has responded to transforming innovation in the past, the types of responses that they have used, and the avenues selected for pursuing those responses can be applied to understand connections between emerging areas of rapidly advancing science and specific patterns of societal response that may emerge" (p.101). By deliberately considering the histories of analogical technologies across sectors, our method identifies relevant social patterns in how technologies develop and are implemented. It also allows us to identify successful social and policy approaches to managing technological harms.

Our analytic approach to LLMs builds on the method we developed to study facial recognition technologies and vaccine hesitancy (Galligan et al., 2020; Wang et al., 2021). We began our work in May 2021 by training the research team, composed of a diverse group of faculty, staff, a postdoctoral fellow, and undergraduate and graduate students, in some of the basic concepts related to the history and sociology of technology and the ACS methodology. To help stimulate our creativity, we read some speculative fiction that imagines how AI might shape the future (Jemisin, 2011; Jemisin, 2012). We also reviewed the scholarly and journalistic literature to understand the projected implications of LLMs. However, because the technology is at such an early stage of development, this literature is small and it has been produced almost exclusively by NLP researchers and journalists. Team members used this literature as well as primary sources to develop an understanding of the history, political economy, and technical dimensions of LLMs.

We then brainstormed two types of analogical cases. Type 1 cases are similar to LLMs in terms of their function (i.e., processing large amounts of data, often with the purpose of prediction), while Type 2 cases have similar implications as those projected for LLMs (e.g., racial bias, massive energy use).

We investigated these cases, which intentionally draw from both historical and more recent technologies, in areas both similar to and different from LLMs. For example, to help us understand the implications of potential biases embedded in this emerging technology, we looked at medical technologies including the spirometer and pulse oximeter. To understand how LLMs might pose challenges to how we understand expertise and professional competence, we looked at traffic lights, which removed traffic management from the domain of law enforcement officers. We also looked at biobanks, large scale repositories of DNA and other forms of data used for the purpose of facilitating biomedical research and ultimately predicting and alleviating human disease. We adopted an iterative process: after we worked our way through the initial set of cases and presented our insights to one another, we reflected about the potential implications of LLMs. We then generated an additional list of cases, and so on until we were confident that we had exhausted the social, ethical, equity, and environmental implications that we could anticipate.

Based on our analysis of dozens of cases, we identified six broad implications of LLMs: three related to their construction and three related to their use. Because LLMs are likely to transform many high-skilled professions–albeit in different ways–we then focused on the implications for one: scientific research.

Background: How do Large Language Models Work?

KEY POINTS

- LLMs are considered more "intelligent" than previous NLP efforts due to their capacity for complex language patterns and ability to behave appropriately in novel situations.
- LLMs learn language from datasets of human-written text from the internet and digitized books that are so large the developers who assemble them often do not know the entirety of their contents.
- While LLM developers may be able to assess the performance of their models, there is no standard approach.
- LLM developers can fix the model's behavior through fine-tuning, but they must both identify problems and develop solutions manually.

LLMs emerged from the field of NLP, where models have grown dramatically in size and sophistication in recent years with the availability of data and more computing power. Increased computing power, in particular, has made it easier and quicker for researchers to collect and categorize data and perform more sophisticated operations. Today's language models, from the simplest to the most advanced, share key features with the original efforts: a training data set, a process for learning patterns in the data set, and the use of the model to generate new text.

Today's language models, from the simplest to the most advanced, share key features with the original efforts: a training data set, a process for learning patterns in the data set, and the use of the model to generate new text.

LLMs differ from their predecessors in two critical ways. LLMs are much larger, both in terms of the massive amounts of data developers use to train them, and the millions of complex word patterns and associations the models contain. LLMs also more closely embody the promise of "artificial intelligence" than previous NLP efforts with smaller data sets because they can complete many types of tasks without being specifically trained for each one. They are "intelligent" because the complexity of the association model allows LLMs to respond to questions and tasks they have never seen before, in the same way they process other language inputs, and identify appropriate responses. This characteristic in particular makes any single LLM more widely applicable than previous NLPs.

LLMs are few-shot learners. This means that they do not need additional fine-tuning to be able to perform different types of useful operations (Brown et al., 2020). To use an LLM, the user inputs a task description, a few examples, and then the prompt for the model to continue. The model is then able to predict the next words, sentences, or paragraphs. This makes LLMs versatile, with the ability for users to apply them to tasks that the LLM developers did not necessarily anticipate.

Developing an LLM involves three steps, each of which can dramatically change how the program models language, and therefore how it will function when it is used. First, developers assemble a dataset, or "corpus", of text-based documents. Second, the algorithm learns about word relationships from this data. Finally, developers iteratively assess and fine-tune the model as needed to fix specific problems. In theory, models can continue to be fine-tuned even once they are in use, although the time and manual labor required of such a change may render ongoing maintenance of this kind impractical. Developers assess model performance against a number of formal and informal metrics such as how well it completes sentences, how accurately it translates to a different language, and whether a human can tell if text was written by the model or by a human.

In theory, models can continue to be fine tuned even once they are in use, although the time and manual labor required of such a change may render ongoing maintenance of this kind impractical.

Once the language model is trained, it can generate and translate text and answer questions, among other tasks based on initial input given by a user. For example, someone could feed an LLM the headline of a hypothetical news article and it would be able to generate a possible body of the article based on text that followed similar phrases in the training set. While LLMs do not understand the text they generate in the way a human does, because they are

repeating patterns developed from human-written text, they are able to generate text that closely resembles what a human might write. For any initial input given to the model, it is as though the model is asking itself "if I were to encounter this text in a document in my training set, what would I expect to see next?" Their ability to do this convincingly is based on the amount of data they are trained on and the size and sophistication of the algorithms.

For the past few years, LLMs have been very rapidly increasing in size. Specifically, the number of pieces of information about language each model stores has been increasing by ten times each year (Li, 2020). As the size of the models increase, the definition of a "large" language model is also shifting. When we discuss large language models, we are also anticipating future models that are even larger without knowing how large they will become or what emergent capabilities they will have: GPT-3, one of the largest LLMs ever developed, has better performance and new behaviors even though the only difference from previous models is its size (Brown et al., 2020). For example, GPT-3 can generate snippets of code based on human description of the desired code functionality, which was neither intended as a feature of the model, nor was it a characteristic of smaller models with the same training process.

Gathering the training data or "corpus"

Each LLM is trained on a large dataset, called a corpus, consisting of many text-based documents such as books, newspaper archives, and websites. In order to comprehend LLMs, we must first understand these datasets and the decisions behind them because they fundamentally shape the output. The large size also limits who can create and maintain an LLM.

The corpora used to train LLMs are massive compared to those used in previous NLP endeavors. The corpus for GPT-3, for example, was 570 gigabytes (GB) in size (Brown et al., 2020). For a sense of scale, 1 GB is about 1,000 400-page books, or about 10 yards of physical books on a shelf (Gavin, 2018). The smaller corpora of past language models could be stored and even processed on ubiquitous computer hardware, but massive corpora require computing space that few have access to. Creating such a large corpus from scratch is not practically possible, and even gathering and curating a collection of this size requires a long time and many human and financial resources. Instead, LLM developers typically take advantage of already-written and curated bodies of text such as collections of digitized books and the user-created text that comprises the internet. Each LLM is thus trained on a collection of different sources, and each type of source is given a weight, or a percentage that represents how much of the final dataset contains data from that source. Weighting addresses the challenge of balancing quality

reliable sources of text with ensuring there is diverse text from a range of different sources. The corpus GPT-3 was trained on, for example, contains text from Wikipedia, the internet, and online collections of books, with greater weight put on the text from Wikipedia even though the corpus contained a greater volume of data from other sources (Brown et al., 2020). Putting greater weight on Wikipedia was a way for the LLM developers to ensure emphasis on text they trusted more (Brown et al., 2020).

Many LLMs (including GPT-3 and BERT) use text from the internet in their corpora. The most common way that developers incorporate internet text is by way of a large dataset called the Common Crawl. Although the Common Crawl is managed by a non-profit organization of the same name, it has deep ties to Google and other large tech companies, and is hosted by Amazon Web Services. The Common Crawl dataset includes millions of GB of data (Common Crawl, n.d.), and is updated once a month. Each update contains 200 to 300 terabytes (TB; a TB is equal to 1000 GB) of textual content scraped via automated web crawling (Luccioni & Viviano, 2021, p.1). It is constructed from the text of websites, but its archive represents only part of each website crawled. The organization argues that this creates a "representative" sample of the internet, but this approach also allows it to claim a fair use exception to copyright laws by only using a portion of each site, instead of the whole thing (Luccioni & Viviano, 2021, p.1). Overall, the Common Crawl dataset represents the text of the internet's most frequent users, who are disproportionately younger, English speaking individuals from Western countries who often engage in toxic discourse (Luccioni & Viviano, 2021, p.5). Therefore it includes a significant amount of harmful data including text that is violent, targets marginalized groups, and perpetuates social biases (Luccioni & Viviano, 2021, p.3). Because LLMs identify and replicate patterns, the inclusion of this data creates a significant risk that without explicit additional case-by-case training, LLMs will produce language that is similarly harmful and biased. LLMs

LLM training data includes a significant amount of harmful material including text that is violent, targets marginalized groups, and perpetuates social biases.

PDF: Click to return to top

do not at present have the capacity to automatically detect this kind of language without specialized training.

OpenAI, the former non-profit-turned-private tech venture that built GPT-3, has another approach to overcoming the quality challenge of internet text. Its internet corpus, WebText, is a 40GB dataset that contains the text from outbound Reddit links that received at least 3 "karma," or upvotes from users (Radford et al., 2019). Their rationale is that these linked and upvoted webpages are more likely to contain quality text because someone bothered to link and upvote them. This may be true, but these linked web pages are also more likely to represent the values and ideology of Reddit users, which are also not representative of the general population (Morales et al., 2021). Meanwhile, because WebText is not available in full for use by people outside of OpenAI, others have tried to construct a publicly accessible version of the same corpora, also based on Reddit links (Gokaslan & Cohen, 2019).

While Common Crawl's corpus is open and available for anyone to use, it is huge, complex, and heterogeneous. As a result, it requires a large amount of computational resources to download and process the data, which means it requires high compute (Luccioni & Viviano, 2021, p.2). Therefore, only researchers at elite universities and large companies are likely to have the financial resources, expertise, and personnel to be able to use it to build their own LLMs. The LLMs they build are likely to reflect their values and priorities; this will influence LLMs' capacity to truly democratize text and knowledge.

LLM developers also draw on collections of digitized books, but often offer few details about the composition of the collections or the rationale behind it. This matters because some datasets may be more or less appropriate for training an LLM. The BooksCorpus, for example, part of the corpus used to train BERT, was originally curated for a completely different project designed to train an AI to generate rich descriptive text when given video or images (Zhu et al., 2015). It contains over 11,000 free web-based books written by unpublished authors, but the curators do not include much detail about its contents other than a breakdown of a few of the genres (2,865 romance books, 1,479 fantasy books, etc.). The developers who used this dataset to train the BERT LLM also did not not provide additional details such as what ideas the corpus includes, whose voices it represents, or why it is an appropriate part of their training corpus. Even more opaque are the Books1 and Books2 datasets, which are part of the training data for OpenAI's GPT-3 (Gokaslan & Cohen, 2019). There is no discussion at all of what these datasets contain, how they were constructed, or what they represent in the context of training the LLM (Scareflow, 2020).

Furthermore, these corpora are often private; most LLM developers do not allow others to inspect or build on their dataset. A rare exception is Eleuther AI, which, as we describe above, takes a more democratic approach to its LLM overall. Eleuther AI developers created the Pile, a publicly available English-text corpus that is about 886 GB and made up of existing corpora such as OpenWebText2 and Books3, as well as internet based datasets such as a filtered

version of Common Crawl (Gao et al., 2020). The Eye, a non-profit, community driven and funded group, which archives a variety of creative materials, hosts this corpus (The Eye, 2020).

These bodies of text are often so large that not even the developers know what is in them.

Overall, our observations about the composition of LLM corpora echo what Bender and Gebru et al. (2021) have said about LLMs; these bodies of text are often so large that not even the developers know what is in them.

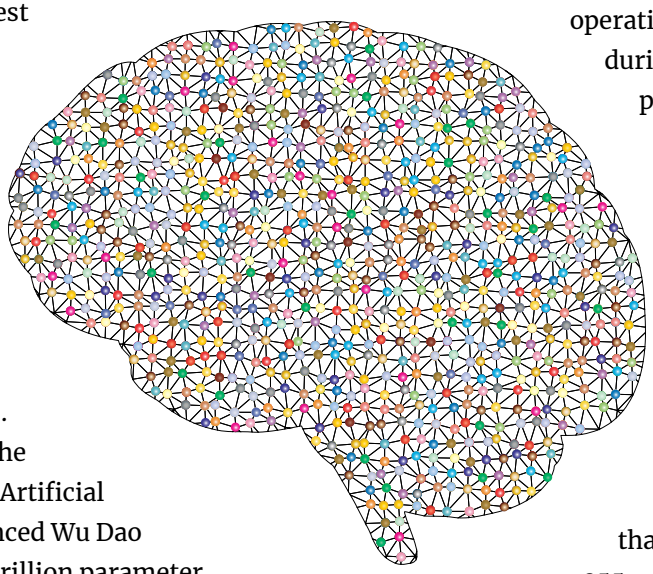
Training the model

When training a language model, developers first make decisions about the model's setup and the process the algorithm will use to learn from the training data. This includes the architecture they will use for training. LLMs are able to operate on a massive scale thanks in part to the invention of the transformer architecture, which was introduced in 2017 and allows the model to learn the relationships between any two words in a sentence as opposed to only the one or two neighboring words as was the previous norm (Vaswani et al., 2017).

In many LLMs, words are represented as vectors: lists of numbers that represent coordinates in a many-dimensional space. This allows computers to use math to understand the relationships between words and sentences and predict what words should be used to complete a sentence or paragraph. There are different methods for generating these word vectors, but recent advances have developed techniques that take into account the fact that different words have different meanings in different sentences. Everything the LLM learns about the relationship between words is based on what is written in the training data. For example, it will only associate "sky" with "blue" if there are many sentences in the corpus that demonstrate an association between those words. Bender and Gebru et al. (2021) caution that although LLMs might appear to be intelligent and coherent, they have no understanding of the underlying properties or relationships between the concepts that words represent and often generate nonsense as a result.

LLM developers must also decide on a strategy for tokenization when they are training an LLM. Tokenization involves breaking the text from a document into pieces for analysis, or tokens. Some tokenization algorithms convert each word into a token, others break words apart, creating, for example, separate tokens for "sing" and "ing" in the word "singing". Different approaches for tokenization may affect model performance down the line, but, like many decisions made when training an LLM, it is hard to predict what the impact will be.

The number of parameters, or pieces of information about language stored by the model, is also crucial. Like neurons in a human brain, the parameters work together to store and process complex information. Models that have more parameters are able to remember more complex patterns from the training data, such as how sentences, paragraphs, and documents are structured, but they require more compute to train. The number of parameters in state of the art LLMs is increasing by a factor of about 10 each year (Li, 2020). At its creation, GPT-3 was the largest at 175 billion parameters, and at the time developers were already eyeing the possibility of a 1 trillion parameter model (Brown et al., 2020). A short time later, the Beijing Academy of Artificial Intelligence announced Wu Dao 2.0, which is a 1.75 trillion parameter model (Zhavoronkov, 2021). Parameter size seems to have significant impact on LLM performance. GPT-3 (175 billion parameters) performs far better than GPT-2 (2.7 billion parameters) along several measures (Brown et al., 2020). Likely as a result, competition among developers is primarily about the size of the LLM.



After the developers have determined what architecture to use, the number of parameters, and the strategy it will use for tokenization, the model is ready for training. The actual training process involves the model traversing the training corpus piece by piece and updating its parameters according to what it learns from each phrase it encounters.

Training LLMs requires a massive amount of compute, which is only available to a handful of people who have access to high performance computers or cloud computing services through their institutions (Ahmed & Wahed, 2020). The developers of GPT-3 report how many computational operations were performed during the training process, but they do not give details on the hardware setup, the cost of training, or how long the training actually took. Based on the reported numbers, other researchers estimate that it could have taken 355 years to train GPT-3 on a single graphics processing unit (GPU) or could have cost $4.6 million with the necessary parallel hardware for faster training (Li, 2020). There is no doubt that training LLMs is expensive. In fact, the developers of GPT-3 noticed errors in their data during the training process, but did not start the training process over because it would have been too expensive and time consuming (Brown et al., 2020). While initiatives such as the federally-funded National Research Cloud aim to broaden access to compute, it is likely to mostly

only help those who already have access to significant amounts of compute on their own because the gap in access is so great (AI Now Institute, 2021).

Fine-tuning the model

After an LLM is trained, it can be fine-tuned. This is a relatively light-weight process that involves feeding the model a few hand-picked examples, and is often used to train the model on socially sensitive topics. The model learns from these examples and changes a few of its parameters without changing the core of the model. OpenAI, for example, created a "values-targeted" version of GPT-3, which they fine-tuned using 80 additional human-written examples that illustrate preferred behavior so it would answer questions about subjective topics such as violence, injustice, human characteristics, and political opinion in what they deemed a desirable manner (Solaiman & Dennison, 2021). After learning from these additional examples, GPT-3 generated drastically different responses to questions about these topics.

Fine-tuning allows dynamic updates to address problems with LLM outputs that are far faster and require less compute than model training. However, this process depends on the sensibilities and knowledge of developers, who will need to decide which topics require additional training and carefully curate the examples. In many cases, developers may not know which examples will be best for fine-tuning. To address these challenges, some organizations have decided to "democratize" fine-tuning. OpenAI is taking steps to allow anyone to create their own fine-tuned version of GPT-3 (OpenAI, 2021b). But this poses new risks and uncertainties. People could feed the LLM hateful or unethical text to produce a more socially dangerous technology. Meanwhile, those with more positive intentions have no guarantees that socially beneficial fine-tuning will be adopted in the main GPT-3 model.

Understanding the model's output and capability

There is no universal standard for assessing LLM quality. However, developers usually ask their models to perform a common set of tests in order to assess performance. These include asking the model to complete sentences, correct grammar, answer commonsense questions, translate sentences to another language, answer reading comprehension questions and compare the results to other LLMs and to a human on the same tasks.

Developers are under no obligation to disclose either the tests they perform or the results, which makes it difficult for third parties, including consumers, to evaluate performance. But some publicly available assessments of GPT-3 help us develop a better understanding of LLM capabilities and potential areas of concern.

First, tests suggest that humans are not able to identify LLM-generated text. Developers asked GPT-3 to generate a news article based on a headline, and then asked people whether the article had been generated by a human

or a computer (Brown et al., 2020). Humans were only able to guess with 52% accuracy, which is barely better than random chance. While there are potential benefits to being able to perform so well, it is also concerning because LLMs could circulate false or damaging information that would be very difficult to trace (Better Language Models, 2019).

Second, LLMs demonstrate gender, racial, and religious bias. When asked to perform a variety of word association tasks, GPT-3 produced violent associations with Islam and negative associations with Black people, which reflects biases in the training data (Brown et al., 2020). And this problem may not be solvable: thus far, even though fine-tuning in general can produce very different outputs, efforts to remove toxic language about a marginalized group from an LLM means removing any mention of that group, even if it is positive (Welbl et al., 2021).

Overall, LLMs constitute a significant leap forward for Natural Language Processing, and the field continues to expand and advance rapidly. LLM developers may continue to compete primarily on the size of their models, or they may eventually shift to improving the quality. New uses and functions of LLMs, as well as new ways to capitalize on them, are emerging regularly, but the basic operation of the technology described here remains consistent. In the following sections, we discuss the implications of widespread growth and adoption of LLMs, both as the technology exists today, and as we expect it to advance.

PDF: Click to return to top

Section 1: Exacerbating Environmental Injustice

KEY POINTS

- LLMs will require the construction of more data centers, which will increase energy and water consumption.
- Data centers will displace and disrupt the lives of already marginalized communities.
- Those living near data centers will experience resource scarcity, higher utility prices, and pollution from backup diesel generators.
- Data centers will be classified as "critical infrastructure", which will make them more difficult to challenge. This will erode the civil rights of affected communities.

Although we tend to think of artificial intelligence as purely digital and "in the cloud", LLMs rely on physical data centers to process the corpora and run the algorithms (Bender & Gebru et al., 2021). They are, in other words, part of the LLM's "sociotechnical system", which includes not just the immediate technology but also its developers and users, and the other artifacts, institutions, relationships, and people that make an LLM work (Hughes, 1983). Data centers themselves are also sociotechnical systems, containing between fifty to eighty thousand servers that store and process data. These servers use graphic processing units (GPUs) that contain silicon chips as well as rare earth elements that are mined around the world (Johnson, 2017; Morgan, 2020). They are supported by ventilation systems, cooling systems, and backup generators.

Data centers rely very heavily on natural resources. They already make up approximately 2% of U.S. electricity use, and contributed to 0.5% of the country's total emissions in 2018 (Oberhaus, 2019; Siddik et al., 2021). They require large amounts of water to cool and power the servers, with a single medium-sized, high-density data center requiring 360,000 gallons of water a **day** (Ensmenger, 2018). For comparison, the city of Ann Arbor, Michigan, with a population of nearly 130,000 people, uses 5 billion gallons of water per **year** (The City

of Ann Arbor, n.d.). Not surprisingly then, data centers are already having significant impacts on the US water supply: they draw water from 90% of US watersheds, and 20% of data centers rely on watersheds that are moderately to highly stressed (Selsky and Valdes, 2021). Furthermore, much of the consumed water is potable, derived from local public utilities (Moss, 2021).

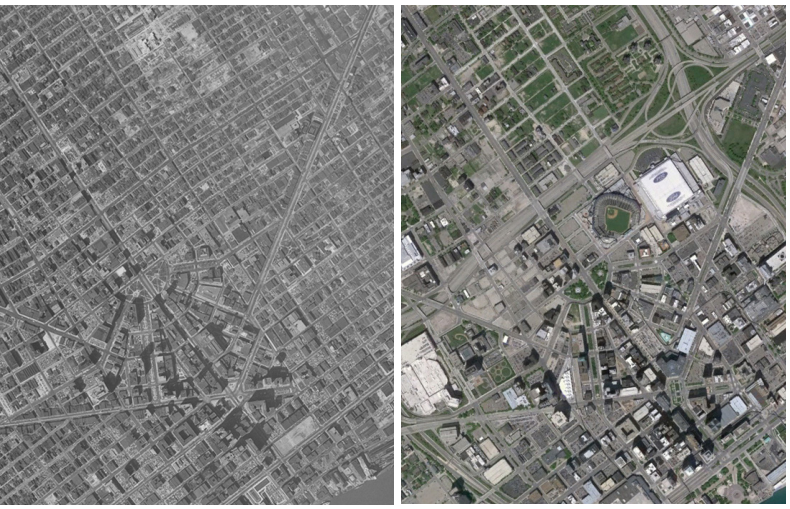
Despite this already-high consumption of resources, the current capacity of data centers is inadequate. To accommodate the rise of LLMs and other types of AI, tech companies will need many more of these large facilities. In fact, data centers owned by hyperscale providers like Amazon Web Services, Google Cloud, and Microsoft Azure have doubled from 2015 to 2020 (Haranas, 2021), and data center investment is projected to increase by 11.6% to $226 billion in 2022 (Haranas, 2022). They are typically 100,000 square feet in size, but the largest data center in the world (in China) is over 6 million square feet, and the largest data center in the United States is over 3 million square feet (Allen, 2018). The United States currently has the most data centers, with hubs around Washington D.C., New York City, Chicago, Los Angeles, San Francisco, and Dallas (Data Center Map, 2022; Berry, 2021); in Europe, there are hubs outside London, Amsterdam, Frankfurt, and Paris, and in Asia in Hong Kong, Mumbai, and Singapore (Data Center Map, 2022). Companies choose data center locations based on their power grids, labor markets, transportation networks, water access, and other social and geographical factors (Ensmenger, 2018). As a result, they have historically chosen more densely populated areas. But with the rise of distributed computing, rural areas–which are invariably cheaper–are becoming more attractive (Isberto, 2021). Microsoft is even experimenting with underwater data centers (Roach, 2020). As data centers increase, they will require additional infrastructure including roads and people but also a massive increase in natural resources.

As we noted in the Introduction, Emily Bender, Timnit Gebru, and their colleagues observed in their "Stochastic Parrots" paper that LLMs–due to their thirst for data and need for processing in data centers–would increase pressure on our energy systems and exacerbate climate change (Bender & Gebru et al., 2021). We agree with their conclusions, but expect that the environmental impacts of LLMs will be much greater and extend beyond climate change. In what follows, we rely on analogical case studies to suggest that the rise of data centers will disproportionately affect marginalized communities through displacement, direct harms, and curtailing their civil rights to protest.

Data Centers will Displace Marginalized Communities

In their quest to find the cheapest land available to build data centers, LLM developers will likely displace low-income and marginalized people in both urban and rural areas. This kind of displacement has a long history. In the middle of the twentieth century, city planners took advantage of the federally funded highway program to eliminate areas they saw as areas of urban "blight," a vague term that could suggest

congestion, property vacancy, vandalism, unkempt vegetation, graffiti, and more (Lopez, 2012; Stumpf, 2018; Mock, 2017). These tended to be Black and immigrant neighborhoods, which had experienced decades of disinvestment due to redlining (Semuels, 2016; Miller, 2018). In their place, city planners built highways that supported automotive transportation from and to the suburbs, which benefited wealthy, white, car commuters (Semuels, 2016). It also hurt marginalized communities by destroying their homes and businesses, physically dividing them, and polluting the local area (Lopez, 2012). This initiative displaced approximately 32,400 families per year in the early 1960s (Pritchett, 2003).



Detroit, 1951 (left) and 2010 (right)

This is only one of many examples of displacement in the service of technology. In order to build the Tucuruí Hydropower Complex, one of the world's largest hydroelectric dams, the Brazilian government displaced over 25,000 people (Downing, 2002). Local community members and activists who opposed the dam were murdered in the process (Environmental Justice Atlas, 2019). Mining projects in Honduras, Argentina, and Colombia, among many other countries, have also led to forced displacement (Working Group on Mining and Human Rights in Latin America, 2014). Mining projects have ruptured the social fabric of local communities in Chile and Mexico. Even if a technology's development does not trigger direct displacement, the economic ripple effects can. Fracking in the Willison Basin, in the northwestern United States, led to a sharp increase in housing prices which displaced longtime residents (Stangeland, 2016). Those with fixed incomes were at the greatest risk. Although tech companies may entice city leaders to accept data centers because they will bring jobs to an area and contribute tax revenue (Day, 2017; Glanz, 2012; Peterson, 2021), their mere construction is likely to interrupt neighborhoods and change mobility patterns. They will damage community cohesion and property prices, and ultimately increase socioeconomic inequalities.

Data Centers will Expose Marginalized Communities to Disparate Harms

Data centers will also subject marginalized communities to direct and disparate harms in two ways. First, already vulnerable individuals living near data centers will bear

the brunt of the negative effects directly. This is a common story. Although developers invariably claim that such facilities bring good jobs to the area (Day, 2017), and cities often provide tax breaks and other incentives, in the long term communities must manage a range of ill effects (Rayome, 2016; Fairchild & Weinrub, 2017). Power plants, oil and gas refineries, factories, and other toxic release sites have a long history of being located in communities that lack the political power to fight back, but must endure the consequences. Perhaps most notorious is Louisiana's "Cancer Alley", an 85-mile stretch of petrochemical plants and refineries where nearby residents are at a much higher risk of contracting cancer (Allen, 2003). Marginalized communities are subject to other kinds of risks as well. As Montana and North Dakota opened up their lands for oil extraction from the Bakken Formation, male laborers flooded the area (Stern, 2021). These new employees stressed the resources of economically fragile areas, and rates of human trafficking, sex trafficking, and missing and murdered Indigenous women rose (First Peoples Worldwide, 2019). Long legacies of discrimination, coupled with land-use, housing, and transportation policies, make it difficult for these communities to escape these "sacrifice zones" (Fairchild & Weinrub, 2017; Baker, 2019). Similarly, oil pipeline construction has caused the destruction of culturally significant sites like Native American burial grounds, inflicting significant harm on Indigenous communities (Whyte, 2017).

The process of extracting natural resources causes environmental degradation and resource scarcity in the immediate area. Companies mining lithium in Argentina and Chile worsened water shortages in the region (Frankel & Whoriskey, 2016). Mining practices themselves produce chemical residue, particularly of sulfuric acid, dissolved iron, copper, lead, and mercury, and this acid runoff can pollute both groundwater and surface water (U.S. Geological Survey, 2018). Similarly, pipeline construction and use cause negative health outcomes through the contamination of water sources and degrade ecosystems of plant and animal life (Betcher et al., 2019; Mall, 2021).

Communities... worry that data centers will stress their scarce resources and cause pollution.

Communities across the country have already begun to worry that data centers will stress their scarce resources and cause pollution. Google recently gained approval to develop several data centers in The Dalles, Oregon, a region experiencing severe drought (More Perfect Union, 2021). Citing trade secrets, the company refuses to disclose how much water its facilities will use, and local residents worry that the company's resource needs will be prioritized over their own. Google says that it will drill wells, build water mains and develop an aquifer to store water and increase supply during drier periods, but this could create additional risks to the community. Rural and

drought-prone Quincy, Washington, home to data centers for Microsoft, Yahoo, and Dell, has seen some of these problems. The town attracted Microsoft, for example, not only with the usual tax breaks but also by offering the company much lower electricity rates than the national average and promising to build a new substation (Glanz, 2012). When the company deemed the public utility too slow in building the substation, it began to waste millions of watts of electricity as a pressure tactic. Residents worry that behaviors like this will lead to a shortage of power and high prices, especially because Microsoft and Yahoo together used 41.8 million watts while all residential and small commercial accounts used only 9.5 million (Glanz, 2012). There are also concerns that Microsoft's 24 diesel generators, which the company uses for backup power for the data center, will create toxic air pollution that may cause cancer. Microsoft's Santa Clara, California, data center was one of the largest stationary diesel polluters in the Bay Area (Glanz, 2012).

Meanwhile, the increased need for data centers will continue, and perhaps even exacerbate, exploitation of the communities that mine minerals—including cobalt, tin, gold, copper, aluminum, tungsten, boron, tantalum, and palladium around the world (Euromines, 2020). These elements are required to construct the servers housed in data centers. Consider what is already happening in the Democratic Republic of Congo (DRC), which produces approximately 70% of the global cobalt supply (Sovacool, 2021). 20% of these producers are informal workers who have very low incomes and look for the mineral under hazardous conditions, known as artisanal and small-scale mining (ASM) (Buss et al., 2019; Lawson, 2021). These miners do not have access to adequate safety equipment, and experience significant negative health impacts due to the pollution of the mines (Amnesty International, 2020b). Furthermore, fatal accidents occur frequently (Al Jazeera, 2020). Wages are low, and miners are often subject to verbal, physical, and even sexual abuse (Sovacool, 2021), but they do not leave because cobalt mining provides wages in an area where there are few opportunities for employment. More data centers will mean increased demand for hardware like GPUs, which will only increase the need for these elements. This, in turn, will worsen the working conditions for these desperate miners as corrupt employers push them to increase their yields. Even though these mines clearly violate international standards, the major tech companies have avoided scrutiny or responsibility for their activities for years (Amnesty International, 2016).

Affected Communities will have Limited Civil Rights

Because of their central role in maintaining cloud computing and other services, we expect the US government to legally designate data centers as "critical infrastructure": physical or cyber systems that are seen as so essential to the country that their incapacitation would have significant negative effects on public health, safety, or the economy (Cybersecurity and Infrastructure Security Agency, n.d.). In 2021 for example, Australia classified

the "data storage and processing sector" (The Parliament of the Commonwealth of Australia, 2021), a category which includes data centers (Barbaschow, 2020; Hirst, 2020), as critical infrastructure. In the United States, 85% of this critical infrastructure is privately owned (Brown et al., 2017); these companies must collaborate and share information with the government in order to receive its protection (Monaghan & Walby, 2017). But while the critical infrastructure classification enhances the security of these facilities and increases the likelihood that they can maintain operations under adverse conditions, it also makes it more difficult for communities to protest against them.

Both governments and companies have used the critical infrastructure framing to surveil and persecute environmental activists (Monaghan & Walby, 2017). Multinational company Shell, for example, pumped oil from the Niger delta for decades. Oil spills were frequent, which led to extensive air, water, and soil pollution and ultimately damage to human and ecosystem health (United Nations Environment Programme, 2011). Communities began to protest systematically, but were invariably met by security forces who intimidated them, attacked them physically, damaged their property, arrested, and even murdered them (Frynas, 2001; Amnesty International, 2020a). In other words, the oil and gas projects didn't just hurt communities physically, they damaged their rights as citizens (Frynas, 2001). This kind of "security" is not unique to Nigeria. Both companies and governments have deployed security forces to protect oil infrastructure in Canada, Bangladesh, Mexico, Kenya, and Australia, and the United States (Savaresi & McVey, 2020).

The Royal Canadian Mounted Police used this framing to label protestors a mix of "peaceful activists, militants, and violent extremists," particularly emphasizing the "extremists," who are characterized as having an "anti-petroleum ideology" (Royal Canadian Mounted Police, 2014; Spice, 2018). Similarly, the Bangladesh government has suppressed freedoms of assembly and speech in response to protests against the Rampal Coal Power Plant (Savaresi, 2020).

In fact, in recent years over a dozen US states have passed "critical infrastructure protection" laws to criminalize anti-oil and gas pipeline protests. While they focus on violence or property damage, many worry that their true aim is to deter nonviolent civil disobedience that is protected by the First Amendment to the US Constitution (Colchete & Sen, 2020; Cagle, 2019). Consider protests over the Bayou Bridge pipeline, which moves oil between Texas and Louisiana (Cagle, 2019). Critics worry that the pipeline will leak and cause environmental damage, particularly in swamplands. In 2018, Energy Transfer, the company building the pipeline, directed "private duty" law enforcement officers from the Louisiana Department of Probation and Parole to arrest three protestors traveling by canoe and kayak who were observing and challenging pipeline construction. The company claimed "unlawful entry of a critical infrastructure," a recently enhanced felony under Louisiana law. The charges were later dropped, but similar cases are pending around the country (Baurick, 2020).

Whether or not data centers are formally designated as critical infrastructure, we expect the concerns of marginalized communities to hold less weight in siting decisions. As we discuss above, dangerous rare earth mining practices continue despite frequent community dissent (Business & Human Rights Resource Centre, 2021).

Whether or not data centers are formally designated as critical infrastructure, we expect the concerns of marginalized communities to hold less weight in siting decisions.

Similarly, Indigenous Americans have repeatedly raised concerns about developing sacred lands–whether for laying pipelines or constructing mines–with little success. And recently, a US federal court rejected attempts by the Paiute and Shoshone communities in Nevada to prevent the construction of a lithium mine on their ancestral lands, for example, largely because the US legal system does not recognize Indigeneous religious perspectives and by extension, cannot protect their sacred sites (Golden, 2021).

LLMs will place enormous pressure on current data processing capacity, which will trigger the development of data centers around the world. This will require not only the development of built infrastructure but also massive resource extraction. Our analogical case study analysis has suggested that already marginalized communities–both low income areas and communities of color–are likely to experience the negative impacts disproportionately. Many of them will be displaced, and their neighborhoods and towns transformed. Those who remain will have to manage new health and ecosystem risks, as well as economic burdens due to the data center's energy and water use. However, they will have limited opportunities to challenge this dynamic. City leaders will be enticed by the promise of jobs and regional economic development, and likely classify the new facilities as "critical". This designation will provide additional security, which will likely be used to curtail free speech and, ultimately, eliminate opposition.

Section 2: Accelerating the Thirst for Data

KEY POINTS

- LLMs will further test the model of individual informed consent which currently governs data sharing and privacy.
- Publics will increasingly hesitate to share personally identifiable information online, negatively impacting not only LLMs but other institutions that work with personal information.
- LLM developers may use unethical tactics to diversify the corpora, placing a disproportionate burden on marginalized communities.
- Users who formerly relied on human interpreters will feel that LLMs offer more privacy than relying on another person.

In addition to data centers and natural resources, LLMs require vast amounts of data. Much of it will come from us, the users of the internet. As we discussed in Background, LLMs already extract text from old books and across the web, including text from links posted on social media. In turn, this raises data security and privacy issues. While LLM developers have adopted some practices to filter out personally identifiable information (PII, which can include full name, social security number, zip code, and more) in LLM training corpora, such methods are neither effective nor commonplace (Privacy Considerations in Large Language Models, n.d.). This presents a serious vulnerability to third-party extraction attacks and unintentional leaks of PII. However, even if LLMs successfully screen out PII,

LLMs require vast amounts of data. Much of it will come from us, the users of the internet.

LLMs might still be able to triangulate bits of disconnected information such as mental health status or political opinions that appear in the corpora to develop a full, personalized picture of an actual individual, their family, or community (Kulkarni, 2021). Thus far, there has been little transparency as to whether the most popular LLMs have been security-tested, but the vulnerabilities are likely to increase as model development increases.

Meanwhile, Americans are increasingly concerned about data security: 79% of adults worry that companies are using their personal information and 64% are worried about government data collection (Auxier et al., 2019). These concerns are valid as data security is a challenge: while Illinois and California have passed data privacy laws, the United States lacks federal legislation and much of the population remains unprotected by data privacy or security policies.

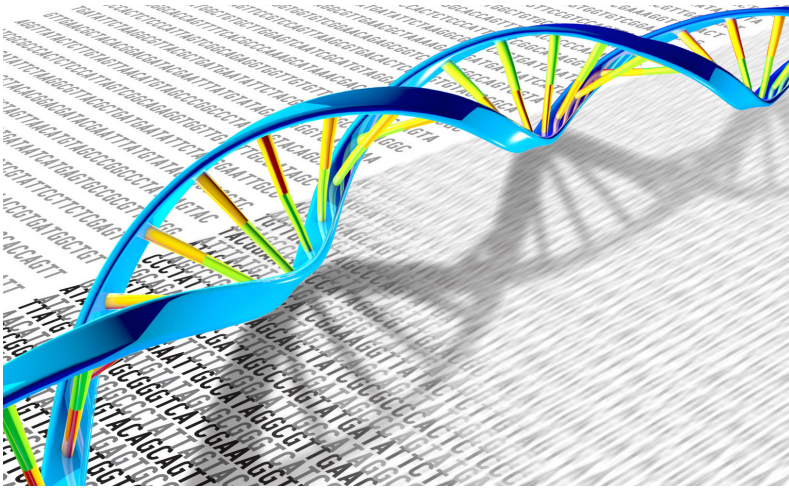
In this Section, we analyze how LLMs will affect the privacy and security of personal information and accelerate a thirst for data. We conclude that LLMs will likely be able to produce information about individuals and communities even if they are barred from including personally identifiable information (PII). As a result, publics will become more hesitant to share information about themselves online. These information practices will have uneven impacts for marginalized groups: those who are underrepresented in the corpora are likely to be pressured to participate in LLMs and may lose some civil liberties if they do not. But others, including those who currently rely on interpreters or translators to communicate and travel (e.g., those who are hearing impaired) may actually be able to better maintain certain forms of privacy.

LLMs will Transform Informed Consent

Most LLM corpora are created using a data collection method called web crawling, which involves systematically traversing the entire internet to gather text. Much of this text was provided by the population through their online activity when they upload web pages or post comments. But few of us have any idea that our text is included in the corpora, much less which information is used or how it is used. In many cases, we may have already provided our consent. We agree to complex and lengthy "click through" agreements to use online services, such as WordPress or Reddit, that allow third parties to have access to the text we post, including for LLMs. This is problematic because few people read user agreements and are therefore unaware of the scope of their consent (Cakebread, 2017). And LLMs pose particular challenges. As noted above, the text we post can be triangulated to develop a full picture of us or to predict our behaviors. If we post information about our community or family, we are also consenting to data collection on behalf of others without their knowledge. In sum, while some of us may consent to the use or sale of some of the information we post, LLMs bring it all together and expand the scope. This makes the information more powerful and has potentially serious implications even for those who are careful about what text they post online.

The field of genomics has been dealing with these kinds of challenges for decades. With the rise of mapping and sequencing technologies and the infrastructure to build and process large databases of information about an individual's genome as well as their health, environment, and lifestyle, there has been growing concern about individual, family, and community privacy. An individual's decision to get a genetic test has ripple effects for their family members. If someone tests positive for a gene mutation related to Huntington's Disease, for example, then not only will their family members feel additional stress or anxiety that their loved one will soon experience a debilitating neurodegenerative condition, but their children will also be more concerned that they too will have the disease (Oliveri et al., 2018). However, these individuals have no say in whether their family member gets a genetic test. Similarly, when a handful of members of a racial group or ethnic community choose to participate in genomics research, all members of the group are all affected by the findings. In the 1970s, the US federal and state governments created screening programs to identify African Americans at risk of sickle cell disease, a painful blood disorder, and ensure that they receive appropriate services (Duster, 1990). Unfortunately, the program resulted in stigmatization and employment exclusion based on race; the US Air Force Academy, for example, erroneously used the data to exclude sickle cell carriers from the applicant pool.

An individual's participation in a DNA database can also have broad criminal justice implications. Law enforcement agencies across the world have created forensic databases that include the DNA information of all individuals convicted of (and sometimes, even arrested for) a crime (National Conference on State Legislatures, 2014; Interpol, 2020). When they find DNA at a crime scene, they then search these databases to find not only matches but also "familial matches", i.e., people whose DNA partially matches the DNA found at a crime scene. This helps police officers narrow down the pool of potential suspects and focus on



Credit: Darryl Leja, NHGRI

a specific family. But, it also means that individuals who never agreed to participate in the database are affected by its findings. This took place in the infamous Golden State Killer case, where investigators identified the killer via the genetic profiles of distant relatives dating back to the 1800s. Clearly, these relatives never consented to upload their genetic profiles, but there is no option to opt-out (Zabel, 2019). Some might argue that this erosion of privacy is permissible in the name of public safety, but studies show that these

databases have a disproportionate number of samples from historically overpoliced communities of color and are thus more likely to affect them (Murphy & Tong, 2019). We have also begun to see similar use of health or ancestry-focused DNA services, such as 23andMe. Even though individuals provide their DNA in a non-criminal context, the services have demonstrated a willingness to share this information with law enforcement. For example, researchers investigating the Y-chromosome Haplotype Reference (YHRD) forensic database, which contains 300,000 anonymous genetic profiles, have raised ethical concerns over a lack of informed consent for the Uyghur and Roma populations. Without knowledge of where their genetic information will go, these minority ethnic groups stand at an increased risk for persecution (Schiermeier, 2021). In all of these cases, the decisions of a few people to share information had widespread impacts.

In almost all these cases, individuals provided free and individual informed consent, a framework developed in the latter 20th century in response to scandals about unethical medical experimentation and practice. But this framework is clearly insufficient for situations where one person may share information that has implications for others. In response, researchers have pioneered new approaches that take human connection seriously. Consider, for example, the Human Genome Diversity Project initiated in the 1990s. Excited about the opportunity to use new techniques to map genomic diversity, scientists identified populations around the world and began to ask for their DNA using well-established consent procedures (Reardon, 2005). Many of these communities were quite isolated, and perturbed by the Western scientists making these requests. They were also distressed by the concept of individual, informed consent. After all, one person's DNA would provide information about the whole community. What if others in the community questioned the use of this information? In response, some of the scientists proposed a new approach that would include informed consent from both individuals and the group through "culturally appropriate authorities" (North American Regional Committee, 1997). Later attempts to map human genomic diversity, including the International Haplotype Mapping Project, have tried to implement this approach (The International HapMap Consortium, 2004).

Another similar framework involves engaging community members throughout the life of a project. Researchers from the University of Oklahoma followed this model when conducting genomic research with Native American populations. They surveyed participants about health decision-making, explained intentions behind the project during public meetings, and established community review boards to review manuscripts. While this approach can slow down research and is often logistically challenging, it ensures public trust and effective guidelines can help minimize negative effects (Foster et al., 1997). The way researchers obtain consent can have long term impacts particularly if the researchers want to return to the community for further data collection. For example, in the case of the Havasuppai in Arizona, the tribe provided their DNA to Arizona State University researchers with the understanding that they would use it for diabetes research (Harmon,

2010). However, the researchers gave the DNA to another group of scientists who used it to investigate the tribe's genetic origins and whether members of the tribe had genetic links to mental illness. When the tribe discovered this misuse, they sued. Eventually, they settled the case, but not before damaging the trust between the university and the tribe.

LLMs will raise similar concerns and controversy, as people realize that their text is being used for purposes that they did not intend or with which they disagree. If developers do not address these concerns at the outset, they risk further erosion of trust in the tech industry, and ultimately, resistance as we discuss in Section 6. However, they can learn from the genomics and medical arenas,

If developers do not address these concerns at the outset, they risk further erosion of trust in the tech industry, and ultimately, resistance.

which have been experimenting with new forms of consent. This includes both group consent, described above, as well as qualified granular consent (Simon et al., 2011) which is designed to provide users with more authority over how their data is used.

People Will Hesitate to Share Personal Information

We expect that as people interact with LLMs and realize the depth and breadth of the data they are trained on as well as their potential to disclose sensitive information, they will hesitate to provide personal information online.

In 2017, Equifax, an American credit reporting agency, suffered a cyber attack that uncovered and downloaded sensitive PII of over 140 million customers. Negligent Equifax security officials were at fault because they failed to install a security patch from their software provider, Apache Struts, which had been released two months before (Baird Equity Research, 2017). When it disclosed the attack a few months later, Equifax lost significant credibility – shares dropped 13% in early trading, people were outraged over the lack of transparency about the attack, and hundreds sued the agency for damages and won. After these sanctions, 54.2% of those publicly surveyed believed that Equifax should no longer serve as a credit bureau (Brown, 2018). One year later, Equifax attempted to remedy the issue by providing consumers with free credit monitoring but only 30.6% said that these steps had improved their perception of the company.

As a result of data breaches, companies across the globe have improved their security

standards and data governance practices – but keeping PII private continues to be a challenge. In just the first quarter of 2021, 4 billion online accounts were hacked worldwide, with LinkedIn and Facebook being the most vulnerable (C., 2022). Occasionally, the party misusing the data is the one collecting the data itself. For example, employees of the ride-sharing app Uber used the company's database to track the locations of politicians, celebrities, and even ex-spouses. They exploited this "God View" feature for over 2 years with little to no user knowledge (Evans, 2016). The cumulative effect is one of user mistrust across companies, and a feeling that the onus is on the user to take preventative measures.

But the effects go beyond loss of trust. User behavior often changes dramatically. Consider recent concerns over email trackers in the most popular email clients (Google's Gmail, Microsoft's Outlook, Yahoo Mail), which enable third-parties to extract a user's email address and activity on a user's web browser. With this information linked together, the third party trackers can target ads based on any future online activity across all devices. The practice is widespread - an estimated 70% of emails embed at least one tracker (Englehardt et al., 2018). In response, users are flocking to a less-established platform that prioritizes security, DuckDuckGo. DuckDuckGo's Email Protection feature strips emails of trackers, sets up a disposable email to forward spam, and prevents disclosure of personal information (Gershgorn, 2021).

LLMs present similar risks, especially because the training dataset is large and, in some cases, contains text from private sources. Hackers could extract specific parts of training data that the LLM has memorized, known as a training data extraction attack. An adversary with access to an LLM would simply have to input probable phrases (e.g. "The phone number of John Doe is"...) and let the model complete information that might reveal PII. Using confidential data to train the LLM is dangerous, as it risks revealing information that users intended to keep secret. This technique has already been put into practice, as Gmail's auto-complete model is trained on private text communication between users (Privacy Considerations in Large Language Models, n.d.). We expect that public-facing LLMs will in part use confidential data for training, which means personal data breaches will be possible. But not all breaches of privacy will rely on private data and sensitive PII. Even an LLM that connects a person's professional online presence with their personal one could have implications if their online presence includes information about things like health, sexual orientation, or immigration status.

As a result, users will lose trust and ultimately hesitate to provide personal information online and in other communication channels. As we discuss in Section 6, this breakdown in trust will have implications for social fragmentation. This could also hurt the accuracy of LLMs and the development of new apps. But it will also hurt institutions that require access to PII to function (e.g., hospitals, banks) as well as the individuals who rely on them and on accurate digital technologies. Users are likely to hesitate to give PII and may create new ways to stay anonymous online, such as tools that prevent

web scraping, although this may not be accessible to or benefit everyone (Zou et al., 2018).

LLMs will Create New Forms of Data Exploitation

With the rise of surveillance capitalism (Zuboff, 2019), all digital data has increased in value. LLMs exemplify this trend as they take advantage of the freely generated data of millions of individuals to produce commercial technologies that summarize, generate, and predict language. But in this ecosystem, not all data has the same value. At present, LLM corpora overwhelmingly include English or Chinese language texts, and many of these texts are quite old. The racist, sexist, and homophobic output described in the Introduction is one result. In order to ensure that LLMs are more useful and less offensive, developers are keen to expand the corpora to include more languages and dialects, genders, cultures, and populations. History suggests that the texts least likely to be currently represented in the corpora, and thus most likely to be valuable in the future, come from marginalized communities and cultures. But as developers try to improve their models by expanding corpora in these directions, they will create new forms of exploitation that will disproportionately affect already marginalized communities.

Facial recognition technologies have posed similar problems. They are famously inaccurate among all populations–including people of color, women, children, gender non-confirming people–except white male adults (Grother et al., 2019). But they are increasingly being used by law enforcement, schools, airlines, and even, briefly, the Internal Revenue Service (Epstein et al., 2022; Galligan et al., 2020). To deal with the technology's accuracy problems, developers have sought out pictures of individuals from marginalized communities. Most famously and problematically, a contractor hired by Google targeted attendees of the BET Awards, college students of color, and even homeless Black people in Atlanta for facial scans. The practice was exploitative, as volunteers were rushed through consent forms and misled about what would be done with the scans (Dillon, 2019).

Similarly, as we suggest above, the rise of genomic science has also made particular genomes valuable. This has, in turn, triggered unethical practices and created new burdens for already marginalized communities. A 2019 controversy at the UK's Sanger Centre, the UK's premier genomics institute, echoes the Havasuppai and Human Genome Diversity Project cases described above (Stokstad, 2019). The Centre was trying to develop and commercialize a "gene chip" that would identify genetic links to common diseases, and needed African DNA samples in order to ensure that this technology was adequately representative. So it entered into agreements with scientific institutes in Africa that had collected indigenous DNA. But it did not disclose that the DNA would be used commercially, and many of the original DNA sharing agreements had forbidden this kind of use. The African scientists who

collected the data worried that it would alienate communities who had just begun to participate in genomics research (AT Editor, 2019).

Some communities have learned how to take advantage of the importance of their own data for emerging technologies.

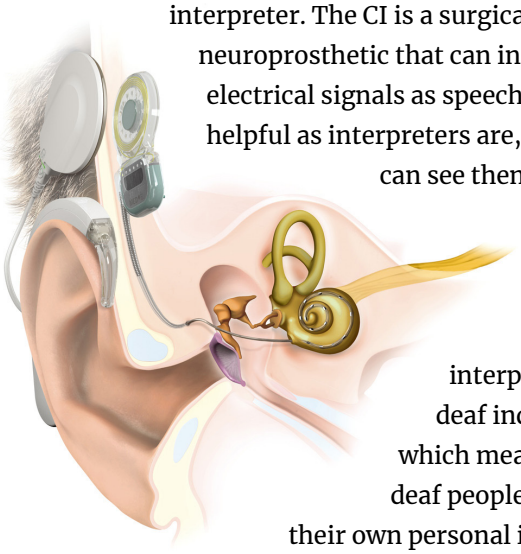
However, some communities have learned how to take advantage of the importance of their data for emerging technologies. Groups representing individuals with rare diseases have negotiated with scientists to own technologies produced using their and their childrens' DNA (Terry et al., 2007). Indigenous communities have developed benefit-sharing agreements with Western companies seeking to commercialize their knowledge (Foster, 2018). Other groups seek to keep the economic benefits to themselves, such as Te Hiku Media, an Indigenous-owned tech nonprofit that has refused to share hundreds of hours of valuable Maori language audio. Instead, by building speech recognition technology internally, Te Hiku has ensured that only the Maori people will control the use of and profit from their language (Coffey, 2021). Similarly, recognizing that non-Black creators reap financial benefits from co-opting their dances, Black TikTok stars boycotted the platform in hopes of receiving proper credit and compensation (Muller, 2021).

In the coming years, LLM developers are likely to prioritize collecting texts from marginalized communities in the name of increasing accuracy. This might mean purchasing access to non-digitized texts in a variety of languages, or deploying speech-to-text apps to capture the rare dialects of some communities. But given the power of these companies, there is a great risk of exploitation. Marginalized communities will need assistance from NGOs and the government in order to ensure that any data-sharing agreements are appropriately balanced and serve community needs.

LLMs create a sense of privacy for some vulnerable communities

Some vulnerable communities, including immigrants who do not speak the dominant language, and people with auditory disabilities, rely on human interpreters to access social services from healthcare to legal aid. Currently, they have to disclose personal, potentially embarrassing information to another person, and trust them to protect it. But in the future the user could interact with an LLM or other technology that feels more private. However, they would still be sharing personal information with the company providing the LLM-based service.

For the deaf community, the cochlear implant (CI) has enhanced privacy by eliminating the need for a human sign language



Credit: WikiCommons user Hear Hear! (CC BY 4.0)

interpreter. The CI is a surgically implanted neuroprosthetic that can interpret electrical signals as speech and sound. As helpful as interpreters are, deaf people can see them as a burden (Reinhardt, 2015). First, there are far fewer interpreters than deaf individuals, which means that few deaf people can afford their own personal interpreter. As a result, they must disclose personal information to multiple interpreters, which increases the risk that interpreters might misuse the information. Interpreters might be familiar with one another and share information about the deaf person, or the interpreter may simply not be fully fluent in sign language which makes it impossible to establish trust from the beginning. Interpreters often report feeling anxious when having to translate serious discussions, such as marriage therapy (Levinger, 2020). Deaf people must trust that the interpreter will adhere to their professional obligations and maintain their privacy. With a CI, however, this trust is unnecessary: the human intermediary is displaced along with privacy concerns (The British Psychological Society, 2017). CIs cannot store information and thus confidential information remains between the deaf individual and the intended party.

Similarly, the elderly population living in nursing homes are often heavily surveilled in order to provide proper assistance. Nursing home staff are informed about medical conditions as well as medication requirements, both of which may be seen as intrusive. They and other staff also use cameras to surveil residents to identify those who are in distress and provide appropriate assistance. However, this data can be easily misused: staff might use personal information to impersonate residents, or to prey on them (Berridge & Levy, 2019). To avoid this potential invasion of privacy, elderly individuals replace nursing home personnel with in-home technologies including virtual assistants, stair lifts, and telemedicine applications (Kelly, 2021). At the very least, these technologies delay the need for an older person to move to a nursing home or assisted living facility. For the blind population, sighted guides similarly share, through vocal or physical cues, information about orientation and navigation when traveling in unfamiliar areas. This help can be crucial. However, as with previous examples, the cost of this assistance is the disclosure of personal information which can, again, be easily misused (Merry-Noel, 2015). Blind people might use guide dogs, braille, or white canes instead, to maintain some privacy.

LLMs offer a similar sense of privacy for some communities. As we describe in other sections, LLMs may be able to translate text across languages and linguistic styles, making the world more accessible for those particularly for those who do not primarily use spoken English. While the interpersonal dimension of human communication might be lost, the case studies we have reviewed here suggest that this will produce privacy benefits. Without a third party translator, the communication experience will be less intimidating especially for those disclosing

embarrassing personal information to the intended recipient. However, it is important to note that even in this case, the user is disclosing information to the LLM or LLM-based app, which could incorporate this information into a corpus that may only be partially protected. Thus, there is still a privacy risk, but it may be indirect and attenuated in comparison to the benefit.

At their core, LLMs rely on data about the way humans communicate through language and thus, LLM developers will continually need data to maintain their model's accuracy. However, under current conditions, communities are likely to become increasingly reticent to share their information. In the long run this could affect our health care, finances, security, and even rights. It could also produce controversies that damage trust in LLMs. Meanwhile, LLM developers are likely to use highly unethical practices to extract data, especially from marginalized populations, in the name of enhancing accuracy in their models. Finally, the communities who have traditionally relied on assistive technology may gain some immediate privacy but will now be disclosing their personal information to an LLM.

Section 3: Normalizing LLMs

KEY POINTS

- To gain user acceptance, LLMs will be framed as empowering and modular.
- Developers will try to incorporate LLMs into existing sociotechnical systems, particularly those governed by trusted institutions, in order to ensure their longevity.
- When LLMs produce hateful language or errors, developers will deflect blame onto infrastructure or human users.

Developers have made experimental LLMs available to both researchers and publics, which has stimulated early excitement about the technology's text summarization, generation, and translation capabilities. And yet, as we have noted repeatedly thus far, there is already concern about the social harms. Journalists worry that LLMs will be able to generate articles and further damage their job prospects (Seabrook, 2019). Others worry that like cryptocurrency, LLMs will require the use of so much energy that they will make it harder to fight climate change (Bender & Gebru et al., 2021). There is already evidence that LLMs will reflect the historical biases of English-language texts by using racist and sexist language and reproducing harmful assumptions about marginalized communities (Abid et al., 2021).

But these conversations have been largely restricted to the field of artificial intelligence and specialist technology publications. An important exception is the 2021 controversy over Google's firing of Timnit Gebru and Margaret Mitchell, which we discuss in the Introduction. Newspapers across the globe covered Gebru's firing, likely due to combined concerns about the practices of the major technology companies, artificial intelligence and algorithmic bias, and heightened media attention to discrimination against Black people in the wake of George Floyd's murder in June 2020.

These events have likely triggered some skepticism about LLMs, most significantly within Black and research communities. Despite these concerns, based on our

analysis of analogical cases we expect Google and other developers to emphasize LLMs' democratizing and even empowering potential, as well as their modularity. In fact, they have already begun to highlight the broad social benefits particularly in terms of increased access to crucial services such as legal aid and health advice. They will also try to make LLMs ubiquitous quickly, and promote their use particularly among established authoritative institutions. In the process, they will continue to dismiss the technology's limitations and errors, deflecting blame onto infrastructure and other users.

LLMs as an empowering technology

In keeping with a long history of technology, particularly those focused on communication, developers will emphasize LLMs' capacity to empower their users. The One Laptop Per Child (OLPC) program is an instructive analog. Founded in 2005 by Nicholas Negroponte, founder and chairman of the MIT Media Lab, OLPC aimed to transform education for children around the world by providing them with extremely cheap (~$200), rugged computers, and accompanying software and content (Ames, 2019). To gain support, Negroponte presented his ideas and solicited investments across the world, including at the World Economic Forum in Davos, Switzerland. He claimed that the technology would allow children to teach themselves and their parents, providing them both with an education that would allow them to lift themselves out of poverty.

In the project's early days he said to the *MIT Technology Review* that OLPC "is probably the only hope. I don't want to place too much on OLPC, but if I really had to look at how to eliminate poverty, create peace, and work on the environment, I can't think of a better way to do it" (Ames, 2019). As Negroponte and his team tried to sell the technology first to investors and then to the governments of Southern countries, he framed it as not just transformational but as leveling the playing field across the world.

Although OLPC was explicitly designed for humanitarian purposes, we expect LLMs to be framed in similar ways. Consulting firm Deloitte has already suggested that LLMs will be able to more efficiently and accurately synthesize public comments on pending policies (Eggers et al., 2019). Others have emphasized that the technology could provide legal aid to those who could not otherwise afford it (Bommasani et al., 2021). We might even expect developers to encourage the first apps on "public interest" oriented technologies, such as therapy chatbots. Despite emerging concerns about LLMs, we do not expect corporate developers to voluntarily take steps to build public trust by making the corpora or algorithms transparent or bringing in community knowledge to develop more politically legitimate technologies. While technologists have begun to take such steps in highly controversial areas such as geoengineering and human gene editing (Stilgoe et al., 2013; Gusmano et al., 2021), LLMs have not yet risen to that level of public attention or scrutiny.

Creators of technologies also sometimes emphasize the empowering potential of their machines by allowing them to have "interpretive flexibility" particularly in their initial rollout. Rather than dictating use, these developers allow users to integrate technologies into their current work and lives however they wish, to increase uptake and excitement (Bijker et al., 1987). Early car manufacturers used this approach to increase acceptance in the early 20th century (Kline & Pinch,1996). Farmers were initially very skeptical of the automobile, which was loud and scared away livestock, and made it difficult for them to use their horse-drawn buggies on the roads. The technology also brought urbanites into their towns, whom rural residents found irritating and sometimes even scary. So, farmers used a variety of strategies to keep out what many called the "devil wagon", and an anti-car movement began to flourish. However, some farmers reinterpreted the technology as a source of power, and demonstrated how its engine could be used to facilitate farm tasks including corn shelling, sheep shearing, and grinding. Soon, manufacturers made changes to new car models, developed new accessories, and changed their advertising strategy to capture this understanding of the automobile (Kline & Pinch,1996). They knew that by endorsing these interpretations of their technology, they could increase demand and ultimately entrench it in American life. LLM developers, by encouraging targeted apps designed for a range of purposes, are already starting to construct a technological ecosystem geared towards this kind of flexibility.

Connecting to Authoritative Institutions

LLM developers will also try to establish the legitimacy of the technology by quickly integrating them into the existing infrastructure, including connecting it to authoritative institutions. Facial recognition technologies have followed this path. First used for security on a large scale at the 2001 Super Bowl, when law enforcement used it to detect potential threats among the crowds, facial recognition has spread

rapidly particularly over the last 10 years with almost no regulation (Galligan et al., 2020). Security companies convinced police, universities, K-12 schools, and airlines across the United States and around the world to adopt the technology in the name of public safety, even in the face of growing evidence that is inaccurate among marginalized communities and often ineffective. With law enforcement and academia as early adopters

PDF: Click to return to top

and advocates, the technology has become harder to challenge. While civil society groups and even policymakers have tried to ban or otherwise regulate the technology, they have met limited local success. And as a result, facial recognition's reach is growing: in 2022, Clearview AI, with one of the largest indexes of faces, announced massive expansion of their services beyond law enforcement (Harwell, 2022). The situation is similar with the breathalyzer, which is used to evaluate cognitive impairment due to alcohol (Cowley & Silver-Greenberg, 2019). Despite extensive evidence that it generates inaccurate results, it is still widely used by law enforcement.

This technological entrenchment is not unique to law enforcement. Consider the pulse oximeter, which assesses blood oxygen levels, and is crucial to diagnosing severe cases of COVID-19. In 2020, an anthropologist published an article observing that the device was likely to be less accurate among people of color because its reading is based on light refraction (Moran-Thomas, 2020; Sjoding et al., 2020). A few months later, a group of physician scientists validated this hypothesis through a randomized controlled trial: they found that people with darker skin tones tended to have higher readings than their white counterparts. This means that when already marginalized people of color went to the hospital unable to breathe, a pulse oximeter reading might suggest to the health care professional that they were not in distress. Likely as a result, in the early days of the COVID-19 pandemic Black patients were turned away from hospitals because their blood oxygen was not low enough (Lothian-McLean 2020; Rahman 2020). *The New York Times* and other prominent media outlets published these scientists' findings (Rabin, 2020; Harris, 2020). But today, doctor's offices and hospitals still regularly use the pulse oximeter as part of their health care, to determine the severity of their patient's condition. It seems too difficult to change professional practices, despite the human cost.

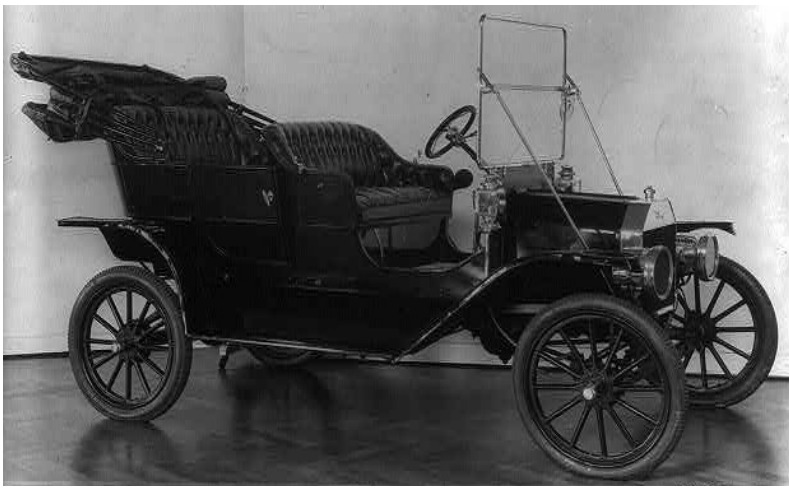
Deflecting Blame for the Technology's Problems

Especially in the early days of LLMs, we expect users to identify a range of errors and problems with the technology. Developers will first try to maintain the technology's credibility by ignoring these problems. If that proves impossible, they will likely blame the infrastructure or users. Let's return to the OLPC. It never had the positive impacts that Negroponte envisioned. Demand for the machine was less than anticipated, and even when governments or civil society groups donated them to low-income children, many broke or simply went unused. And yet, OLPC's developers largely do not acknowledge the technology's failure (Ames, 2019). When they do, they suggest that the technology simply lacked the needed support structure.

Or, consider Boeing's introduction of the Maneuvering Characteristics Augmentation System (MCAS) system in its 737 MAX planes, and the subsequent crashes of two planes in Indonesia and Ethiopia in 2018 and 2019. 346 passengers died in total. Boeing had installed the technology in its planes without alerting regulators in the US or elsewhere. But when the first plane crashed in October 2018, the

company denied responsibility. It responded that the plane was "as safe as any airplane that has ever flown the skies", and pointed instead to human error (Robison, 2021). When observers began to question its MCAS system, the company persisted, arguing that the pilots should have known how to handle the emergency. Taking advantage of age-old Western prejudices towards people in low and middle-income countries, it suggested that the Indonesian pilots had insufficient expertise. If they had followed the established emergency procedures, Boeing argued, pilots would have been able to reverse the plane's downward spiral and keep the vehicle aloft (Glanz et al., 2019). In November 2018, the company explicitly advised pilots to take corrective action if the MCAS system engaged. But it was only after a second plane crashed in Ethiopia in March 2019 that governments around the world grounded Boeing's 737 MAX planes. Boeing changed the aircraft design in response, and in December 2020 the US Federal Aviation Administration allowed the planes to fly again. Other countries quickly followed. Until governments stepped in, Boeing kept deflecting blame for their technology.

Similarly, since the automobile's earliest incarnation, automakers have refused to address known safety concerns or deploy safety features until forced by legislation (Singer, 2022). For example, Hugh DeHaven, a former pilot who survived a deadly airplane crash and then spent decades trying to improve airplane and automobile safety, collected reports from hundreds of crashes to identify the most dangerous parts of a car: rigid steering columns that did not collapse on impact, unpadded dashboards, pointed



Credit: Library of Congress (CCo)

knobs, and a lack of seatbelts. He presented this information to the auto industry at a conference in 1953, along with remedies that would improve "crashworthiness" including a collapsible steering column and a 3 point seatbelt. However, the industry insisted that the problem was not with their products, but rather with "the nut behind the wheel": "reckless" drivers who were the cause of deadly crashes. And they were successful for a time. It took external pressure—Ralph Nader's 1965 book *Unsafe at Any Speed* (Nader, 1965) and the resulting outcry—to trigger regulatory action and ultimately changes to the technology's design. But in the years between DeHaven's presentation and the publication of *Unsafe at Any Speed*, over half a million people had died in car crashes.

Right now, LLMs are unknown to many except for the few who paid attention to the controversies over Timnit Gebru and Margaret Mitchells' firings. Given this, we expect LLM and app developers to try to shape early public opinion about the technology

both by emphasizing its empowering potential and humanitarian benefits, and by encouraging flexible interpretations of its use. They will also try to integrate it into existing sociotechnical systems and infrastructure, particularly those that enjoy high public trust. As the technology becomes ubiquitous, we expect that developers will be able to dismiss any problems or errors and avoid real public or policy scrutiny–unless there is a catastrophic outcome. In the meantime, the costs will likely be borne by users, particularly those who are marginalized as we discuss in the next section.

As LLMs become ubiquitous, we expect that developers will be able to dismiss any problems or errors and avoid real public or policy scrutiny–unless there is a catastrophic outcome. In the meantime, the costs will likely be borne by users, particularly those who are marginalized.

Section 4: Reinforcing Social Inequalities

KEY POINTS

- LLMs will exacerbate the inequalities faced by marginalized communities.
- Individuals, particularly those from already marginalized communities, will bear the blame for LLM error, rather than the developers or the technology itself.
- LLMs will reinforce English language, and ultimately Anglo-American, dominance, and alienate those outside these cultures.
- Because they seem technical and objective, LLMs will obscure systemic biases embedded in their design. This will make inequities even harder to identify.

In this section, we shift from the construction of LLMs to the implications of their use. AI developers tend to emphasize the objectivity of their technologies, but scholars point out how technologies always reflect the societies that make them (Benjamin, 2019; Parthasarathy, 2007). The history of technology provides numerous examples of tools and systems that are presented as value-free and yet are skewed by built-in biases including racism, sexism, and xenophobia. The same is true for LLMs. As described in the Introduction, they are trained on vast datasets composed of internet text and historic literature; both contain enormous amounts of prejudiced and hateful language towards minoritized groups. In other words, they reflect historical biases. Not surprisingly, then, LLMs that generate "new" content end up reproducing these biases often in the form of violent language (Abid et al., 2021; Tamkin et al., 2021). But fixing these problems isn't just a matter of including more, better data. LLMs are built and maintained by humans who bring prejudices and biases to their work, and who operate within institutions, in social and political contexts. This will shape the biases that developers perceive, and how they choose to fix them. Meanwhile, researchers have already brought attention to how artificial intelligence is exacerbating what they call a

"compute divide": wealthy companies and academic institutions have greater resources to invest in emerging technologies, which is likely to reflect their worldviews, needs, and biases (Ahmed & Wahed, 2020).

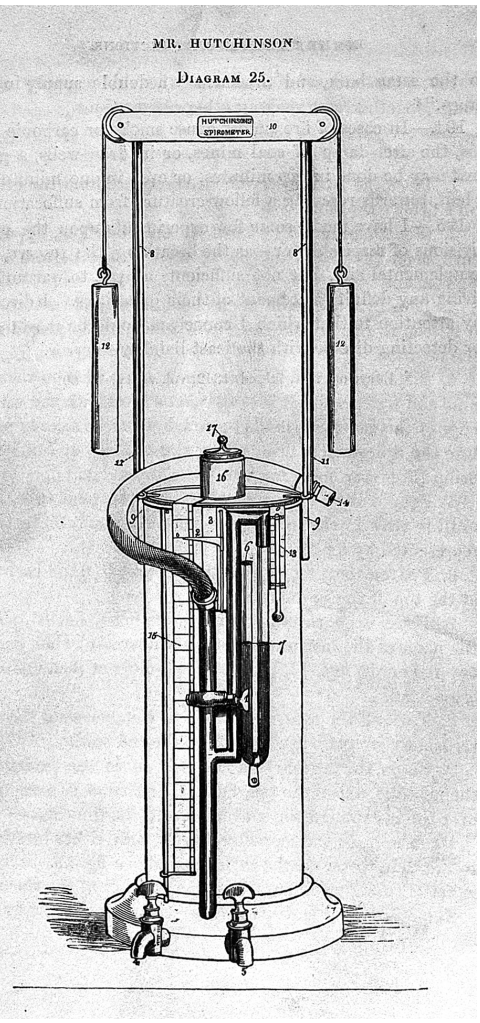
In what follows, we suggest that LLMs are likely to reproduce social biases in a variety of ways beyond what observers have already identified. Trained on texts that have marginalized the experiences and knowledge of certain groups and were produced by a small set of technology companies primarily in the United States and China, LLMs are likely to systematically misconstrue, minimize, and misrepresent the voices of some groups, while amplifying the perspectives of the already powerful. In addition to producing language that contains racist, sexist, xenophobic tropes, they may fail to include representations of minoritized groups altogether. These implications are particularly problematic for two reasons. First, LLMs have a wide range of possible uses across fields, so there is broad potential to replicate and perpetuate racism and other biases. Second, people are likely to assume that because they are based on vast amounts of data and produced by highly technical, proprietary algorithms, they will be objective. Therefore, these biases will be harder to identify and challenge. This section identifies these biases at a societal level; other sections discuss how LLMs will affect equity in the workplace (Section 5) and environmental justice (Section 1).

Trained on texts that have marginalized the experiences and knowledge of certain groups and produced by a small set of technology companies primarily in the United States and China, LLMs are likely to systematically misconstrue, minimize, and misrepresent the voices of some groups, while amplifying the perspectives of the already powerful.

LLMs will Perpetuate Inequitable Distribution of Resources

LLMs will reinforce the inequitable distribution of resources, continuing to favor those who are privileged over people who need aid the most. The racism embedded in the very design of many technologies including common medical diagnostic tools has had similar impacts. Consider the spirometer, a device widely used to measure the volume of air inspired and expired by the lungs. It is used to diagnose diseases such as asthma and emphysema and to identify the cause of shortness of breath, including environmental contamination. Patients breathe into a tube, and the machine both measures lung function and assesses whether it is "normal" using software.

However, these assessments differ by race, based on false beliefs that race affects lung function, and that Black people naturally have lower lung function than white people (Braun et al., 2013).



Spirometer diagram. Credit: Wellcome Library, London

The racist belief that Black people have lower lung function can be traced back to slavery. White slavers used early versions of spirometry to "prove" that Black people were physically inferior; they found that enslaved Black people had lower lung function than free white people, without accounting for the many factors—such as the effects of extreme physical labor and abuse—that could contribute to such a disparity (Braun, 2021). These assumptions were then embedded into assessments of "normal" lung function, first in tables and then in the software of the spirometer. Today, most health care professionals operating spirometers have no idea that assessments of normal and abnormal lung function are based on this racist science, instead viewing it as an objective measurement. And despite numerous scholarly publications describing this bias, there have been no changes in medical diagnosis. Still today, lower lung function in a Black person is often considered normal and does not trigger further action.

The bias built—and perpetuated—in spirometry machines means that Black patients need to be sicker, with more severe illness, in order to qualify for many treatments and insurance coverage. For example in 1999, employees of an insulation manufacturer filed for disability payouts related to asbestos-caused lung disease. In a bid to limit compensation, the company set different standards for Black employees, who had to demonstrate more severe disease and lower lung function than their white coworkers to qualify for compensation (Braun, 2014). After all, the company argued, their Black employees started out with lower lung function according to the spirometer. This made it difficult for Black employees to challenge their employer, and it continues to affect the quality of care that Black patients receive to this day: the severity of their illness is not recognized by the machine. As noted in the previous section, the story of the pulse oximeter similarly requires Black patients to be sicker to receive care.

We can expect similar scenarios if, for example, health care professionals use LLMs to assist with diagnoses. Imagine an LLM app designed to summarize insights from previous scientific publications and generate health care recommendations accordingly. But if previous publications rely on racist assumptions, or simply ignore the needs of particular groups as in the case of the pulse

PDF: Click to return to top

oximeter, the LLM's advice is likely to be inaccurate too. And while these cases focus on medicine, we can imagine other domains including criminal justice, housing, and education where biases and discrimination enshrined in historical texts are likely to generate advice that perpetuates inequities in resource allocation. Aadhar, India's biometric identification system, has already begun to highlight such inequities. In order to receive an Aadhar number, citizens must provide their fingerprints, iris scans, and their photograph. For many of India's marginalized communities, such forms of identification are impossible to provide: their fingerprints have rubbed off due to years of manual labor, or they cannot provide an accurate iris scan due to a disability (Singh & Jackson, 2021). And yet, these are the communities that need Aadhar the most: in order to access any social services, they must provide not just their Aadhar number but also their biometric information.

Cochlear implants (CIs), which we introduced in Section 2, also demonstrate how technology can distort needs and ultimately erode services for marginalized populations. The FDA first approved CIs for adults in 1984 and for children in 1990 and consist of two components: a permanently implanted internal set of electrodes that interface directly with the nervous system, and an external processor that picks up sounds and translates them to patterns of electrical impulses for the electrodes. While many may assume that CIs cure deafness by giving the wearers the same level of aural capacity as those with natural hearing, in fact they only have a small range of audial inputs and patients must spend months or years of therapy to develop new connections in the brain to accommodate the CIs and learn to associate the signals with different sounds. In other words, deaf adults may still need auxiliary services even when people assume their problem has been solved.

Meanwhile, just as CIs were approved, academics and activists developed the concept of Deaf Culture (Denworth, 2014). Deaf Culture is based on shared values, experiences and beliefs of people influenced by deafness, and serves as a form of organizing political power. Activists define deafness as a neutral trait rather than a disability. Because most deaf children are born to hearing parents, acculturation happens primarily within deaf organizations and schools. Activists fear that the perception of CIs as a cure might lead to the reinterpretation of deafness as a voluntary disability, which could result in the defunding of deaf institutions (Tucker, 1998). This could also make it more difficult to access accommodations in employment or education such as interpreters, and alienate deaf people for whom CIs do not work or who choose not to get CIs (Cooper, 2019). Similarly LLMs, as a cheap and fast translation and interpretation tool, could actually lead to a reduction in other kinds of support, including human interpreters, written materials offered in languages other than English, and even language learning programs. This would create harm for Indigenous groups, marginalized groups that use dialects not covered well by LLMs, and immigrant communities. After all, because the LLM corpora are mostly in English and Chinese, they will be less accurate in other

languages. However, most people may not know about these deficiencies. In high stakes settings such as hospitals and courtrooms, human translators, though fallible, can rely on a variety of cues to ensure the person they are assisting understands what is happening, and can ask and answer clarifying questions. LLMs cannot do those things. If the LLM has a poor understanding of a particular language, or is otherwise unable to accurately translate technical medical or legal terminology, individuals are left without support, which depending on the setting could result in a variety of negative outcomes.

We can also expect similar outcomes when LLMs are used in social service provision. LLMs may be used to automatically screen applicants, or they might be used as a chat function on websites to assist people seeking resources or help. But the historical use of automated decision making tools by social service agencies produced results that are biased or inequitable in ways the tool is meant to prevent. Allegheny County, based in southwestern Pennsylvania, adopted the Allegheny Family Screening Tool (AFST), a computer-based program designed to assess the risk that a child might be experiencing harm and require intervention (Eubanks, 2018). The AFST uses a wide array of historical data about children and families, including data from local housing authorities, the criminal justice system, and local school districts, to produce a risk score that assesses the urgency of individual reports that come in through a child welfare hotline. Though it is designed to be used in tandem with a human screener, in practice the algorithm tends to train the humans, and over time the screeners' scores begin to match the algorithm's. In other words, independent human oversight is diminished.

AFST is supposed to be objective and evidence based, but its results overrepresent poor and working class families in ways that become a self-fulfilling prophecy (Eubanks, 2018). Simply asking for support from public services including childcare, tutoring, or therapy increases a family's risk score in the system. When wealthier families get that same support, they do so privately so it does not affect their score. As a result, simply being poor or requesting help become "risk factors". When a family's score is higher, it increases the likelihood that a report will result in a home visit, and pulls parents into a system of increased state surveillance, which

When institutions such as social service agencies, hospitals, insurers, or banks use LLMs to determine eligibility for or recommend products and services, we can expect that LLMs will make recommendations rooted in historical biases that will then produce inequitable outcomes.

itself increases the risk of further harms such as removal of children from the home, that parents will be arrested or lose their jobs, or even lose their housing. These outcomes then get fed back into the algorithm as evidence that its risk assessment was correct, even though in fact the risk assessment caused the outcomes.

When institutions such as social service agencies, hospitals, insurers, or banks use LLMs to determine eligibility for or recommend products and services, we can expect that LLMs will make recommendations rooted in historical biases that will then produce inequitable outcomes. They may systematically miscategorize or misinterpret people based on language use patterns, or fail to include key elements of their situation. Overall, like all conclusions drawn from huge datasets, the LLM is likely to focus on correlations in historical data (Bender & Gebru et al., 2021; Spinney, 2022). When evaluating eligibility for social services, LLMs might reinforce stereotypes about people who have previously used social services such as food banks or have a criminal record. They may also be used to collect and store additional information about people, similar to the way facial recognition technologies are being used in housing under the guise of ease of use (Strong et al., 2020). Or, institutions might use LLMs to determine whether a request for services is sincere or to provide advice in the interest of lowering the workload of social workers. LLMs might recommend against issuing a loan to some historically disadvantaged communities of color based on historical evidence that they might default. But this correlation is the outcome of prejudice and stereotypes in the training data, rather than a characteristic of these communities. Regardless, the consequences are dire: these decisions will generally favor the powerful, and further perpetuate inequitable distribution of resources.

LLMs will Reinforce Dominant Cultures

Developers have argued that LLMs hold tremendous promise for language translation (Brown et al., 2020), which will ultimately promote closer relationships across communities and international cooperation. We discuss how this capacity might transform scientific work in Section 7. However, as noted above, the largest and most powerful LLMs are being built in the United States and China, and their corpora are overwhelmingly dominated by English and Chinese language texts. Although developers are optimistic that LLMs will be able to translate across languages based on minimal training text, mistranslation is still likely. We are also likely to see language dominance (often English) reinforced and cultures distorted and even erased.

Let's return to the case of cochlear implants. Some deaf activists argue that they are an attempt by hearing medical professionals and parents to erase deaf culture (Ramsey, 2000). They fear CIs will lead to fewer students participating in deaf organizations where acculturation to deaf culture and visual language learning happen, and that a generation of deaf children will grow up without learning sign language and will

have difficultly communicating in the future because CIs are only a partial fix. LLMs could have similar impacts on other marginalized cultures. While some developers have argued that LLMs could help preserve languages that are disappearing (Coffey, 2021), they are also likely to contribute to the erasure of marginalized cultures and languages. Because their corpora are English and Chinese dominant, LLMs learn the rules of language through the lens of these languages, and are thus likely to be most accurate in their dominant training language. Eventually this will reinforce the dominance of standard American English in ways that will expedite the extinction of lesser-known languages, and contribute to the cultural erasure of marginalized people. Of the 300 Indigenous languages that were once spoken in the United States, only 175 remain today with most of them at risk of extinction (Cohen, 2010).

LLMs might also distort our understanding of other cultures. Consider the ongoing controversy over a museum to preserve and exhibit the history of Chinese Americans, built in New York City's Chinatown in 1980 (de Freytas-Tamura, 2021). The museum recently received a $35 million grant from the city in exchange for allowing the expansion of a jail in the neighborhood. Opponents argued that while the initiative is well-intentioned, funding the museum supports the preservation of only a narrow slice of Chinese culture while not considering or supporting the ongoing vitality of the community itself particularly as it faces gentrification pressures (de Freytas-Tamura, 2021). Similarly, LLMs could preserve limited, historically suspended understandings especially of the non-American or Chinese cultures represented in its corpora. And they could erroneously perpetuate these limited understandings, even as these cultures are changing, which could exacerbate cultural misunderstandings at the expense of the people of those cultures.

Responsibility for the Technology's Errors will fall on Marginalized Communities

Given the limitations in LLMs and the corpora behind them, we expect that the technology will be less accurate in already marginalized communities. But these inaccuracies may not always be clear. Let's return to the example of the spirometer. It is the gold standard for diagnosing and monitoring a broad range of common lung conditions, but patients need intact cognitive abilities, muscle coordination, and a certain level of physical strength to use it correctly. If it produces an inaccurate reading or misdiagnosis, the patient is usually blamed (Braun, 2021). But the problems with spirometry are systemic; in addition to its inaccuracies in Black people, it consistently fails for people with certain disabilities. One could imagine that the spirometer could be redesigned to accommodate these ground realities. Instead, individuals bear the responsibility for the technology's failures. If they cannot get an accurate spirometer reading, patients may undergo more invasive means of assessing lung function or lose access to important

social or health care services. And all the while, they may not know that the problems are systemic and instead blame themselves.

The same is true for pulse oximeters, which we discussed in Section 3. Health care professionals trusted the technology and perhaps also mistrusted the patients due to systemic biases (Fitzgerald & Hurst, 2017), and marginalized communities had to manage the consequences without knowing that the seemingly objective technology had failed them. We anticipate similar outcomes with LLMs. They might produce biased text, or some communities will not be able to use them due to financial limitations, disability, or language barriers. But the technology will not be blamed, especially as it becomes ubiquitous. Instead, any problems will be individualized and treated as a personal failing. In some cases, the problem might not even be clear and even more difficult to identify and solve.

Finally, we know that building and maintaining LLMs is extraordinarily expensive, and requires an enormous amount of computing power. Even using them requires access to a computer and high speed internet; as LLMs become embedded in more areas of life, lack of access will deepen existing inequalities, but the cause will not be visible. In the US, for example, both infrastructure and architecture are largely built for cars. Both road design and transportation policy favor the speed and convenience of people driving private vehicles over the safety and wellbeing of people who walk, use public transportation, or ride bicycles (Shill, 2020). In many regions of the country, including urban, suburban, and rural areas, private cars are the only available method of transportation, but cars are inaccessible to large portions of the population. They are expensive to purchase, maintain, insure, and store, and 40% of people with disabilities in the U.S. cannot or do not drive (Bureau of Transportation Statistics, 2018). Meanwhile, many forms of employment, important services, and markets are only accessible by car. But governments and businesses rarely acknowledge or do anything to address these inequities. As a result, these communities are not only further marginalized, but also alienated because their concerns seem rare and out of the mainstream (Schmitt, 2020).

In addition, car crashes kill nearly 40,000 Americans every year, and seriously injure millions more (National Highway Traffic Safety Administration, 2021); Black and Latinx pedestrians and bicyclists are disproportionately the victims of car crashes, even though they are less likely to have access to cars (Schmitt, 2020). However, local, state and federal governments do little to improve other forms of transportation or protect the safety of more vulnerable road users (Shill, 2020; Singer, 2022). Lack of access to a car, as well as being killed or injured by one, is treated as a personal shortcoming, rather than as a societal failure. Lack of access to LLMs, as well as any negative impacts an LLM might have on a person's life, will similarly be blamed on the individual, rather than the systems that produced those impacts.

In sum, we imagine that LLMs will reproduce societal biases in a few ways. First, because

they rely on historical texts, they are likely to reproduce systemic biases reflected in those texts. But, these biases will not be clearly visible to LLM users because they will be reproduced by seemingly objective algorithms. Second, they are likely to reinforce already dominant cultures, while creating historically arrested caricatures of others. Finally, as they become ubiquitous their limitations and errors will become less clear. Users will absorb the responsibility and blame, sometimes without even realizing it. This phenomenon is likely to be much more acute in marginalized populations.

Section 5: Remaking Labor and Expertise

KEY POINTS

- LLMs will transform, rather than replace, most occupations. In most cases, humans will shift to more complex and risky tasks.
- LLMs will transform authorship and associated standards for certification and evaluation.
- LLMs will eliminate some tech-based professions and enable others.
- Workers, supported by consumers, will resist these technologies.

For years, observers have predicted that the rise of artificial intelligence would trigger significant job losses, particularly for those in lower skilled occupations (West, 2019). Our analogical case study analysis validates these concerns, and suggests that LLMs will transform some professions completely. But we expect LLMs to have a major impact on higher skilled professions, which will use the technologies to summarize, generate, and translate text. While LLMs will be initially introduced as assistive technologies, they will eventually take over more common and predictable tasks. This will leave more challenging labor to humans which will carry physical, psychological, and social risks. We also expect these changes to trigger popular unrest and mobilization, both organized and informal.

Transforming Professions

As they become more accurate, LLMs will change work across a range of jobs, from translation to customer service. Over time, they will likely perform central parts of even high-skilled professions including constructing legal arguments and thus replacing the lawyer's typical tasks (Blijd, 2020). Consider how technology has transformed the medical profession (Howell, 1995). Over the last two decades, the internet has allowed patients to search for information related to their concerns and join online support groups to develop knowledge about the conditions that directly affect them. They then visit their physicians armed with this information, ready to ask

for particular services or treatment plans. In market environments like the United States, some are even prepared to visit another physician to confirm their diagnosis and facilitate their proposed treatment. Thus far, the "new expert patient" has not led to massive deprofessionalization of medicine. However, the doctor-patient relationship has changed (Tan & Goonawardene, 2017; Broom, 2005). Rather than providing their clients with information, physicians are increasingly focusing on helping them manage and interpret it. This new role comes with new expectations, as it requires physicians to stay up to date on new medical research. And patients may make more demands for access to diagnostic, prevention, and treatment technologies. LLMs are likely to increase patients' access to biomedical knowledge. As more scientific research is incorporated into the corpora (Else, 2021), the models will be able to summarize recent findings at a level that lay people can understand. This will exacerbate the trend identified here, in which physicians play a more supportive, interpretive role than a didactic one.

LLMs will also perform more mundane tasks and shift the risky work onto humans. In the early part of the 20th century, genetic services were only offered to the public through geneticists or genetic counselors who had extensive graduate training and worked at specialized clinics (Hogan, 2016), usually based in academic medical centers. These experts work with families to understand their histories of and experiences with particular diseases and then use this information to predict whether a disease might emerge in subsequent generations and advise how to avoid such circumstances.

By the middle of the 20th century, new technologies such as chromosomal analysis assisted their work, but counselors still played the primary role in interpreting the results and guiding people through difficult decisions

Credit: Sven Dowideit (CC BY SA 2.0)

about marriage, reproduction, estate planning, and communication and disclosure among loved ones (Rapp, 1999). But by the end of the century, companies had begun to offer genetic testing directly to consumers. In contrast to genetic counseling services available mostly at universities for relatively high prices, genetic tests could be ordered online for a much lower fee. And testing companies claimed greater accuracy than the human interpretation of family histories of disease (Parthasarathy, 2007). Today, these direct-to-consumer genetic tests play a central role in assessing susceptibility to disease. While specialized genetics clinics remain, they are small, unknown to many, and tend to focus on complex cases. Primary care physicians are often not able to answer

PDF: Click to return to top

patients' genetics-related questions if they are not equipped with specialized genetics expertise. The tests have thus offloaded the mundane task of genetic testing from experts and put untrained patients and physicians in the risky position of interpreting the results.

Such changes are not unique to high-skilled professions. Despite the rise of point-of-sale (POS) systems in supermarkets and other stores, cashiers have not become obsolete (Mateescu & Elish, 2019). Instead, they have been retrained to help customers use "self" checkout systems. However, as with physicians and genetics professionals, their labor now focuses on facilitating the user's interaction with the technology (e.g., difficulty scanning an item or a coupon), which substantially changes the nature of their jobs. They are more likely to encounter customers who are tense and frustrated in their interactions with the technology, which puts them at higher risk and removes the pleasure of mundane, informal interactions. In this and many other cases, technological automation transforms a consistent and predictable job to one that deals primarily with exceptions and other problems that the technology cannot solve (Chui et al., 2015).

LLMs and Gatekeeping

LLMs will also transform the social understanding of authorship and professional standards, particularly in fields that prize writing including law, academia, and journalism. Authorship and credit are socially constructed, influenced by the circumstances of the time and place. For much of modern history, for example, authorship–as defined by copyright law–was restricted to individuals legally recognized as fully human: white men (Vats, 2020). Similarly, major scientific prizes tend not to recognize the contributions of women, even today (Lincoln et al., 2012).

LLMs will interrupt our understandings of authorship and require us to reconfigure our systems of evaluation and certification.

But technology also plays an important role in constructing authorship. The invention of the typewriter in the mid-19th century raised a serious question: how could you be sure who authored a document? Previously, both individuals and legal authorities trusted the authenticity of documents because they were handwritten and could be scrutinized using formal handwriting analysis. But the typewriter triggered forgeries and even fraudulent transactions (Moore, 1959). This led to the establishment of a "document examiner" who was widely recognized as an expert in determining the genealogy of

a document. Based on a document's page alignment, spacing, ribbon, color, overtyping, type variation, retyping, and more, the examiner could link it to a particular typewriter. In other words, with the rise of the typewriter, document examiners became central to the construction of authorship.

The rise of the premier scientific journal *Nature* is another useful analog. In its early years, the journal targeted a wide audience including laypeople. But over time, it began to focus exclusively on scientific professionals; it excluded political topics and published frequently to compete with other scientific journals (Baldwin, 2015). In the process, it defined science itself, constructing it as a technical domain of interest to a narrow set of practicing experts.

Like typewriters and *Nature*, LLMs will interrupt our understandings of authorship and require us to reconfigure our systems of evaluation and certification. Everyone, from high school teachers to the judges of the Pulitzer Prize competition will need to decide whether they will accept LLM-generated work, how much, from whom, and under what circumstances. They may also have to change their standards accordingly. However, if these institutions accept LLMs as legitimate authors, this could increase inequality for researchers or writers who do not use or have access to LLM-based tools and strip the authors who wrote the text the LLMs were trained on of due credit and value.

The Fall and Rise of Tech-Based Work

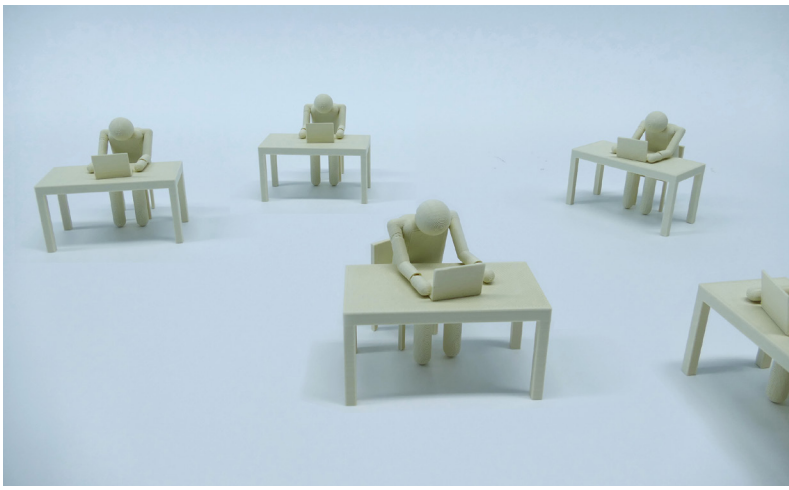
We believe that in many cases, LLMs are likely to change rather than eliminate most occupations. Before the invention of the traffic light, police teams managed flows of traffic manually, physically standing at intersections and directing traffic (McShane, 1999). Later, traffic lights used timers to automate light cycles, removing police from the activity and making it the responsibility of electricians and engineers. Police officers' responsibilities then shifted to dealing exclusively with violations and enforcing traffic laws.

However we expect that LLMs will eliminate some types of work completely, and trigger the creation of new types. The rapid social acceptance of the telegraph during the 19th century, for example, allowed buyers and sellers to communicate directly and inexpensively, and removed the need for both wholesalers and supply chain middlemen who had previously facilitated commerce (du Boff, 1984). Meanwhile, telegraphs also created new problems that required whole new categories of work. Telegraphs made it possible for all kinds of information to travel very quickly, regardless of its veracity. Malicious misinformation triggered financial panics. In response, companies developed the new field of 'business intelligence' to validate and distribute trustworthy information about commodities via telegraph (du Boff, 1984). Similarly, we expect that even as LLMs reduce or eliminate the need for some types of human labor, they will also prompt

the development of new professions and categories of expertise.

More recent paradigm shifts in communications technology have also created new professions and categories of work. In response to the flood of violent and often harassing vitriol unleashed on social media platforms, an entire ecosystem of work and expertise has developed. Social media companies created the content moderator to police and filter what users post on their platforms, and formed teams dedicated to developing principles and best practices for content moderation (Roberts, 2021; Gray & Suri, 2019). Legal departments handle lawsuits related to freedom of speech, harassment, and defamation. Researchers in academia study the impacts of toxic content on society, the experiences of workers responsible for moderating toxic content, and the effectiveness of different content moderation approaches (Roberts, 2021; Gray & Suri, 2019).

Within social media companies, the work of content moderation spans a global labor hierarchy similar to what we anticipate will happen with LLMs. Employees at the top of the hierarchy, dedicated to developing content policy and making organization-wide decisions, are highly paid, and typically located in San Francisco or another major city in the Global North. Workers at the bottom, who review content and process complaints, are typically located in countries with less expensive labor such as the Philippines or India. These "clickworkers" handle an immense volume of disturbing content at a rapid pace, which takes a psychological



Credit: Max Gruber / Better Images of AI (CC-BY 4.0)

toll (Roberts, 2021; Perrigo, 2022). In other words, workers in the Global North do safer, higher paid jobs while workers in the Global South, who have to manage the failures of the technology, are paid less and subject to psychological, emotional, and moral costs. We anticipate that LLMs will create similar divisions of labor, especially because many of the companies that created the need for content moderation are the same ones developing LLMs.

Labor Unrest

The history of technology is a history of labor unrest as workers worry how technological changes will affect their jobs. In fact, Luddites – now a term to describe someone resistant to emerging technologies – were 19th century textile workers who destroyed new machinery because they worried about their job security (Sale, 1995). Similarly, police officers initially resisted the introduction of traffic lights operated by engineers (McShane, 1999). Given this, we can expect some affected

workers to resist LLMs as well. Labor mobilization is likely to emerge first among unions, and then spread more widely.

We can expect some affected workers to resist LLMs.

Workers, both unionized and not, have consistently opposed the implementation of automated checkout devices at grocery stores, arguing that they eliminate jobs and disproportionately impact people of color (Harnisch, 2019). Some cashiers opposed them by going on strike (Lombrana, 2019; Pilon, 2019), and the United Food and Commercial Workers International Union's (UFCW) developed public campaigns against Amazon's cashierless grocery store model (UFCW, 2020). Consumers have also refused to engage with automated POS systems in solidarity with grocery workers, as well as in opposition to the loss of tax revenue that results from replacing tax-paying human workers with machines (Harris, 2018). These fears of job loss and degradation are not unfounded; grocery stores in France used automated POS systems to circumvent national labor laws, keeping stores managed by machine check-out counters open past legal working hours (France24, 2019; Alderman, 2019). The case of POS automation suggests that LLMs might similarly incite resistance from workers and consumers based on fear of job loss, violations of social norms, and reduced income taxes.

In addition to automation, there are several recent cases of workplace surveillance technologies prompting resistance from workers. In Amazon fulfillment centers, inventory scanners also track the workers; the scanners calculate the workers' efficiency and productivity and penalize them if they do not meet targets determined by algorithms (Guendelsberger, 2019). Amazon has also implemented software to track union activities in facilities (Del Rey & Ghaffary, 2020). In response, employees started using encrypted communication platforms to organize (Palmer, 2020). Sometimes, labor unrest is less coordinated. Truck drivers have responded to electronic monitoring by finding creative modes of resistance that are less confrontational and more identity-affirming (Levy, 2016). One trucker, for example, simply found a way to reprogram his vehicle's surveillance technology so that he could play solitaire. During the Covid-19 pandemic when office workers shifted to remote work en masse, companies deployed tracking software on employee computers that captured when people were using their mouse or keyboard. In response, employees started using mouse jigglers, keyboard tappers, and special apps to trick the trackers (Cole, 2021). When bosses use technology to control their workers, workers use technology to evade control. In the next Section, we describe how this technological one-upmanship leads to a loss in social trust overall.

LLMs are poised to remake labor and expertise across a wide swath of industries and professions. Some of these changes are likely to expand access to knowledge

or services that were previously limited by certification and professional gatekeeping, as in the case of direct to consumer genetic testing. Some will improve convenience for users at the expense of workers whose job or expertise becomes obsolete. Meanwhile, whole new forms of work and expertise will grow around the design, management, and use of LLMs. Those new roles will map onto existing corporate hierarchies rather than disrupting or overturning the current economic order. Even so, we expect that LLMs will exacerbate existing tensions between workers and corporations. Because of their flexibility, LLMs will function as both automation and surveillance technology, producing many areas for worker and consumer resistance.

Section 6: Increasing Social Fragmentation

KEY POINTS

- Publics will use LLMs to gather information that aligns with their interests and values, and ultimately challenge traditional expert authorities.
- The tailored information provided by LLMs will erode shared realities.
- LLMs will produce widespread public suspicion about legitimate authorship.
- LLMs will help outsider groups participate more actively in highly technical discussions related to science, technology, medicine, and the economy.
- LLMs will be less useful to already marginalized groups, increasing their social alienation.

Thus far, we have described LLMs primarily as a workplace technology that can improve professional life. But we also expect that publics will find great use in these technologies. As we suggested in the last Section, patients may use LLMs to access technical information about their medical conditions, which will empower them in their interactions with physicians. Governments might use LLMs to extract insights from large volumes of public comments about a proposed regulation, as a step towards more politically legitimate policies.

But this movement towards public empowerment will likely have negative impacts for institutional trust and social cohesion. In recent years, the United States has seen declining trust in authoritative institutions, from the government to science (Kennedy et al., 2022). We also have less trust in one another (Rainie & Perrin, 2019). Citizens feel that decisions made by elite institutions do not reflect their knowledge and lived realities (Parthasarathy, 2017). Media fragmentation has contributed to this: citizens can now find information that fits with their needs and values. We expect that LLMs will accelerate this trend. Publics will solicit information that aligns with their interests and values, which may contradict expert knowledge authorized by, for example,

scientific, legal, and medical establishments. LLMs will also be a crucial information-finding tool for social justice groups traditionally excluded from technology and technology policy domains, which will continue to erode institutional legitimacy and accelerate social fragmentation. Meanwhile, LLMs will generate questions about authorship that will erode interpersonal trust. Overall, we expect that the least privileged groups will experience the greatest social alienation. These groups already feel ignored by authoritative institutions and as we discuss earlier in this report, LLMs are likely to reproduce historical biases about, for example, the physiology, health, and lives of marginalized communities.

LLMs will destabilize institutional authority

For generations, experts across a range of fields, from science to the law, have controlled the production and interpretation of information. In order to file a complaint against their landlord, tenants usually need the help of a lawyer. To combat local air pollution, residents need scientific specialists who can help them interpret their symptoms quantitatively. These knowledge monopolies have given these professions economic, political, and cultural power. But the history of communications technologies suggests that LLMs will disrupt these monopolies.

The emergence of one of the earliest communication technologies, the printing press, is an excellent example. In 16th century Germany, Catholic priests held a monopoly over the Bible's teachings because the text was in Latin. Worshippers could not access the Bible's teachings directly. German priest Martin Luther was frustrated that his fellow priests abused this power; for example, they claimed that the Bible endorsed the exchange of money for admittance to heaven (Edwards, 2004). Luther responded by translating the Bible into vernacular German and using the newly developed printing press to disseminate the text, and his critique of the Catholic Church, across the country. His actions triggered the decline of the Catholic Church and rise of Protestantism in Germany and other parts of Europe (Dickens, 1974).

Similarly, patients today use the internet and social media to challenge the knowledge monopolies of their physicians. These technologies allow them to research their symptoms, diagnose themselves, and come to their medical appointments armed with information. This allows patients to advocate for themselves, develop a better understanding of their condition, and change the types of conversations they have with their physicians, as suggested in the previous Section (Tan & Goonewardene, 2017). However, patients lack the training and experiences of physicians, leading them to sometimes rely on incorrect information or magnify the importance of websites that fit with their preconceptions. When physicians are open to talking through this information, they are able to maintain their patient's trust (Tan & Goonewardene, 2017). But physicians often resist or refuse to discuss the information, which leaves patients frustrated (Stevenson et al., 2007; McMullan, 2006). Ultimately, this leads patients to seek out other physicians who might validate

their concerns and the evidence they have uncovered (Murray et al., 2003). In some cases, following misleading online advice can cause physical harm (Bessell et al., 2003), as in the recent case of COVID-19 treatments (Mariana, 2020). Overall, because physicians no longer have exclusive access to medical knowledge, the internet is eroding some of their power. Patients now feel emboldened to ask questions, and sometimes get other medical opinions if they are unsatisfied with the first.

In Section 5, we discuss how LLMs are likely to reshape knowledge-based professions. Here, we conclude that in the aggregate this will also destabilize the cultural power of professionals. Developers may create apps that provide individuals with medical or legal advice, or scientific information. LLMs could go as far as generating contracts such as drafting an amicus brief in a court case, or offering chat-based psychiatric services, thus providing individuals with direct access to some of the services that they would normally only access through experts and often for a significant fee. We expect that this will profoundly challenge social understanding of expertise and authority structures, and may even challenge certification systems. While authority figures will still be necessary for some tasks such as writing prescriptions, they will need to evolve to the changing information landscape.

LLMs will destabilize the cultural power of professionals.

LLMs as a mobilization tool

We also expect activists to use LLMs to challenge particularly technical areas of science, technology, and public policy. In recent decades, communities–particularly those that are low-income and historically disadvantaged–have become frustrated that science and technology do not reflect their needs and priorities, and have mobilized in response. But in order to influence decision making, they need to develop a technical understanding of the issue at hand and also translate their concerns into quantitative or scientific language (Parthasarathy, 2010). In the 1980s, fed up with the lack of research and treatments for AIDS, activists taught themselves immunology, microbiology, public health, and the science of clinical trials in order to translate their concerns to scientists and policymakers (Epstein, 1996). They used this knowledge to force the National Institutes of Health (NIH) to fund more HIV/AIDS research, demand that the Food and Drug Administration expedite approval of potentially useful treatments, and change how biomedical scientists tested the effectiveness of drugs. A decade later breast cancer activists, similarly concerned about the scale and ferocity of the disease and what they perceived to be a weak scientific and government response, took a similar path (Dickersin et al., 2001). They set up special

workshops to train people with breast cancer about the science of the disease, diagnostics, treatments, and prevention measures, and then successfully lobbied the government to include these people on committees that reviewed applications for breast cancer research funding.

More specifically, activists have a long record of using technology to bring data and gain public attention to their issue of concern. For years, residents of Norco, Louisiana, had complained about the air pollution coming from a local chemical plant. But the government did not take these concerns seriously; according to its measurements, which analyzed average air quality over 24-hour samples taken once every six days, local residents were not at high risk (Ottinger, 2010). But residents worried about the impacts of short term pollution flare ups that the government's sensors did not capture. So they taught themselves not only about the science of air pollution and its health impacts, but also about monitoring technologies. They constructed their own bucket monitoring system that took measurements over shorter duration and during the flare ups (Ottinger, 2010). Ultimately, this influenced air pollution monitoring systems not only at the US Environmental Protection Agency but around the world (Scott & Barnett, 2009). Similarly, patients have used technology to challenge expert understandings of disease. Recently, long COVID sufferers used Twitter to identify themselves, find one another, and crowdsource symptoms and treatments months before scientists or physicians acknowledged that the condition existed (Callard & Perego, 2021). However, although they were pleased that the biomedical community finally began to notice, and Congress authorized a research program dedicated to the disease, they were frustrated that their own knowledge-gathering and expertise were quickly dismissed once authoritative figures stepped in.

We expect that outsider activists will use LLMs in a variety of ways. Communities that feel unheard by scientists, engineers, and policymakers will use the technology to summarize the state of knowledge in a particular area in order to participate more confidently in public debate. Some will use LLMs to extract insights from or evaluate information that experts have traditionally ignored, such as pollution impacts, and then bring them to the attention of other citizens and decision makers with the additional legitimacy of the technology.

LLMs will increase social fragmentation and mistrust

LLMs will help people access information that fits with their interests and values, which means that neighbors might end up consuming rather different media. Of course, this phenomenon is not new. Benedict Anderson (1983) once famously wrote that newspapers construct "imagined communities", but we are now seeing how the explosion of online media and cable news, by providing content according to the user's needs and priorities, can actually produce the opposite. While this diverse media landscape broadens the perspectives involved in public and policy discussion by allowing

more people access to information and the ability to find communities where they are most comfortable, it also increases social fragmentation. We expect that LLMs will further erode the shared realities that were once constructed through common media consumption and ultimately, erode social trust.

The impact of the fragmentation of US television news provides a cautionary tale. For decades, Americans gathered around their television sets at 6pm to watch the evening news on one of three broadcast networks: NBC, ABC, or CBS. Executives at these networks had spent the day deciding which news was important to tell the American public, and curated the text and images to speak to the nation. But in the late 20th century the number of channels available to households increased dramatically, and people eating dinner could choose all sorts of programming to accompany them. In 1980, CNN appeared, and gave viewers the opportunity to watch the news all day. Media executives and advertisers soon learned that there was an audience for 24-hour news, and launched both Fox News and MSNBC in 1996 to take advantage of the market. Now, there was a battle for viewership among the multiple broadcast and cable networks, and each tried to cater to a different audience (Fox to conservative viewers, MSNBC to liberal viewers, CNN to establishment viewers) (Morris, 2007). They also focused on sensational stories in order to capture and maintain attention (Gordon, 2000).

Cable news channels were successful, building loyal audiences over time. Fox News, in particular, became a favorite for conservative audiences and eventually integral to their identities (Hoewe et al., 2020). These channels don't just provide different perspectives on the same issues. They often report on different issues entirely, which gives their viewers different understandings of what is happening in the world and makes it difficult to maintain a shared reality, which contributes to political polarization (Gordon, 2000). In recent years frequent viewing of Fox News, for example, shaped attitudes towards building a wall on the US border with Mexico and government action regarding climate change (Hoewe et al., 2020).

LLMs are likely to produce greater social fragmentation than cable news.

LLMs are likely to produce greater social fragmentation than cable news, as users will be able to use LLMs to distill the text they consume into forms that fits their individual needs and priorities. A user could use an LLM to filter news articles or summarize the key pieces of information. Consumers would thus no longer be exclusively shaped by the priorities of media executives, at least until media executives figure out how to integrate LLMs into their offerings. Websites or social services that use LLMs to generate text may cater to different demographics with new specificity and accuracy, but in the process of providing bespoke information they will

erode shared realities even further. As noted above, unlike social media, LLMs will not even be able to build new communities. Meanwhile, developers will have an incentive to intentionally tune LLMs to generate text that is as attention-grabbing as possible, in order to capture more users.

In addition, as LLMs get better at writing text that is indistinguishable from something a human could have written, they will not only challenge the cultural position of authors but also trust in their authorship. As noted in the previous section, every technology that enables new ways to make, copy, and distribute creative work produces a new round of both cultural and legal negotiations about how authorship is defined and authenticated. LLMs will be used for writing tasks that range from enhanced spelling and grammar checkers to producing entire paragraphs or even articles from whole cloth. This will make it much more difficult to determine just how much human effort was involved in the creation of a given text and will enhance social suspicion related to authorship.

For example, many schools and universities today use plagiarism detection technologies to prevent student cheating. One such service, Turnitin, compares the submitted paper against its massive database of previous student papers, as well as common sources like encyclopedias and textbooks (Foster, 2002; Davies, 2022). If the software determines that some or all of the text is substantially similar to material in the database, it flags the paper for cheating. Students are required to submit their papers to Turnitin for plagiarism review before they go to the instructor for grading. However, this has triggered a technological arms race. A variety of services have emerged to help students cheat while evading detection by Turnitin, from websites full of how-to advice to paid essay writing services. They are all findable with a quick web search.

LLMs will trigger a similar dynamic. As students use the technology to write better papers, instructors will employ more and more sophisticated methods of detecting LLM assistance, and students will fight to stay one step ahead. On both sides, companies will be ready to stoke and profit from mistrust by selling tools and services that promise to detect or obscure the use of an LLM in writing. All of this will erode trust between students and their educational institutions. This phenomenon will not be unique to educational institutions. The more writers of all kinds use LLMs for assistance, the more efforts to authenticate whether they "really" wrote their article or book, and the more writers will find new ways to take advantage of LLM capabilities without detection. In the long run, this will foster cultures of suspicion on a massive scale.

LLMs are likely to foster cultures of suspicion on a massive scale.

LLMs will have disparate impacts on social trust

Although we expect LLMs to be framed as mechanisms for empowerment and increased access to knowledge, the composition of the corpora coupled with the priorities of developers will likely result in a technology that is more useful for dominant groups than for communities that are already marginalized. Since they are trained primarily on text written by the majority, LLMs better reflect their views and linguistic style. There is already evidence that LLMs reflect racial and other forms of social bias against marginalized populations (Abid et al., 2021; Greene, 2021). Meanwhile, privileged members of society are likely to have more opportunities to shape the ways LLMs are integrated in daily life. This, in turn, will create distance between social groups. The more that LLMs shape public and private sector services, the more marginalized communities will feel alienated from them and from society.

In Section 4, we discuss how technology can reinforce systemic bias. Here, we emphasize how the same technology can be seen completely differently by dominant and marginalized groups, which has serious impacts for public trust. For decades, the United States and other countries have used cameras to ensure public safety. However, in practice, they tend to extend and reinforce surveillance over historically disadvantaged communities of color while making dominant communities feel protected. Amazon, for example, portrays its Ring doorbell–a motion-sensing, video-recording doorbell connected to an app–as fun, a way to be a good neighbor, and stay safe (Selinger & Durant, 2021). Part of the marketing behind the devices focuses on "surveillance as a service" (West, 2019). Scholars call this "luxury surveillance", because it enables self-reflection, empowerment, or care for a select group (Gilliard & Golumbia, 2021). This framing focuses on white and wealthier members of society, who tend to view law enforcement and surveillance technologies as protecting their interests. Furthermore, many individuals dislike surveillance, they may feel like they have nothing to hide and are in control of the technology (rather than the opposite). By contrast, people of color and other disadvantaged communities tend to experience "imposed surveillance", such as Detroit's facial recognition technology program Project Greenlight, which is imposed on the local population. They tend not to trust police or surveillance technologies because they have a long legacy of being victimized by them (Browne, 2015).

Credit: WikiCommons user Abas Gemini (CC BY SA 4.0)

PDF: Click to return to top

Biometric technologies, which track location and bodily measurements such as heart rate, pulse, and sleep, have similar disparate impacts (Gilliard & Golumbia, 2021). Law enforcement officials have long used ankle monitors to keep track of individuals caught up in the criminal justice system, whether out on bail, on house arrest, or on parole. But the ankle monitor is quite similar to the FitBit or Apple watch, except that the agency of the user differs. As we discuss in Section 4, even roads have had these kinds of impacts. In the 1950s, city planners across the United States used the emerging interstate highway system to segregate Black and white communities. This quickly became a crucial dividing line that allowed white neighborhoods to attract investment and feel protected, while their Black neighborhoods were isolated and economically starved (Miller, 2018).

Marginalized individuals may not get to decide when or how they encounter an LLM.

We expect LLMs to have similar impacts. Privileged communities might be able to choose to use them, to help write a blog post, summarize technical information, or file a legal complaint against a service provider. But already marginalized individuals may not get to decide when or how they encounter an LLM. For example, government agencies might use them to evaluate someone's application for social services, or as a chatbot that answers public questions. But because the technology is less likely to be accurate or useful to these communities, it may make it more difficult for these individuals to access crucial programs that might improve their lives.

This is likely to increase social alienation of already marginalized communities. Thanks to the long legacy of racism in biomedicine, Black communities distrust both scientists and physicians. This has had serious impacts during the COVID-19 pandemic, for example, when the US Black community had a particularly low vaccination rate (Willis et al., 2021). Similarly, frustrated by a long history of media bias (Race Forward, 2014), Black Americans tend to trust media sources based on how they portray their racial group (Kilgo et al., 2020).

LLMs' capacity to summarize and generate text will undoubtedly benefit users by answering their complex queries and making technical information more accessible. This will empower people to fight for their needs in their individual interactions with medical and legal experts, and to mobilize against technical organizations. But, this individual and community empowerment has social costs. We expect that LLMs' capacity to produce tailored text will further fragment society, as publics can essentially generate information or look at information through a lens that fits with their needs and values. This is likely to hurt already marginalized communities the most, since LLMs are likely to be the least useful for their needs and even reproduce biases against them.

LLM CASE STUDY

Section 7: Transforming the Scientific Landscape

KEY POINTS

- LLMs will transform both the kind of research scientists do, and how they do it.
- Academic publishers are likely to develop LLMs to maintain their monopoly power over most scientific literature.
- Using LLMs to conduct scientific evaluation will generate controversy among scientists.
- LLMs will reinforce Anglo-American dominance in science.

Throughout this report, we have anticipated the social, political, and equity implications if LLMs are adopted across a range of sectors. In this chapter, we examine how LLMs might transform one sector in particular: science. In this analysis, it is crucial to remember that the major LLMs currently under construction are based on corpora composed primarily of open access texts available online. But, most recent research publications–particularly scientific journal articles–are owned by academic publishing companies such as Elsevier and JSTOR. Therefore, they are not part of these corpora. We expect that these publishers might develop their own LLMs that leverage their proprietary text databases, particularly at a moment when universities are frustrated by their high fees (Resnick & Belluz, 2019). These proprietary LLMs are likely to be of greatest interest to the scientific community because they will be the most up-to-date, in contrast to publicly available LLMs that may contain slightly older scientific knowledge. As they become more important to academic researchers, universities may be forced to maintain their subscriptions. Less likely is that academic publishers will sell their texts to the large companies for inclusion in their corpora, because it would make their texts essentially available to everyone.

In this new environment, LLMs will transform scientific practices, including authorship and citations. They may also transform peer review systems, which have increasingly come under scrutiny. LLMs will also reinforce Anglo-American dominance in

PDF: Click to return to top

science. While they may help some scientists from low and middle income countries participate more actively in the international scientific community and engage in cross-national collaboration, the English and Chinese language dominance of the corpora will limit efforts to "decolonize" science. Finally, LLMs will limit the power of the open access movement, as academic publishers are likely to have more resources than governments, non-profit organizations, and individuals to generate LLMs.

LLMs will transform scientific practices

Remaking scientific authorship and methods

Given their capacity to process and summarize huge amounts of text, we expect LLMs to have a profound impact on authorship and scientific methods as well as evaluation. As we describe in more detail below, researchers in non-English speaking countries are likely to use LLMs to more accurately translate texts or check their grammar or spelling. This might make it easier for them to publish in top journals, which are invariably published in English. Even English-dominant researchers might use LLMs to generate more generic parts of scientific texts, including materials and methods, and parts of introductions and conclusions. As we discuss in Section 5, we expect that these uses will trigger questions about rightful authorship.

We also expect LLMs to profoundly shape scientific practice. The development of particle accelerators in the 1930s allowed physicists to investigate the structure of the atomic nucleus, and more recently to investigate subatomic particles (Ishkhanov, 2012). The polymerase chain reaction technique, which makes millions of copies of small pieces of DNA, transformed genetics and biotechnology research and enabled mapping and sequencing the human genome, the study of ancient DNA, and gene manipulation including CRISPR gene editing (Rabinow, 2011). And the internet has already had profound impacts on research. It has made it easier for scholars to read research across fields, and thus promote interdisciplinary thinking (Herring, 2002). It has also helped researchers contact a wider array of potential subjects, whether for clinical trials or for surveys and interviews. Social scientists, for example, use email, social media, and even the "crowdworking" platform Mechanical Turk (MTurk) owned by Amazon to publicize their studies and recruit subjects. MTurk allows researchers to access a fairly representative population for a small fee (less than half of minimum wage) (Fort et al., 2011).

LLMs will similarly enable new forms of research, perhaps most notably in the humanities. Historians and scholars of English literature will be able to quickly generate summary information about historical texts or genres in the major corpora or new texts they wish to consider. However, scholars may be reticent to use these sources for two reasons. First, scholars accustomed to using archives and carefully documenting the provenance of texts are likely to be wary

of LLMs as data sources at least initially, because of the lack of transparency about the texts contained in the corpora and the inability to cite them specifically. Scholars and academic publications will likely have to develop conventions about whether and how LLMs are used and documented. Wikipedia, for example, has become an important source introducing scholars to a particular topic, but is generally not acceptable as a reference in serious scholarly work (Chen, 2009). Second, because corpora predominantly include dominant and privileged voices, they may be of less utility in fields that are increasingly trying to capture the perspectives and experiences of those who have been historically marginalized.

LLMs will also continue to transform the nature of scientific inquiry. In recent years, there has been an explosion in enormous datasets and the computing power needed to process them. As a result, scientists can now use algorithms to identify correlations in huge datasets rather than starting with hypotheses (Huang, 2018; Kitchin, 2014). However, these correlations tell them neither about causality nor how such relationships emerge. In addition, just because a correlation appears in the data doesn't mean it is real or meaningful (Zhang, 2018). Researchers could also use LLMs as a new tool for data analysis, using them to extract insights from or summarize large amounts of text. Qualitative researchers are often constrained by the laborious manual processes of thematic coding, for example, but LLMs would allow them to analyze greater quantities of data or draw insights from data sources such as social media posts that were previously too large to consider as research sources.

Psychologists and political scientists could use data from the corpora to assess public attitudes and concerns. Given academic pressures to publish ("or perish"), we expect the proliferation of articles identifying data correlations. However, without changing statistical methods, this could also increase the production of spurious data that cannot be reproduced.

Scientific Credit Systems will Change

Scientists identify the lineage of their interests, theories, and methods through explicit citations to earlier work. This is an important method of providing credit. It has also become crucial to measuring scholarly impact. Scientists use "citation counts" to decide whether a publication is worth reading, or citing in their own publications. Hiring, tenure, and promotion committees use these indicators to judge a scientist's impact. Meanwhile, journals have developed "impact factors" based on the average number of times their articles are cited; these impact factors in turn affect scientists' decisions where to publish and university decisions on how to evaluate employees and applicants. However, citation practices are also highly political; white men tend to be the most cited across fields (Caplar et al., 2017; Dworkin et al., 2020).

We expect LLMs to reduce citations overall, and ultimately reinforce existing biases in research fields. While LLMs currently do not have the technical capability to identify which text from the corpus informed the generated text, if a future LLM is able to provide

citations along with the text summaries, we expect it to privilege highly cited articles which are not likely to represent the field's diversity or its most novel findings. But in the more likely scenario, scientists might query an LLM about the prevailing knowledge related to a particular phenomenon and simply treat the output as general knowledge that doesn't need to be cited. Consider the recent controversy over sharing data about COVID-19 genomic variants. Western scientists advocated putting this information into an open database that could be used across the world, to facilitate quicker understanding of disease progression and development of prophylactics, diagnostics, and treatments (Van Noorden, 2021). However, scientists from Southern countries protested, arguing that the open approach would rob them of the opportunity to receive credit for their hard work identifying variants such as Omicron (Maxmen, 2021). They worried further that scientists from wealthy nations would publish papers based on—but not citing—their results, because they had the resources to do further analysis, write up their findings, and submit them for publication. More generally, they were frustrated that as soon as they had begun to build expertise and resources to participate in the transnational world of science, Western leaders seemed to be changing the game. Similarly, marginalized scientists might worry that LLMs will make it more difficult for them to receive credit and for their ideas to become recognized as part of a mainstream corpus of knowledge.

They were frustrated that as soon as they had begun to build expertise and resources to participate in the transnational world of science, Western leaders seemed to be changing the game. Similarly, marginalized scientists might worry that LLMs will make it more difficult for them to receive credit and for their ideas to become recognized as part of a mainstream corpus of knowledge.

Transforming Peer Review

We also expect research funding agencies, scientific publishers and editors, and even patent systems to consider incorporating LLMs into their review processes. These institutions depend on technical experts to assess the novelty of a study or invention, the appropriateness of the methods, and the plausibility of findings. Invariably, these experts also advise researchers how to consider and address counterfactuals,

strengthen their claims or findings, or simply improve their writing. But peer reviewers are unpaid, and as academic pressures increase it is difficult to find good peer reviewers; editors say that they spend an enormous amount of time searching, and even then the reviewers may be uninformed, provide insufficient evaluation, or take too long and delay publication (Benos et al., 2007; Severin & Chataway, 2021). LLMs could solve many of these problems. Developers could create algorithms based on the backlists of all scholarly publications, or smaller ones targeted to a particular field or a particular journal, in order to identify high-quality publications and even advise authors how to improve their publications or fit better with the journal's standards. In fact, researchers have already begun to develop algorithms that claim to predict the grantability of patent applications, and even which patents are likely to be the most consequential (Candia & Uzzi, 2021). The next step would be to use LLMs to determine patentability, a particularly attractive option as patent offices struggle to hire and retain their personnel.

In the short term, editors might use LLMs as a half-measure, to help identify peer reviewers. They might ask the LLM: "who is an expert in X topic?" Editors have long used email and the internet in this way, which has allowed them to diversify their pool of reviewers. However, because LLM corpora are composed of historical texts, this use might actually eliminate the gains in reviewer and field diversity made in recent years. Unless the LLM is used very carefully, and with additional checks, this use could also affect a field's trajectory. An LLM might define reviewer expertise in terms of the number of citations in a particular journal (or set of journals), which may not represent a field's cutting edge.

If humans begin to use LLMs to conduct peer review itself, this could become a bigger problem. LLMs are likely to produce conservative peer reviews. We expect editors to use LLMs to scaffold parts of the peer review process—that is, to train the technology to look for particular elements in a paper, such as particular methods—to ensure quality reviews. However, this scaffolding could produce inflexible standards and slower recognition of truly novel results. It could also transform scientific practices. Consider the history of the IRB, in which narrow definitions of risk, benefit, and generalizable research have become hurdles for researchers (White, 2007). Or, educators in K-12 schools, who have increasingly had to twist their instructional strategies to accommodate standardized testing (Shelton & Brooks, 2019). Overall, LLMs might be good at evaluating papers in a field where the conventions, materials, and methods are well-established. However, it is hard to imagine how a corpus based on historical texts could adequately evaluate new and evolving science (Kuhn, 1962); we already know that this is a challenge for human reviewers (Pontis et al., 2017). As a result, widespread use of LLMs for primary peer review could limit creativity. It could also perpetuate biases against certain types of investigation, such as on structural racism or systemic inequality (Hoppe et al., 2019).

Scientific Evaluation by LLMs will Create Crises of Credibility

LLM-based scientific evaluation systems could also erode trust both within and beyond science. Today, peer review is the predominant form of scientific evaluation. Experts in a subfield review grant applications and scientific publications, and validate the ideas or findings as credible and worthy of funding or further circulation through scholarly journals or academic presses (Latour, 1987). Media outlets and governments often expect research to be peer reviewed before reporting on it or using it as the basis for policymaking. But this approach to evaluating scientific results is not natural or self-evident; it is the product of social negotiations and settlement. And it could certainly be otherwise. In the 17th century, wealthy gentlemen were assumed to be trustworthy–and producing credible scientific findings–because they were free from economic pressures (Shapin, 1995). They maintained their credibility by employing probabilistic discourse and minimizing precision, so as to avoid direct conflict with their peers. Scientists also trusted others' findings because they could witness the experiments themselves (Shapin & Schaffer, 1985). As the scientific enterprise grew, witnessing became "virtual", through standardization of methods, research publications, and peer review (Baldwin, 2018). These changes, however, came from within the scientific community, invariably when they concluded that they needed to establish credibility among new audiences.

Credit: Philadelphia College of Pharmacy and Science (CC BY 4.0)

In fact, professional communities respond quite poorly to externally imposed evaluation systems, and these external impositions tend to be less successful when the community is powerful. For example, in 1836 the US Congress passed a law requiring the Patent Office to employ examiners with science and engineering backgrounds, to replace the clerks who had previously handled patent applications. It was concerned that the bureaucracy was issuing too many patents based on old, unoriginal, and non-workable ideas, and believed that highly trained technical experts would solve the problem (Swanson, 2009). However, when these new examiners applied scientific standards for novelty and nonobviousness, they found that very few applications should be granted. Patent agents and lawyers, who were accustomed to a bureaucracy that had only legal criteria for granting patents, protested vigorously and threatened that if no patents were granted, the fledgling US economy would fail. They were ultimately successful; Patent Office administrators negotiated with

the new examiners to lower their standards. Physicians launched similar protests when the United States began to consider a national health care system in the mid-20th century, because they worried that it would lead to new forms of oversight and evaluation (Starr, 1982).

Especially because many scientists have already begun to criticize the business models of academic publishing–and ultimately distrust their intentions–we expect that if these companies build LLMs to replace peer review it will create a similar crisis among scientists. Scientists will not trust the technology to replace their judgment, and will likely point out the types of limitations that we have outlined above. We also expect publics to question scientific results that LLMs have evaluated, particularly in the early days of the technology or in response to the publication of particularly controversial ideas. And if communities don't trust evaluation systems then they will challenge the institutions promoting them. Prescription drug recalls have engendered not only mistrust in the US Food and Drug Administration, but hesitancy towards vaccines (Goldenberg, 2021). Similarly, distrust in the US Centers for Drug Control and Prevention has exacerbated resistance to mask wearing and other protection measures during the COVID-19 pandemic.

LLMs will Reinforce Public Myths about Science

As we have discussed in earlier sections of this report, we expect LLMs will increase the trend towards open and free information facilitated by the internet. Patients will be able to query disease symptoms and receive summaries of related medical articles. Curious individuals can generate lay summaries about the most technical topics, from astrophysics to artificial intelligence. In many respects, this will, as developers argue, democratize access to knowledge.

But as the technology presents complex scientific findings in comprehensible language, we expect that it will flatten important nuance, caveats, error rates, and uncertainties. This, we fear, will reinforce the illusion that scientific findings are objective, stanceless, value-free, and are generated with a view from nowhere. Ultimately, this could exacerbate public skepticism of science. We have seen this with previous efforts to popularize science. Scientific journalism, for example, tends to minimize what scholars call the "translational gap": the amount of additional research needed before scientific findings can lead to better medical practice (Summers-Trio et al., 2019). Instead, they tend to overestimate the importance of early stage studies. For example, many early biomedical studies are performed on mice. This can provide general indicators about the safety or effectiveness of a particular treatment, or shape of a particular phenomenon, but

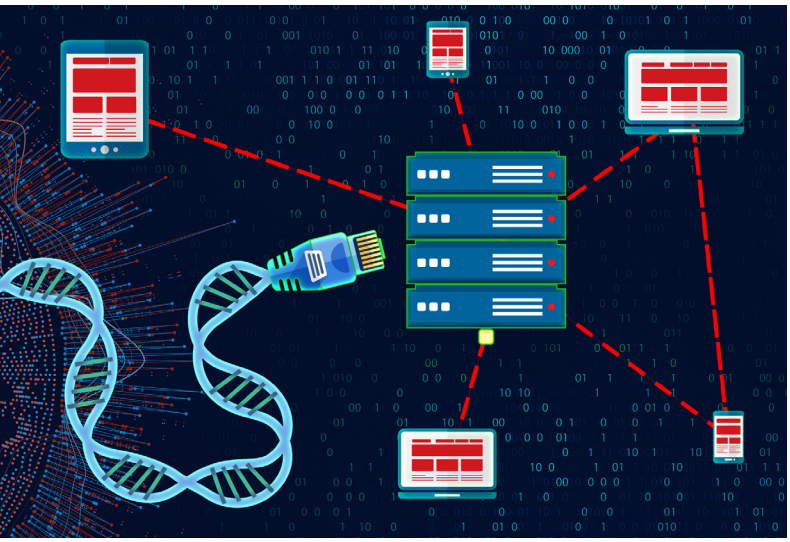
mice are quite different physiologically than humans. However, media articles still report these results with breathless excitement, creating false expectations about the imminence of treatments and the power of science (Chakradhar, 2019). Similarly, museums and other exhibitions such as World's Fairs tend to produce idealized images of cultures and countries, reinforcing distorted public understandings with real geopolitical consequences (Swift, 2019). We expect LLMs to reinforce a similarly idealized image of science, which will leave publics bewildered and frustrated when they confront its realities. Ultimately, this could exacerbate problems of public trust and alienation particularly among publics already questioning scientific findings (Funk, Kennedy, & Tyson, 2020; Funk, Kennedy, & Johnson, 2020).

LLMs will Hurt Open Access Movements

Finally, we expect LLMs to become another tool for academic publishing giants to maintain their control over scientific knowledge. In recent years, researchers have become increasingly concerned about how journal subscription costs hurt access to knowledge. This, they argue, limits who can participate in scientific knowledge production and ultimately, the quality of science itself. In response, universities are canceling huge journal subscriptions (Resnick & Belluz, 2019). Researchers are sharing preprints on their own websites, or on portals such as Sci-Hub and ArXiV.org (Nicholas et al., 2019). They are publishing in "open access" journals. Journals may implement new forms of monetization by charging LLM developers who use their university subscriptions to incorporate journal articles into training corpora. But we believe that LLMs will increase the attractiveness of Elsevier and other academic publishers themselves. Given their financial resources and monopolies over huge volumes of scientific texts, publishers could create their own LLMs for researchers and bundle them in their services to academic institutions. They might even require universities to purchase all of their journals in order to access their LLM. Indeed, companies frequently leverage emerging technologies to maintain or enhance their monopoly power. Monsanto spliced "terminator gene" technology into its genetically modified crops in order to prevent them from replicating (Masood, 1998). This meant that farmers could not replant their seeds after the growing season, which they had done for hundreds of years. Similarly, academic publisher JSTOR, in conjunction with MIT, used its internet surveillance capabilities to track down and stop excessive downloads of journal articles it owned. An MIT student activist Aaron Swartz downloaded these articles in order to promote their open access; he was later criminally charged for this act and died by suicide (Schwartz, 2013).

Given the vitality of the open access movement, we expect scientists to resist by creating grassroots LLMs. They might build on the work of non-profit initiatives such as Eleuther AI and rely on pro bono expertise and donated pre-prints and other text to develop apps. Scientists made similar attempts to gather data about disease-causing mutations in genes linked to breast

and ovarian cancer (known as the BRCA genes), to compete with biotechnology company Myriad Genetics' virtual monopoly on BRCA gene testing in the United States (Conley et al., 2014). Myriad used its testing monopoly to build a proprietary database of information about the genomic variants discovered, their association with disease, as well as individual and family health histories. Even though it lost its US testing monopoly in 2013 after patients, physicians, and scientists contested its patents (Parthasarathy, 2017), Myriad maintained its intellectual property through this database; patients and physicians preferred to use Myriad's



Credit: Ernesto Del Aguila III, NHGRI

testing service rather than others because the database could provide better interpretations about the implications of the genetic variants for disease. In order to build their alternative, scientists had to rely on word of mouth, and voluntary submissions of test results and other information from patients and physicians. This made it virtually impossible to build a database as powerful or useful as Myriad's, which in turn made it difficult to challenge the company's monopoly. We expect scientists developing grassroots LLMs to confront similar challenges, even if they have access to adequate technical expertise and financial resources.

LLMs Will Reinforce Anglo-American Scientific Dominance

Like the telephone and the internet, LLMs may facilitate global scientific communication and even cooperation. However, given the technology's capacity to summarize and translate text, some may assume that it could facilitate real international inclusion and even the "decolonization" of science. Consider how the internet has changed science. Internet search engines, scientific databases, and social media have helped scientists learn about and build upon one another's work, regardless of where they are in the world. Email has facilitated communication, allowing researchers to contact one another and even collaborate despite living in different time zones or on distant continents. Indeed, there is evidence that international scientific collaboration has increased significantly in recent years, allowing scientists to share project costs, gain access to expansive or unique physical resources, share more data, and enhance creativity (Matthews et al., 2020). And yet, technology-mediated communication also increases misunderstandings. Whereas previous

collaborations may have required scientists to visit laboratories for extended periods of time to learn methods, now such collaborations can occur without any in-person contact. This makes it much more difficult to transfer tacit knowledge–intangible scientific practices–which is essential for proper collaboration (Collins, 1992). However, scientists may not be aware that this knowledge is lost.

While LLMs may help some scientists in low and middle income countries, the prevailing political economy of science is likely to prevent true mutual learning and engagement.

In the abstract, LLMs could allow scientists across countries to read texts in their native languages, facilitating communication. In practice, however, the picture already looks more complicated. As we have noted repeatedly throughout this report, LLM corpora–particularly those being built by the major companies–are primarily in English, and to a lesser extent, Chinese. This is crucial when considering the impacts of LLMs for international scientific cooperation; it means that the technology's translation capabilities are likely to be poor, particularly for the languages where there are fewer digitized texts. While scientists in non-English speaking countries may initially use them for translation purposes, the outputs will likely be filled with errors and this practice will stop. However, we do expect scientists to use LLMs to improve their English writing, to facilitate journal publication. While scientists in former British or US colonies could also use them to gain easier access to knowledge, they may still not have access to the proprietary LLMs sold by academic publishing companies. Thus, while LLMs may help some scientists in low and middle income countries, the prevailing political economy of science is likely to prevent true mutual learning and engagement.

Instead, we expect LLMs to reinforce Anglo-American dominance in science while also helping Chinese scientists. In fact, it may also promote international collaboration between the two. Our research suggests that most efforts to promote mutual understanding across nations cannot escape geopolitical power struggles. Consider the World's Fairs, international platforms to showcase national scientific and technological achievements and facilitate cultural exchange, which began in the late 18th century. Cities hosting these yearly events brought global attention to their activities, and the sites also usually featured themed pavilions from a variety of countries that allowed them to showcase themselves and perhaps even develop grounds for collaboration (Molella & Knowles, 2019). However, countries used these as opportunities to advance their priorities. In 1993, South Korea's fifth largest city Daejeon hosted a Specialized Expo which produced international investment, and brought

attention to another region beyond the large and prosperous city of Seoul (Knowles, 2019 p. 207). Similarly, while both the United States and Soviet Union focused on similar themes of technological progress and cultural diversity in the 1958 World's Fair, the United States took a less serious approach in order to downplay the perception of its strength and power during the Cold War (Swift, 2019 p. 38). Similarly, *Nature* has always characterized itself as a premier scientific journal that explicitly serves an international community despite its British base. However, in its early decades it saw the world through a British lens (Baldwin, 2015). Contributors adopted a voyeuristic approach to foreign science, and often used it as a foil to comment on national affairs.

The more common LLMs become as a scientific tool, the more they will reinforce English as the lingua franca of science. This will likely also mean that the values and concerns of the English-speaking world– particularly the United States and Britain– will dominate global scientific priorities. Furthermore, knowledge produced in English may be viewed as more generalizable than knowledge produced in other languages. And yet, these political implications may remain hidden because LLMs will be promoted as a technology that will be able to truly globalize science.

In this section, we have explored the range of implications that LLMs will have on scientific knowledge and practice. We expect LLMs to transform scientific priorities and practices, and systems of authorship, credit, and evaluation. This may produce crises of credibility, not only within science and beyond. It will also strengthen the power of scientific publishers, despite growing frustration about their knowledge monopolies. Finally, while we are hopeful that LLMs could facilitate international cooperation and inclusion, we fear that this will not materialize unless the corpora become much more diverse.

PDF: Click to return to top

Policy Recommendations

LLMs have great potential to benefit society. However, the priorities of the current development landscape make it difficult for the technology to achieve this goal. Below, we articulate how both LLMs (the models themselves, corpora, and output) and LLM-based apps must be regulated in order to maximize the public good. We also recommend greater scrutiny of LLMs' impacts on labor and the environment. Finally, we recommend that the National Science Foundation (and similar science funding agencies around the globe) invest more heavily in research related to LLMs and their impacts, to balance attention in an area currently dominated by the private sector.

1

RECOMMENDATION 1

The US government must regulate LLMs, for example through the Federal Trade Commission. This should include:

a. Clear definition of what constitutes an LLM.

b. Evaluation and approval of LLMs based on: 1) process of corpus development and ongoing procedures for maintenance and quality assurance; 2) diversity of the corpus; 3) LLM performance including accuracy particularly in terms of output related to marginalized communities; 4) transparency of the corpora and algorithms; and 5) data security.

c. Evaluation of efforts to diversify corpora. Government should monitor data extraction practices to ensure that efforts to diversify the corpora are ethical.

d. A complaint system that allows users to document their negative experiences with an LLM. These complaints should be publicly available. Developers must articulate in writing how they have addressed all complaints.

e. Ongoing oversight and monitoring of LLMs. Developers must make the corpora available to regulators for periodic testing. This should include both basic accessibility and comprehensibility to someone with a basic understanding of data and computer science.

f. Requirement to label all LLM output as such and include information about the developer.

POLICY RECOMMENDATIONS (CONTINUED)

2

RECOMMENDATION 2

The US government must regulate all apps that use LLMs, for example through the Federal Trade Commission, according to their use. The more consequential the LLM output, the greater the regulatory scrutiny (e.g., LLM-based apps related to criminal justice and patient care receive more extensive evaluation). Evaluation should consider:

a. Whether app developers are using the right LLM for their needs.

b. Likelihood that the app will generate false or dangerous results.

c. Potential benefits for the user.

d. Social, equity, and psychological implications, including potential harms to end users.

3

RECOMMENDATION 3

Either a national or international standard setting organization (e.g., National Institute for Standards and Technology, International Standards Organization) must publish yearly evaluations of LLMs. They should assess: 1) diversity of the corpora; 2) performance; 3) transparency; 4) accuracy; 5) data security; and 6) bias towards marginalized communities.

4

RECOMMENDATION 4

The US government must enact comprehensive data privacy and security laws.

5

RECOMMENDATION 5

Under no circumstances should LLM-based apps deployed by the government (e.g., chatbots that provide information about social services, pre-trial risk assessment apps in criminal justice proceedings) harvest personally identifiable information.

POLICY RECOMMENDATIONS (CONTINUED)

6

RECOMMENDATION 6

The agencies that regulate LLMs and LLM-based apps, those that incorporate LLMs into its services, and all standard-setting bodies (e.g., the National Institute for Standards and Technology) must employ full-time advisors in the social and equity dimensions of technology. This "Chief Human Rights in Tech" Officer would advise procurement and technology evaluation decisions, monitor the technology once it is used and flag problems, and address disparate impacts.

7

RECOMMENDATION 7

Both national and international intellectual property authorities (e.g., the US Copyright Office, the World Intellectual Property Organization) must develop clear rules about the copyright status of LLM-generated inventions and artistic works.

8

RECOMMENDATION 8

All environmental assessments of new data centers must evaluate the impacts on local utility prices, local marginalized communities, human rights in minerals mining, and climate change.

9

RECOMMENDATION 9

The US government must work with other governments around the world (perhaps under the auspices of the United Nations) to develop global labor standards for tech work (including minerals mining).

10

RECOMMENDATION 10

The government must evaluate the health, safety, and psychological risks that LLMs and other forms of artificial intelligence create for workers, e.g., reorienting them towards more complex and often unsafe tasks. The Occupational Safety and Health Administration can perform this role, but it will require new regulations for workplace safety and an expansion of its purview to include psychological risks.

POLICY RECOMMENDATIONS (CONTINUED)

11

RECOMMENDATION 11

The US government must develop a robust response to the job consolidation that LLMs, and automation more generally, are likely to create. At a targeted level this should include job retraining programs and at a broad level, a guaranteed basic income and universal health care.

12

RECOMMENDATION 12

The National Science Foundation must substantially increase its funding for LLM development. This funding should prioritize:

a. Developing alternative corpora and models, especially those driven by the needs of low-income and marginalized communities (and in partnership with them).

b. Meetings that establish standards for making corpora representative and for incorporating the knowledge of citizens (particularly low-income and marginalized communities)

c. Supporting updates and maintenance of existing corpora and models (in contrast to just making more new models).

d. Support research into building new types of models that are more easily updated and maintained.

e. Research into evaluation of fit between model and use.

f. Research on the equity, social, and environmental impacts of LLMs.

Developers' Code of Conduct

LLMs are likely to trigger profound social change. Both LLM and app developers must recognize their public responsibilities and try to maximize the benefits of these technologies while minimizing the risks. To do this, they should adhere to the following practices:

LLM Developer Responsibilities

- LLM developers should dedicate significant effort and resources to maintaining and improving on existing LLMs rather than exclusively developing new ones. LLMs must be kept up to date with changing language and sentiments.
- LLM developers should curate corpora with care. They should resist appropriating already assembled bodies of text that were created for other purposes. They should instead define standards their corpus needs to meet and build a collection of texts with those standards in mind.
- Construction of the corpora must be ethical and be reviewed by ethics experts before deployment. Authors should be able to opt-out of their texts' inclusion in the corpora.
- LLM developers should make each corpus publicly accessible for other developers and interested stakeholders to scrutinize. They should be open to the problems identified by these stakeholders and make changes accordingly.
- LLM developers should prioritize research in the following areas:
    - Building models that are easily updated and maintained
    - Evaluating the fitness of a model for a particular task
    - Equity, social, and environmental impacts of LLMs
    - Understanding and explaining to end users the rationale behind LLM output

App Developer Responsibilities

- App developers must carefully evaluate the social and equity implications of their products before development, with the help of potential users, relevant stakeholders, and experts who systematically analyze

technology (i.e., science and technology studies scholars). This includes systematic analysis of both positive and negative implications for marginalized communities.

- App developers must label LLM-generated text as such.

Both LLM and App Developers

- Rather than creating a few general purpose LLMs and assuming they are ready to be integrated into a variety of apps, LLMs should be designed and evaluated for specific purposes. Both app and LLM developers should work together or developers should take on both of these roles.
- Both LLM and app developers must support low income and marginalized communities' capacity to drive development. This includes providing funding and technical support so that community organizations can develop their own apps and LLMs. In the process, developers must recognize that the trust of marginalized communities is fragile, and can only be achieved through authentic engagement and long-term relationships.
- LLM developers must be fully transparent about the limitations of their technology, including in their discussions with app developers. App developers, in turn, must not use LLMs to perform tasks they are not suited for. Specifically:
    - LLMs should not be treated as a source of intelligence since they were trained to model language, not understand the world. The fact that LLMs "know" some things about the world is coincidental.
    - Developers should build apps and deploy LLMs only in situations where up-to-date language patterns are not necessary. Since LLMs are conservative, they replicate the past.
    - An LLM cannot speak for everyone. LLMs are universalizing; they favor dominant language patterns and flatten nuance, but language is diverse even within a single language. This means that even an LLM that appears to be "neutral" will serve members of the dominant group as it alienates others.
- Both LLM and app developers should implement a complaint system for end users and other stakeholders to document their negative experiences with an LLM. Developers should be sympathetic and responsive to these concerns.

Recommendations for the Scientific Community

We urge all professions to develop rules and guidelines to accommodate the rise of LLMs. Because we focused our attention on how LLMs might affect science (Section 7), we offer recommendations specific to this community. We hope this will guide researchers, journal editors, scientific publishers, and universities, as they contend with this emerging technology.

Development of LLMs by the scientific community

- If scientific publishers develop LLMs, they should:
    - Provide users with information about how output is generated (i.e., the composition of the corpora and the logic of the algorithm).
    - Ensure that the LLM is accessible to and accurate for non-English speakers.
- The National Science Foundation should support the development of an LLM that includes publicly available journal articles and all results generated from their funding. It should deliberately include texts across all fields. To ensure that it captures the nuances of a variety of fields, experts from multiple disciplines–from the natural sciences to the humanities–should test it before deployment.
- All authors should be permitted to opt-out of their texts' inclusion in LLM corpora.

RECOMMENDATIONS FOR THE SCIENTIFIC COMMUNITY (CONTINUED)

LLM use for evaluation

- If scientific journals and academic publishers use LLMs to evaluate the quality of manuscripts, they must be transparent about this use. This includes clear explanations on the publisher's website so that prospective authors can be fully informed about LLM use before submission.
- Scientific journals and academic publishers should not rely completely on LLMs for "peer review". LLMs are likely to produce conservative evaluations–and therefore be more critical of novel findings and ideas–because they are based on historical texts.

Research using LLMs

- Scientific journals and academic publishers must develop rules for how they–and peer reviewers–will evaluate research conducted using LLMs.
- All publications that rely on LLMs for text analysis should provide detail about the corpora and algorithms on which the results are based.

Scientific communication using LLMs

- Scientific communicators should help publics understand how to use LLMs to interpret science. This includes evaluating which LLMs are the most appropriate for their needs, and how to understand the credibility of LLM output.
- Scientific communicators and publics should test LLMs before deployment to ensure that outputs related to scientific topics are accurate, credible, and comprehensible.

Acknowledgements

The authors would like to thank Shelby Pitts, Daniel Rivkin, and Nick Pfost for their assistance in researching, revising, and producing this report.

The Technology Assessment Project is supported in part through a generous grant from the Alfred P. Sloan Foundation (grant #G-2021-16769)

References

Abid, A., Farooqi, M., & Zou, J. (2021, July). Persistent anti-muslim bias in large language models. Proceedings of the *2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 298-306). https://doi.org/10.1145/3461702.3462624

Ahmed, N., & Wahed, M. (2020). *The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research*. ArXiv. https://arxiv.org/abs/2010.15581

AI Now Institute. (2021, October 5). Democratize AI? How the proposed National AI Research Resource falls short. *AINOW*. https://medium.com/@AINowInstitute/democratize-ai-how-the-proposed-national-ai-research-resource-falls-short-96ae5f67ccfa

AIRC. (n.d.). *About AIRC*. Retrieved March, 13, 2022 from https://www.airc.aist.go.jp/en/intro/

Al Jazeera. (2020, September 12). *At least 50 people feared dead in DR Congo mine collapse.* Al Jazeera. https://www.aljazeera.com/news/2020/9/12/at-least-50-feared-dead-in-dr-congo-mine-collapse

Alderman, L. (2019, December 26). Self-checkout in France sets off battle over a day of rest. *The New York Times*. https://www.nytimes.com/2019/12/26/business/self-checkout-automation.html

Alford, A. (2020, November 3). Large-scale multilingual AI models from Google, Facebook, and Microsoft. *InfoQ*. https://www.infoq.com/news/2020/11/multilingual-ai-models/

Algolia. (n.d.). Retrieved March 13, 2022 from https://www.algolia.com/

Allen, B.L. (2003) *Uneasy alchemy: Citizens and experts in Louisiana's chemical corridor disputes*. The MIT Press.

Allen, M. (2018, June 14). *And the title of the largest data center in the world and largest data center in the US goes to....* DataCenters.com. https://www.datacenters.com/news/and-the-title-of-the-largest-data-center-in-the-world-and-largest-data-center-in

Ames, M. G. (2019). *The charisma machine: The life, death, and legacy of One Laptop per Child*. MIT Press.

Amnesty International. (2016, January 19). Democratic Republic of Congo: "This is what we die for": Human rights abuses in the Democratic Republic of the Congo power the global trade in cobalt. *Amnesty International.* https://www.amnesty.org/en/documents/afr62/3183/2016/en/.

Amnesty International. (2020, February 10). Nigeria: 2020 could be Shell's year of reckoning. *Amnesty International.*

https://www.amnesty.org/en/latest/news/2020/02/nigeria-2020-could-be-shell-year-of-reckoning/

Amnesty International. (2020, May 6). DRC: Alarming research shows long lasting harm from cobalt mine abuses. *Amnesty International.* https://www.amnesty.org/en/latest/news/2020/05/drc-alarming-research-harm-from-cobalt-mine-abuses/

Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso.

AT Editor. (2019, October 17). Wellcome Sanger denies charge of misusing African DNA. *Africa Times.* https://africatimes.com/2019/10/17/wellcome-sanger-denies-charge-of-misusing-african-dna/Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019, November 15). Americans and privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center*. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

Baird Equity Research. (2017). Equifax Inc. (EFX) announces significant data breach; -13.4% in after-hours. https://baird.bluematrix.com/docs/pdf/dbf801ef-f20e-4d6f-91c1-88e55503ecb0.pdf

Baker, S.H., (2019). Anti-resilience: A roadmap for transformational justice within the energy system. *Harvard Civil Rights- Civil Liberties Law Review*, 54, 1-48. https://ssrn.com/abstract=3362355.

Baldwin, M. (2015). *Making "Nature": The history of a scientific journal.* University of Chicago Press.

Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of "peer review" in the Cold War United States. *Isis*, 109(3), 538-558. https://doi.org/10.1086/700070

Barbaschow, A. (2020, November 8). Australia's critical infrastructure definition to span communications, data storage, space. *ZDNet*. https://www.zdnet.com/article/critical-infrastructure-definition-to-span-communications-data-storage-and-space/

Baurick, T. (2020). *Bayou Bridge Pipeline protesters' lawsuit against company can proceed, judge rules*. NOLA.com. https://www.nola.com/news/environment/article_05cacaa4-0358-11eb-b1a2-4303c8dedb22.html

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., ... & Zotov, A. (2007). The ups and downs of peer review. *Advances in physiology education*. 31(2), 145-152. https://doi.org/10.1152/advan.00104.2006

Berridge, C., & Levy, K. (2019, July 24). Webcams in nursing home rooms may deter elder abuse – but are they ethical? *The Conversation.* https://theconversation.com/webcams-in-nursing-home-rooms-may-deter-elder-abuse-but-are-they-ethical-120208

Berry, I. (2021, September 20). *Top 10 countries with the most data centers.* Data Centre Magazine. https://datacentremagazine.com/top10/top-10-countries-most-data-centres

Bessell, T. L., Anderson, J. N., Silagy, C. A., Sansom, L. N., & Hiller, J. E. (2003). Surfing, self-medicating and safety: buying non-prescription and complementary medicines via the internet. *BMJ Quality & Safety*, 12(2), 88-92. http://dx.doi.org/10.1136/qhc.12.2.88

Betcher, M., Hanna A., Hansen, E., and Hirschmann, D. (2019, August 21). P*ipeline impacts to water quality: Documented impacts and recommendations for improvements.* Downstream Strategies and Hirschmann Water & Environment, LLC. https://www.tu.org/wp-content/uploads/2019/10/Pipeline-Water-Quality-Impacts-FINAL-8-21-2019.pdf

Better language models and their implications. (2019, February 14). OpenAI. https://openai.com/blog/better-language-models/

Bijker, W., Hughes, T.P., & Pinch, T. (1987). *The social construction of technological systems: New directions in the sociology and history of technology.* MIT Press.

Birhane, A. (2021). The impossibility of automating ambiguity. *Artificial Life*, 27(1), 44-61. https://doi.org/10.1162/artl_a_00336

Blijd, R. (2020, September 7). Will lawyers be replaced by GPT-3? Yes, and here's when. *Spark Max.* https://www.legalcomplex.com/2020/09/07/will-lawyers-be-replaced-by-gpt-3-yes-and-heres-when/

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S.... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. https://arxiv.org/pdf/2108.07258.pdf.

Braun, L., Wolfgang, M., & Dickersin, K. (2013). Defining race/ethnicity and explaining difference in research studies on lung function. *EUROPEAN RESPIRATORY JOURNAL*, 41(6), 9.

Braun, L. (2014). *Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics.* U of Minnesota Press.

Braun, L. (2021). Race correction and spirometry: why history matters. *Chest*, 159(4), 1670-1675. https://doi.org/10.1016/j.chest.2020.10.046

Broom, A. (2005). Medical specialists' accounts of the impact of the Internet

on the doctor/patient relationship. *Health*, 9(3), 319-338.

Brown, A., Parrish, W., and Speri, A. (2017, June 3) *Standing rock documents expose inner workings of 'Surveillance-Industrial Complex.'* The Intercept. https://theintercept.com/2017/06/03/standing-rock-documents-expose-inner-workings-of-surveillance-industrial-complex/

Brown, M. (2018, October 29). *One year later: The impact of Equifax's data breach.* Transforming Data With Intelligence. https://tdwi.org/articles/2018/10/29/biz-all-impact-of-equifax-data-breach.aspx

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.

Browne, S. (2015). *Dark matters*. Duke University Press.

Bureau of Transportation Statistics. (2018). *Travel Patterns of American Adults with Disabilities*. Retrieved March 13, 2022 from https://www.bts.gov/travel-patterns-with-disabilities.

Business & Human Rights Resource Centre. (2021, February). *Transition minerals tracker: global analysis of human rights policies & practices*. https://media.business-humanrights.org/media/documents/2021_Transition_Minerals_Tracker_Monday_w_numbers_updated.pdf

Buss, D., Rutherford, B., Stewart, J., Côté, G.E., Sebina-Zziwa, A., Kibombo, R., Hinton, J., and Lebert, J. (2019, November). Gender and artisanal and small-scale mining: Implications for formalization. *The Extractive Industries and Society*, 6(4), 1101-1112. https://www.sciencedirect.com/science/article/pii/S2214790X19301522

C., R. (2022). Almost 6 billion accounts affected in data breaches in 2021. *Atlas VPN*. https://atlasvpn.com/blog/almost-6-billion-accounts-affected-in-data-breaches-in-2021

Cagle, S. (2019, July 8) 'Protesters as terrorists': growing number of states turn anti-pipeline activism into a crime. *The Guardian*. https://www.theguardian.com/environment/2019/jul/08/wave-of-new-laws-aim-to-stifle-anti-pipeline-protests-activists-say

Cakebread, C. (2017, November 15). You're not alone, no one reads terms of service agreements. *Business Insider*. https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11?r=US&IR=T

Callard, F., & Perego, E. (2021). How and why patients made Long Covid. *Social science & medicine*, 268, 113426. https://doi.org/10.1016/j.socscimed.2020.113426

Candia, C., & Uzzi, B. (2021). Quantifying the selective forgetting and integration of ideas in science and technology. *American Psychologist*, 76(6), 1067. https://doi.org/10.1037/amp0000863

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, 1:141. https://www.nature.com/articles/s41550-017-0141

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2020). Extracting training data from large language models. arXiv. *arXiv*:2012.07805.

Chakradhar, S. (2019, April 15). It's just in mice! This scientist is calling out hype in science reporting. *STAT*. https://www.statnews.com/2019/04/15/in-mice-twitter-account-hype-science-reporting/

Chen, C.J. (2009). Art history: a guide to basic research resources. *Collection Building*, 28(3), 122-125. https://doi.org/10.1108/01604950910971152

Chui, M., Manyika, J., and Miremadi, M. (2015, November) Four fundamentals of workplace automation. *McKinsey Quarterly*. https://roubler.com/sg/wp-content/uploads/sites/49/2016/11/Four-fundamentals-of-workplace-automation.pdf

Coffey, D. (2021). Māori are trying to save their language from Big Tech. *Wired UK*. https://www.wired.co.uk/article/maori-language-tech

Cohen, P. (2010, April 5). Indian tribes go in search of their lost languages. *The New York Times.* https://www.nytimes.com/2010/04/06/books/06language.html#:~:text=Of%20the%20more%20than%20300,reclamation%20efforts%20have%20shown%20success

Colchete, G., & Sen, B. (2020, October). Muzzling dissent: How corporate influence over politics has fueled anti-protest laws. *Institute for Policy Studies*, https://ips-dc.org/wp-content/uploads/2020/10/Muzzling-Dissent-Anti-Protest-Laws-Report.pdf

Cole, S. (2021, December 8). Workers are using 'mouse movers' so they can use the bathroom in peace. *Vice*. https://www.vice.com/en/article/88gqgp/mouse-mover-jiggler-app-keep-screen-on-active

Collins, H. (1992). *Changing order: Replication and induction in scientific practice.* University of Chicago Press.

Common Crawl. (n.d.). *Want to use our data?* https://commoncrawl.org/the-data/.

Computer AI passes Turing test in 'world first'. (2014, June 9). BBC. https://www.bbc.com/news/technology-27762088

Conley, J. M., Cook-Deegan, R., & Lázaro-Muñoz, G. (2014). Myriad after myriad: the proprietary data dilemma. *North Carolina journal of law & technology*, 15(4), 597.

Cooper, A. (2019). Hear me out. *Missouri Medicine*, 116(6), 469-471. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913847/

Council of the EU (2018). *Copyright rules for the digital environment: Council agrees its position.* https://www.consilium.europa.eu/en/press/press-releases/2018/05/25/copyright-rules-for-the-digital-environment-council-agrees-its-position/#

Cowley, S., & Silver-Greenberg, J. (2019, November 3). These machines can put you in jail. Don't trust them. *The New York Times.* https://www.nytimes.com/2019/11/03/business/drunk-driving-breathalyzer.html

Cybersecurity and Infrastructure Security Agency. (n.d.) Infrastructure security. *Cybersecurity and Infrastructure Security Agency.* https://www.cisa.gov/infrastructure-security

Dale, R. (2017). The commercial NLP landscape in 2017. *Natural Language Engineering*, 23(4), 641-647. https://doi.org/10.1017/S1351324917000237

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113-118. doi:10.1017/S1351324920000601

Data Center Map. (2022). *Data Center Map.* https://www.datacentermap.com/

Davies, W. (2022, February 24). How many words does it take to make a mistake? *London Review of Books*, 44(4). https://www.lrb.co.uk/the-paper/v44/n04/william-davies/how-many-words-does-it-take-to-make-a-mistake

Daws, R. (2020, October 28). *Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves.* AI News. https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/

Day, T. (2017, August 1). Building data centers creates jobs. *U.S. Chamber of Commerce.* https://www.uschamber.com/technology/building-data-centers-creates-jobs

de Freytas-Tamura, K. (2021, August 19). Why some people in Chinatown oppose a museum dedicated to their culture. *The New York Times*. https://www.nytimes.com/2021/08/19/nyregion/chinatown-museum-protests.html

Del Rey, J. and Ghaffary, S. (2020, October 6). Leaked: Confidential Amazon memo reveals new software to track unions. *Vox*. https://www.vox.com/recode/2020/10/6/21502639/amazon-union-busting-tracking-memo-spoc

Denworth, L. (2014, April 25). Science gave my son the gift of sound. *TIME*. https://time.com/76154/deaf-culture-cochlear-implants/

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://arxiv.org/abs/1810.04805

Devlin, J., Change, M., Research Scientists, Google AI Language. (2018, November 2). Open sourcing BERT: state-of-the-art pre-training for natural language processing. *Google AI Blog*. https://

ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

Dickens, A. G. (1974). *The German nation and Martin Luther.* Edward Arnold.

Dickersin, K., Braun, L., Mead, M., Millikan, R., Wu, A. M., Pietenpol, J., Troyan, S., Anderson, B., and Visco, F. (2001). Development and implementation of a science training course for breast cancer activists: Project LEAD (leadership, education and advocacy development). *Health Expectations*, 4(4), 213-220. https://doi.org/10.1046/j.1369-6513.2001.00153.x

Dillet, R. (2021, March 11). Hugging Face raises $40 million for its natural language processing library. *TechCrunch*. https://techcrunch.com/2021/03/11/hugging-face-raises-40-million-for-its-natural-language-processing-library/

Dillon, G. A. O., Nancy. (2019, October 2). Google using dubious tactics to target people with "darker skin" in facial recognition project: sources. *Nydailynews.com*. https://www.nydailynews.com/news/national/ny-google-darker-skin-tones-facial-recognition-pixel-20191002-5vxpgowknffnvbmy5eg7epsf34-story.html

Downing, T.E. (2002, April). *Avoiding new poverty: Mining-induced displacement and resettlement*. Mining, Minerals, and Sustainable Development. https://pubs.iied.org/sites/default/files/pdfs/migrate/G00549.pdf

Du Boff, R.B. (1984). The telegraph in nineteenth-century America: Technology and monopoly. *Comparative Studies in Society and History*, 26(4), 571-586.

Duster, T. (1990). B*ackdoor to eugenics.* Routledge, Cop.

Dworkin, J.D., Linn, K.A., Twitch, E.G., Shinohara, R.T., & Bassett, D.S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8), 918-926. https://doi.org/10.1038/s41593-020-0658-y

Edwards Jr, M. U. (2004). *Printing, Propaganda, and Martin Luther.* Fortress Press.

Eggers, W.D., Malik, N, & Gracie, M. (2019). *Using AI to unleash the power of unstructured government data.* Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html

EleutherAI. (n.d.). *Frequently asked questions.* Retrieved March 3, 2022 from https://www.eleuther.ai/faq/

Else, H. (2021, October 26) Giant, free index to world's research papers released online. *Nature*. https://www.nature.com/articles/d41586-021-02895-8?utm_source=twt_nat&utm_medium=social&utm_campaign=nature

Englehardt, S., Han, J., & Narayanan, A. (2018). I never signed up for this!

Privacy implications of email tracking. *Proceedings on Privacy Enhancing Technologies*, 2018(1), 109–126. https://doi.org/10.1515/popets-2018-0006

Ensmenger, N. (2018, October). The environmental history of computing. *Technology and Culture*, 59(4). http://www.ctcs505.com/wp-content/uploads/2016/01/Ensmenger-2018-The-Environmental-History-of-Computing.pdf

Environmental Justice Atlas. (2019, April 19). *Tucuruí hydroelectric dam, Pará, Brazil*. Institute of Environmental Science and Technology at the Universitat Autònoma de Barcelona. Retrieved March 13, 2022 from https://ejatlas.org/conflict/tucurui-hydroelectric-dam-and-the-assassination-of-dilma-ferreira-silva-para-brazil

Epstein, J., Donnan, S., & Bass, D. (2022, January 28). Treasury weighing alternatives to ID.me over privacy concerns. *Bloomberg*. https://www.bloomberg.com/news/articles/2022-01-28/treasury-weighing-id-me-alternatives-over-privacy-concerns

Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in human behavior*, 17(3). 295-314. https://doi.org/10.1016/S0747-5632(01)00004-8

Epstein, S. (1996). *Impure science: AIDS, activism, and the politics of knowledge.* Univ of California Press.

Eschrich, J., & Miller, C. (2021, March 12). *Cities of Light: A Collection of Solar Futures.* Center for Science and the Imagination, Arizona State University.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Euromines. (2020, May). *The electronics value chain and its raw materials*. Euromines. http://euromines.org/files/key_value_chain_electronics_euromines_final.pdf

European Commission (2021). *A European approach to artificial intelligence*. Retrieved March 13, 2022 from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence#:~:text=The%20European%20approach%20to%20artificial,in%20AI%20and%20trustworthy%20AI.

Evans, W. (2016, December 12). Uber said it protects you from spying. Security sources say otherwise. *Reveal*. https://revealnews.org/article/uber-said-it-protects-you-from-spying-security-sources-say-otherwise/

Fagone, J. (2021, July 23). The Jessica Simulation: Love and loss in the age of A.I. *The San Francisco Chronicle*. https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/

Fairchild, D., & Weinrub, A. (2017). *Energy democracy: Advancing equity in clean energy solutions.* Island Press.

First Peoples Worldwide. (2019, March 14). *New report finds increase of violence coincides with oil boom*. University of Colorado, Boulder. https://www.colorado.edu/program/fpw/2019/03/14/new-report-finds-increase-violence-coincides-oil-boom

Fisher, E., Mahajan, R. L., & Mitcham, C. (2006). Midstream modulation of technology: Governance from within. *Bulletin of Science, Technology & Society*, 26(2), 485-496. https://doi.org/10.1177/0270467606295402

Fitzgerald, C., & Hurst, S. (2017). Implicit bias in health care professionals: a systematic review. *BMC Medical Ethics*, 18(19). https://doi.org/10.1186/s12910-017-0179-8

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine?. *Computational Linguistics*, 413-420. https://hal.archives-ouvertes.fr/hal-00569450

Foster, A. L. (2002, May 17). Plagiarism-detection tool creates legal quandary. The Chronicle of Higher Education. https://www.chronicle.com/article/plagiarism-detection-tool-creates-legal-quandary/

Foster, L. A. (2018). *Reinventing hoodia: peoples, plants, and patents in South Africa.* Wits University Press.

Foster, M. W., Eisenbraun, A. J., & Carter, T. H. (1997). Communal discourse as a supplement to informed consent for genetic research. *Nature Genetics*, 17(3), 277–279. https://doi.org/10.1038/ng1197-277

France24. (2019, August 27). *Protests erupt after French supermarket uses automation to evade labour laws*. https://www.france24.com/en/20190827-protests-erupt-french-supermarket-automation-labour-laws-sunday-laws

Frankel, T.C., & Whoriskey, P. (2016, December 19). Tossed aside in the 'white gold' rush. *Washington Post*. https://www.washingtonpost.com/graphics/business/batteries/tossed-aside-in-the-lithium-rush/

Frynas, J.G. (2001). Corporate and state responses to anti-oil protests in the Niger Delta. *African Affairs*, 100(398), 27–54, http://www.jstor.org/stable/3518371.

Funk, C., Kennedy, B., Johnson, C. (2020, May 21). *Trust in medical scientists has grown in U.S., but mainly among Democrats*. Pew Research Center. https://www.pewresearch.org/science/2020/05/21/trust-in-medical-scientists-has-grown-in-u-s-but-mainly-among-democrats/

Funk, C., Kennedy, B., Tyson, A. (2020, August 28). Black Americans have less confidence in scientists to act in the public interest. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/08/28/black-americans-have-less-confidence-in-scientists-to-act-in-the-public-interest/

Galligan, C., Rosenfeld, H., Kleinman, M., & Parthasarathy, S. (2020). *Cameras in the Classroom: Facial Recognition Technology in Schools.* Technology Assessment Project, Science, Technology and Public Policy Program, University of Michigan. http://

stpp.fordschool.umich.edu/sites/stpp.fordschool.umich.edu/files/file-assets/cameras_in_the_classroom_full_report.pdf

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., and Leahy, C. (2020, December, 31). The pile: An 800GB dataset of diverse text for language modeling. *arXiv*. https://arxiv.org/abs/2101.00027

Gavin, B. (2018, May 25). *How big are gigabytes, terabytes, and petabytes?* How-to Geek. Retrieved March 13, 2022 from https://www.howtogeek.com/353116/how-big-are-gigabytes-terabytes-and-petabytes/

Gershgorn, D. (2021, July 20). DuckDuckGo launches new Email Protection service to remove trackers. *The Verge*. https://www.theverge.com/2021/7/20/22576352/duckduckgo-email-protection-privacy-trackers-apple-alternative

Gibson, E. (2020, August 26). *New NSF AI research institutes to push forward the frontiers of artificial intelligence*. NSF. https://beta.nsf.gov/science-matters/new-nsf-ai-research-institutes-push-forward-frontiers-artificial-intelligence

Gilliard C., & Golumbia, D. (2021, July 6). *Luxury surveillance*. Real Life. https://reallifemag.com/luxury-surveillance/

Glanz, J., Creswell, J., Kaplan, T., Wichter, Z. (2019, February 3). After a Lion Air 737 Max crashed in October, questions about the plane arose. *The New York Times*. https://www.nytimes.com/2019/02/03/world/asia/lion-air-plane-crash-pilots.html?partner=IFTTT

Glanz, J. (2012, September 23). Data barns in a farm town, gobbling power and flexing muscle. *The New York Times*. https://www.nytimes.com/2012/09/24/technology/data-centers-in-rural-washington-state-gobble-power.html

Gokaslan, A., and Cohen, V. (2019). *OpenWeb text corpus*. http://Skylion007.github.io/OpenWebTextCorpus

Golden, H. (2021, October 15). Indigenous tribes tried to block a car battery mine. But the courts stood in the way. *The Guardian.* https://www.theguardian.com/environment/2021/oct/15/indigenous-tribes-block-car-battery-mine-courts

Goldenberg, M. J. (2021). *Vaccine hesitancy: public trust, expertise, and the war on science.* University of Pittsburgh Press.

Gordon, M. T. (2000). Public trust in government: The US media as an agent of accountability?. *International Review of Administrative Sciences*, 66(2), 297-310. https://doi.org/10.1177/0020852300662006

Gray, M.L., & Suri, S. (2019). *Ghost work: How to stop silicon valley from building a new global underclass*. Harper Business.

Greene, T. (2021, September 20). DeepMind tells Google it has no idea how to make AI less toxic. *The Next Web*. https://thenextweb.com/news/deepmind-tells-google-no-idea-make-ai-less-toxic?utm_campaign=Neural%20Newsletter&utm_medium=email&_hsmi=162703071&_hsenc=p2ANqtz-9ukUzOvODtsxryos3QqQUtYtzktnr7KI4FUYHBBkqAmxLLBL8bZGYEK9Y5nB3n9b05ishvnRk06Phu1JpWwlzGSPParw&utm_content=162703071&utm_source=hs_email

Grimmelman, J. (2016). Copyright for literate robots. *Iowa Law Review*, 101(2). https://ilr.law.uiowa.edu/print/volume-101-issue-2/copyright-for-literate-robots/

Grother, P., Ngan, M., & Hanaoka, K. (2019). Face recognition vendor test part 3: demographic effects. *National Institute of Standards and Technology*, 8280. https://doi.org/10.6028/nist.ir.8280

Guadamuz, A. (2016). The monkey selfie: copyright lessons for originality in photographs and internet jurisdiction. *Internet Policy Review*, 5(1). https://doi.org/10.14763/2016.1.398

Guendelsberger, E. (2019, July 18). I worked at an Amazon fulfillment center; They treat workers like robots. *TIME*. https://time.com/5629233/amazon-warehouse-employee-treatment-robots/

Gusmano, M.K., Kaebnick, G.E., Maschke, K.J., Neuhaus, C.P., & Wills, B.C. (2021). Public deliberation about gene editing in the wild. *Hastings Center Report*. 51 (S2): S2-S10. https://doi.org/10.1002/hast.1314

Guston, D. H., & Sarewitz, D. (2002). Realtime Technology Assessment. *Technology in Society*, 24(1-2), 93-109. https://doi.org/10.1016/S0160-791X(01)00047-1

Hamlett, P., Cobb, M. D., & Guston, D. H. (2013). National citizens' technology forum: Nanotechnologies and human enhancement. *Nanotechnology, the Brain, and the Future*, 265-283. https://doi.org/10.1007/978-94-007-1787-9_16

Hao, Karen (2020, December 16). 'I started crying': Inside Timnit Gebru's last days at Google--and what happens next. *MIT Technology Review.* https://www.technologyreview.com/2020/12/16/1014634/google-ai-ethics-lead-timnit-gebru-tells-story/

Hao, K. (2021, May 20). The race to understand the exhilarating, dangerous world of language AI. *MIT Technology Review.* https://www.technologyreview.com/2021/05/20/1025135/ai-large-language-models-bigscience-project/

Haranas, M. (2021, March 21). *Microsoft to build new $200M data center as Azure sales soar.* CRN. https://www.crn.com/news/data-center/microsoft-to-build-new-200m-data-center-as-azure-sales-soar

Haranas, M. (2022, January 20). *Data center market 2022 forecast: Private equity takes over.* CRN. https://www.crn.com/news/data-center/data-center-market-2022-forecast-private-equity-takes-over

Harmon, A. (2010, April 21). Indian Tribe Wins Fight to Limit Research of Its DNA. *The New York Times.* https://www.nytimes.com/2010/04/22/us/22dna.html?scp=1&sq=indian%20tribe%20wins%20fight%20to%20limit%20research%20on%20its%20dna&st=cse

Harnish, K. (2019, September 4). Oregon labor union wants voters to limit grocers to two self-checkout stations per store. *Willamette Week.* https://www.wweek.com/news/state/2019/09/04/oregon-labor-union-wants-voters-to-limit-grocers-to-two-self-checkout-stations-per-store/

Harris, R. (2020, December 16). Oxygen-Detecting Devices Give Misleading Readings in People with Dark Skin. *NPR.* https://www.npr.org/2020/12/16/947261192/oxygen-detecting-devices-give-misleading-readings-in-people-with-dark-skin

Harris, S. (2018, December 8). 'They kill jobs': Meet Canadians who refuse to use self-checkout. *CBC.* https://www.cbc.ca/news/business/self-checkout-cashier-jobs-retail-automation-1.4937040

Harwell, D. (2022). Facial recognition firm Clearview AI tells investors it's seeking massive expansion beyond law enforcement. *The Washington Post.* February 16. https://www.washingtonpost.com/technology/2022/02/16/clearview-expansion-facial-recognition/

Heaven, W.D. (2020, October 8). A GPT-3 bot posted comments on Reddit for a week and no one noticed. *MIT Technology Review*. https://www.technologyreview.com/2020/10/08/1009845/a-gpt-3-bot-posted-comments-on-reddit-for-a-week-and-no-one-noticed/

Heaven, W.D. (2021, May 14). Language models like GPT-3 could herald a new type of search engine. *MIT Technology Review*. https://www.technologyreview.com/2021/05/14/1024918/language-models-gpt3-search-engine-google/

Herring, S.D. (2002). Use of electronic resources in scholarly electronic journals: a citation analysis. *College & Research Libraries*, 63(4), 334-340. https://doi.org/10.5860/crl.63.4.334

Hirst, D. (2020, September 9). *How data centers became as important as water and energy.* Data Centre Dynamics Ltd. https://www.datacenterdynamics.com/en/opinions/how-data-centres-became-important-water-and-energy/

Hoewe, J., Brownell, K. C., & Wiemer, E. C. (2020, October). The role and impact of Fox News. *The Forum* 18(3), 367-388. De Gruyter. https://doi.org/10.1515/for-2020-2014

Hogan, A.J. (2016). *Life histories of genetic diseases.* Johns Hopkins University Press.

Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., Hutchins, B. I., Davis, A.F., Lauer, M.S., Valentine, H.A., and Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Science advances*, 5(10). https://doi.org/10.1126/sciadv.aaw7238

Howell, J.D. (1995) *Technology in the hospital: Transforming patient care in the early twentieth century.* Johns Hopkins University Press.

Hristov, K. (2017). Artificial intelligence and the copyright dilemma. *IDEA - The Journal of the Franklin Pierce Center for Intellectual Property*, 57(3), 431-454. https://ipmall.law.unh.edu/sites/default/files/hosted_resources/IDEA/hristov_formatted.pdf.

Huang, S. (2018). The tension between big data and theory in the "omics" era of biomedical research. *Perspectives in biology and medicine*, 61(4), 472-488.

Hughes, T.P. (1983). *Networks of Power: Electrification in Western Society, 1880-1930*. Johns Hopkins University Press.

Hutchins, J. (2003). ALPAC: the (in) famous report. *Readings in machine translation*, 14, 131-135.

Interpol. (2020). *Our 19 databases*. Retrieved March 13, 2022 from https://www.interpol.int/en/How-we-work/Databases/Our-19-databases

Isberto, M. (2021, June 9). *Are there benefits of a rural data center?*. Colocation America. https://www.colocationamerica.com/blog/rural-data-center-benefits

Ishkhanov, B. S. (2012). The atomic nucleus. *Moscow University Physics Bulletin*, 67(1), 1-24. https://doi.org/10.3103/S0027134912010092

Jemisin, N. K. (2011). The Trojan Girl. *Weird Tales #357*. https://nkjemisin.com/2012/08/the-trojan-girl/

Jemisin, N. K. (2012). Valedictorian. *After: Nineteen stories of apocalypse and dystopia* (T. Windling & E. Datlow, Eds.). Little, Brown.

Johnson, P. (2017). With the public clouds of Amazon, Microsoft, and Google, big data is the proverbial big deal. *Forbes*. https://www.forbes.com/sites/johnsonpierr/2017/06/15/with-the-public-clouds-of-amazon-microsoft-and-google-big-data-is-the-proverbial-big-deal/?sh=1a5dc7652ac3

Kelly, H. (2021, November 19). For seniors using tech to age in place, surveillance can be the price of independence. *The Washington Post*. https://www.washingtonpost.com/technology/2021/11/19/seniors-smart-home-privacy/

Kennedy, B., Tyson, A., and Funk, C. (2022, February 15). Americans' trust in scientists, other groups declines. *Pew Research Center*. https://www.pewresearch.org/science/2022/02/15/americans-trust-in-scientists-other-groups-declines/

Kilgo, D. K., Wilner, T., Masullo, G. M., & Bennett, L. K. (2020). *News distrust among Black Americans is a fixable problem.* Center for Media Engagement. https://mediaengagement.org/research/news-distrust-among-black-americans

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1), 2053951714528481. https://doi.org/10.1177%2F2053951714528481

Kline, R. & Pinch, T. (1996). Users as agents of technological change: The social construction of the automobile in the rural united states. *Technology and Culture*, 37(4), 763-795.

Knight, W. (2021, August 24). AI can write disinformation now—and dupe human readers. *WIRED*. https://www.wired.com/story/ai-write-disinformation-dupe-human-readers/

Knowles, S. G. (2019). Does the World's Fair Still Matter?. In Molella, A. P., & Knowles, S. G. (Eds.). *World's Fairs in the Cold War: Science, technology, and the culture of progress.* (pp. 194-212) University of Pittsburgh Press.

Ku, E., McCulloch, C.E., Adey, D.B., Li, L. & Johansen, K.L. (2021). Racial disparities in eligibility for preemptive waitlisting for kidney transplantation and modification of eGFR thresholds to equalize waitlist time. *Journal of the American Society of Nephrology*, 32, 677-685. https://doi.org/10.1681/ASN.2020081144

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Kulkarni, P., & K, C. N. (2021). Personally Identifiable Information (PII) detection in the unstructured large text corpus using Natural Language Processing and unsupervised learning technique. *International Journal of Advanced Computer Science and Applications*, 12(9). https://doi.org/10.14569/ijacsa.2021.0120957

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* Harvard university press.

Lawson, M.F. (2021, September 1). The DRC mining industry: Child labor and formalization of small-scale mining. *Wilson Center*. https://www.wilsoncenter.org/blog-post/drc-mining-industry-child-labor-and-formalization-small-scale-mining

Leahy, C., Hallahan, E., Gao, L., Biderman, S. (2021, July 7). What a long, strange trip it's been: EleutherAI one year retrospective. *EleutherAI*. https://blog.eleuther.ai/year-one/

Levinger, M. (2020). Triad in the therapy room - The interpreter, the therapist, and the deaf person. *Journal of Interpretation*, 28(1). https://digitalcommons.unf.edu/joi/vol28/iss1/5

Levy, K.E.C. (2016). Digital Surveillance in the hypermasculine workplace. *Feminist Media Studies*, 16(2), 361-365., https://doi.org/10.1080/14680777.2016.1138607.

Li, C. (2020, June 3). OpenAI's GPT-3 language model: A technical overview. *Lambda Labs*. https://lambdalabs.com/blog/demystifying-gpt-3/

Liddy, E.D. (2001). *Natural Language Processing*. Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc. https://surface.syr.edu/istpub/63/

Lincoln, A.E., Pincus, S., Koster, J.B., Leboy, P.S. (2012). The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s. *Social Studies of Science*. 42(2). 307-320. https://doi.org/10.1177%2F0306312711435830

Lombrana, L.M. (2019, July 10) Walmart Workers Rebel Against Retailer's Robot Push in Chile. *Bloomberg*. https://www.bloomberg.com/news/articles/2019-07-10/walmart-workers-rebel-against-retailer-robot-push-in-chile

Lopez, R. (2012). Urban Renewal and Highway Construction. In Lopez, R., *Building American public health: Urban planning, architecture, and the quest for better health in the United States* (199-138). Palgrave Macmillan.

Lothian-McLean, M. (2020, April 27). "Black Woman in US dies after being turned away from Hospital she worked at for 31 years." *indy100*. https://www.indy100.com/article/coronavirus-black-health-care-worker-dies-test-detroit-deborah-gatewood-9485341. Downloaded May 20, 2020.

Luccioni, A., and Viviano, J. (2021). What's in the box?: An analysis of undesirable content in the common crawl corpus. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2, doi:10.18653/v1/2021.acl-short.24.

Luitse, D., and Denkena, W. (2021). The great transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society*, 8(2). https://doi.org/10.1177/20539517211047734

Luong, N., & Arnold, Z. (2021). China's Artificial Intelligence industry alliance. *Center for Security and Emerging Technology*. https://doi.org/10.51593/20200094

Mall, A. (2021, May 24). Gas pipelines: Harming clean water, people, and the planet. *National Resources Defense Council, Inc.* https://www.nrdc.org/experts/amy-mall/gas-pipelines-harming-clean-water-people-and-planet

Manjoo, F. (2020, July 29). How do you know a human wrote this? *The New York Times.* https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html

Mariana, S. (2020, May 27). Coronavirus: The human cost of virus misinformation. *BBC.* https://www.bbc.com/news/stories-52731624

Masood, E. (1998). Monsanto set to back down over 'terminator' gene?. *Nature*, 396(6711), 503-503. https://doi.org/10.1038/24949

Mateescu, A., and Elish, M.C. (2019, January 30). AI in context: The labor of integrating new technologies. *Data & Society.* https://datasociety.net/library/ai-in-context/

Matthews, K. R., Yang, E., Lewis, S. W., Vaidyanathan, B. R., & Gorman, M. (2020). International scientific collaborative activities and barriers to them in eight societies. *Accountability in Research*, 27(8), 477-495. https://doi.org/10.1080/08989621.2020.1774373

Maxmen, A. (2021). Why some researchers oppose unrestricted sharing of coronavirus genome data. *Nature*, 593(7858), 176-177.

McMullan, M. (2006). Patients using the Internet to obtain health information: how this affects the patient–health professional relationship. *Patient education and counseling*, 63(1-2), 24-28. https://doi.org/10.1016/j.pec.2005.10.006

McShane, C. (1999) The origins and globalization of traffic control signals. *Journal of Urban History*, 25(3), 379-404., doi:10.1177/009614429902500304.

Merry-Noel, C. (2015). *Sighted/Human Guide: One Instructor's Perspective.* National Federation of the Blind. Retrieved March 13, 2022 from https://nfb.org//sites/default/files/images/nfb/publications/fr/fr34/1/fr340110.htm

Metz, C., & Wakabayashi, D. (2020, Dec 3). Google researcher says she was fired over paper highlighting bias in A.I. *The New York Times.* https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html.

Michelson, E. (2016). *Assessing the societal implications of emerging technologies: Anticipatory governance in practice.* Routledge.

Miller, Johnny. (2018, February 21). Roads to nowhere: How infrastructure built on American inequality. *The Guardian*. www.theguardian.com/cities/2018/feb/21/roads-nowhere-infrastructure-american-inequality.

MIT CSAIL. (n.d.) *Strategic partners*. Retrieved March 3, 2022 from https://www.csail.mit.edu/sponsors/strategic-partners

Mock, B. (2017, February 16).The meaning of blight. *Bloomberg*. https://www.bloomberg.com/news/articles/2017-02-16/why-we-talk-about-urban-blight

Molella, A. P., & Knowles, S. G. (Eds.). (2019). *World's Fairs in the Cold War: Science, Technology, and the Culture of Progress.* University of Pittsburgh Press.

Moltzau, A. (2020, August 2). A short history of natural-language understanding. *Towards Data Science.* https://towardsdatascience.com/a-short-history-of-natural-language-understanding-f1b3c382f285

Monaghan, J., & Walby, K. (2017). Surveillance of environmental movements in Canada: critical infrastructure protection and the petro-security apparatus, *Contemporary Justice Review*, 20(1), 51-70. https://doi.org/10.1080/10282580.2016.1262770

Moore, W C. (1959). The questioned typewritten document. *Minnesota Law Review*, 2585. https://core.ac.uk/download/pdf/217208687.pdf

Morales, G. D. F., Monti, C., & Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Scientific Reports*, 11(1), 1-12. https://doi.org/10.1038/s41598-021-81531-x

Moran-Thomas, A. (2020, August 5). How a popular medical device encodes racial bias. *Boston Review*. https://bostonreview.net/articles/amy-moran-thomas-pulse-oximeter/

More Perfect Union. (2021, November 9). *Google threatens water supply of drought-stricken town*. https://perfectunion.us/google-data-center-water-supply-oregon-drought/

Morgan, T.P. (2020, February 15). The datacenter has an appetite for GPU compute. *The Next Platform*. https://www.nextplatform.com/2020/02/15/the-datacenter-has-an-appetite-for-gpu-compute/

Morris, J. S. (2007). Slanted objectivity? Perceived media bias, cable news exposure, and political attitudes. *Social science quarterly*, 88(3), 707-728. https://doi.org/10.1111/j.1540-6237.2007.00479.x

Moss, S. (2021, October 21). Data center water usage remains hidden. *Data Center Dynamics.* https://www.datacenterdynamics.com/en/analysis/data-center-water-usage-remains-hidden/#:~:text=Direct%20water%20consumption%20of%20US,coming%20straight%20from%20the%20utility

Muller, O. (2021, June 28). Here's why Black TikTok creators are boycotting Megan Thee Stallion's new song. *TODAY.* https://www.today.com/tmrw/here-s-why-black-tiktok-creators-are-boycotting-megan-thee-t223706

Murphy, E. E., & Tong, J. (2019). The Racial Composition of Forensic DNA Databases. *California Law Review*, 108(6). https://doi.org/10.15779/Z381G0HV8M

Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., White, M., Zapert, K., & Turner, R. (2003). The impact of health information on the internet on the physician-patient relationship: patient perceptions. *Archives of internal medicine*, 163(14), 1727-1734. https://doi.org/10.1001/archinte.163.14.1727

Nader, R. (1965). *Unsafe at any speed.* Grossman Publishers.

National Conference on State Legislatures. (2014, August 5). *Forensic science database.* https://www.ncsl.org/research/civil-and-criminal-justice/dna-database-search-by-policy.aspx

National Highway Traffic Safety Administration. (2021). *2020 Fatality data show increased traffic fatalities*

during pandemic. https://www.nhtsa.gov/press-releases/2020-fatality-data-show-increased-traffic-fatalities-during-pandemic.

Nelkin, D. E. (1992). *Controversy: politics of technical decisions.* Sage Publications, Inc.

Ng, A., (2018, September 7). How the Equifax hack happened, and what still needs to be done. *CNET*. https://www.cnet.com/tech/services-and-software/equifaxs-hack-one-year-later-a-look-back-at-how-it-happened-and-whats-changed/

Nicholas, D., Boukacem-Zeghmouri, C., Xu, J., Herman, E., Clark, D., Abrizah, A., ... & Świgoń, M. (2019). Sci-Hub: The new and ultimate disruptor? View from the front. *Learned Publishing*, 32(2), 147-153. https://doi.org/10.1002/leap.1206

North American Regional Committee of the Human Genome Diversity Project. (1997). Proposed model ethical protocol for collecting DNA samples. *Houston Law Rev.* 33, 1431– 1473. http://www.stanford.edu/group/morrinst/hgdp/protocol.html

Oberhaus, D. (2019, December 10). Amazon, Google, Microsoft: Here's Who Has The Greenest Cloud. *Wired*. https://www.wired.com/story/amazon-google-microsoft-green-clouds-and-hyperscale-data-centers/

Oliveri, S., Ferrari, F., Manfrinati, A., & Pravettoni, G. (2018). A systematic review of the psychological implications of genetic testing: A comparative analysis among cardiovascular, neurodegenerative and cancer diseases. *Frontiers in Genetics*, 9(624). https://doi.org/10.3389/fgene.2018.00624

OpenAI. (2021, December 14). *Customizing GPT-3 for your application*. https://openai.com/blog/customized-gpt-3/

OpenAI. (2021, August 10). *OpenAI Codex*. https://openai.com/blog/openai-codex/

Oppy, G., & Dowe, D. (2021). The Turing test. *Stanford Encyclopedia of Philosophy.* http://seop.illc.uva.nl/entries/turing-test/

Ottinger, G. (2010). Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values*, 35(2), 244-270. https://doi.org/10.1177%2F0162243909337121

Palmer, A. (2020, October 24). How Amazon keeps a close eye on employee activism to head off unions. *CNBC*. https://www.cnbc.com/2020/10/24/how-amazon-prevents-unions-by-surveilling-employee-activism.html

Parthasarathy, S. (2007) *Building genetic medicine: Breast cancer, technology, and the comparative politics of health care.* The MIT Press.

Parthasarathy, S. (2010). Breaking the expertise barrier: understanding activist strategies in science and technology policy domains. *Science and Public Policy*, 37(5), 355-367. https://doi.org/10.3152/030234210X501180

Parthasarathy, S. (2017). *Patent politics.* University of Chicago Press.

Patterson, M. (n.d.). GPT-3 and AI in Customer Support. *Help Scout*. https://www.helpscout.com/blog/ai-in-customer-support/

Perrigo, B. (2022, February 14). Inside Facebook's African Sweatshops. *TIME*. https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.

Peterson, T. (2021, October 22). *Google looks to be a go in The Dalles*. Columbia Community Connection. https://www.columbiacommunityconnection.com/the-dalles/google/data/centers/tax-abatements/city-council

Pilon, M. (2019, April 29). Stop & Shop strike reveals concerns about job-killing technology. *Hartford Business Journal.* https://www.hartfordbusiness.com/article/stop-shop-strike-reveals-concerns-about-job-killing-technology

Pontis, S., Blandford, A., Greifeneder, E., Attalla, H., & Neal, D. (2017). Keeping up to date: An academic researcher's information journey. *Journal of the Association for Information Science and Technology*, 68(1), 22-35. https://doi.org/10.1002/asi.23623

Pritchett, W.E. (2003). The "public menace" of blight: urban renewal and the private uses of eminent domain. *Penn Law: Legal Scholarship Repository*. https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=2199&context=faculty_scholarship#:~:text=Blight%2C%20renewal%20proponents%20argued%2C%20was,the%20future%20of%20the%20city.

Privacy Considerations in Large Language Models. (n.d.). *Google AI Blog.* Retrieved January 31, 2022, from https://ai.googleblog.com/2020/12/privacy-considerations-in-large.html

Rabin, R.C. (2020, December 22). Pulse oximeter devices have higher error rate in Black patients. *The New York Times.* https://www.nytimes.com/2020/12/22/health/oximeters-covid-black-patients.html?searchResultPosition=1

Rabinow, P. (2011). *Making PCR: A story of biotechnology*. University of Chicago Press.

Race Forward: The Center for Racial Justice Innovation. (2014, January 22). *Moving the race conversation forward.* https://www.raceforward.org/research/reports/moving-race-conversation-forward

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9. http://www.persagen.com/files/misc/radford2019language.pdf

Rahman, K. (2020, April 24). Michigan man died after being repeatedly denied test just hours after his father died of Coronavirus, family say. *Newsweek*. https://www.newsweek.com/michigan-man-dies-coronavirus-repeatedly-turned-away-1499818Downloaded May 20, 2020.

Rainie, L., & Perrin, A. (2019, July 22). Key findings about Americans' declining trust in government and each other. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2019/07/22/key-findings-about-americans-declining-trust-in-government-and-each-other/

Ramsey, C.L. (2000). Ethics and Culture in the Deaf Community Response to Cochlear Implants. *Seminars in Hearing*, 21, 0075-0086.

Rapp, R. (1999). *Testing women, testing the fetus.* Routledge.

Rayome, A.D. (2016, September 19). Why data centers fail to bring new jobs to small towns. *Tech Republic.* https://www.techrepublic.com/article/why-data-centers-fail-to-bring-new-jobs-to-small-towns/

Reardon, J., & Princeton University. (2005). *Race to the finish : identity and governance in an age of genomics.* Princeton University Press, Cop.

Reinhardt, L.R. (2015). *Deaf-hearing interpreter teams: Navigating trust in shared space.* (Publication No. 21) [Master's Thesis, Western Oregon University]. Digital Commons. https://digitalcommons.wou.edu/theses/21/

Resnick, B., & Belluz, J. (2019, July 10). The war to free science: how librarians, pirates, and funders are liberating the world's academic research from paywalls. *Vox*. https://www.vox.com/the-highlight/2019/6/3/18271538/open-access-elsevier-california-sci-hub-academic-paywalls

Roach, J. (2020, September 14). Microsoft finds underwater datacenters are reliable, practical, and use energy sustainably. *Microsoft News*. https://news.microsoft.com/innovation-stories/project-natick-underwater-datacenter/

Roberts, S.T. (2021) *Behind the screen: Content moderation in the shadows of social media.* Yale University Press.

Robinson, R. (2021, November 16). Boeing built an unsafe plane, and blamed the pilots when It crashed. *Bloomberg Businessweek*. https://www.bloomberg.com/news/features/2021-11-16/are-boeing-planes-unsafe-pilots-blamed-for-corporate-errors-in-max-737-crash

Romero, A. (2021, June 24). Can't access GPT-3? Here's GPT-J-- Its open-source cousin. *Towards Data Science.* https://towardsdatascience.com/cant-access-gpt-3-here-s-gpt-j-its-open-source-cousin-8af86a638b11

Romero, A. (2021, June 5). GPT-3 scared you? Meet Wu Dao 2.0: A monster of 1.75 trillion parameters. *Towards Data Science.* https://towardsdatascience.com/gpt-3-scared-you-meet-wu-dao-2-0-a-monster-of-1-75-trillion-parameters-832cd83db484

Rousseau, A., Baudelaire, C., Riera, K. (2020, October 27). Doctor GPT-3: hype or reality? *Nabla*. https://www.nabla.com/blog/gpt-3/

Royal Canadian Mounted Police. (2014, January 24). Critical Infrastructure Intelligence Assessment: Criminal threats to the Canadian Petroleum industry. *Ottawa: RCMP.* https://www.statewatch.org/media/documents/news/2015/feb/can-2014-01-24-rcmp-anti-petroleum-activists-report.pdf

Sale, K. (1995) *Rebels against the future: The luddites and their war on the industrial revolution: Lessons for the computer age.* Addison-Wesley Publishing.

Savaresi, A., & McVey, M. (2020, February 7). *Human rights abuses by fossil fuel companies.* 350. https://350.org/climate-defenders/

Savero, R. (1981, February 4). Air Force Academy to drop its ban on applicants with sickle-cell gene. *The New York Times.* https://www.nytimes.com/1981/02/04/us/air-academy-to-drop-its-ban-on-applicants-with-sickle-cell-gene.html

Scareflow, A. [@vboykis]. (2020, August 2). NLP People: Anyone know what these two Books1 and Books2 data sources in GPT-3 are? Writing a newsletter about them...[Tweet]. Twitter. https://twitter.com/vboykis/status/1290030614410702848?lang=en

Schiermeier, Q. (2021). Forensic database challenged over ethics of DNA holdings. *Nature*, 594(7863), 320–322. https://doi.org/10.1038/d41586-021-01584-w

Schiffer, Zoe (2021, February 19). Google fires second AI ethics researcher following internal investigation. *The Verge.* https://www.bbc.com/news/technology-56135817

Schmitt, A. (2020). *Right of way: Race, class, and the silent epidemic of pedestrian deaths in America.* Island Press: Washington, DC.

Schurman, R., & Munro, W. A. (2010). *Fighting for the future of food: Activists versus agribusiness in the struggle over biotechnology* (Vol. 35). U of Minnesota Press.

Schwartz, J. (2013, January 12). Internet activist, a creator of RSS, is dead at 26, apparently a suicide. *The New York Times.* https://www.nytimes.com/2013/01/13/technology/aaron-swartz-internet-activist-dies-at-26.html

Scott, D., & Barnett, C. (2009). Something in the air: civic science and contentious environmental politics in post-apartheid South Africa. *Geoforum*, 40(3), 373-382. https://doi.org/10.1016/j.geoforum.2008.12.002

Seabrook, J. (2019, October 14). The Next Word. *The New Yorker.* 95 (31). https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker

Select Committee on Artificial Intelligence. (2019, June). *The national artificial intelligence research and development strategic plan: 2019 update.* NITRD. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

Selin, C. (2011). Negotiating plausibility: Intervening in the future of nanotechnology. *Science and Engineering Ethics*, 17, 723-737. https://doi.org/10.1007/s11948-011-9315-x

Selinger, E., & Durant, D. (2021). Amazon's Ring: Surveillance as a Slippery Slope Service. *Science as Culture*, 1-15. https://doi.org/10.1080/09505431.2021.1983797

Selsky, A., and Valdes, M. (2021, October 25). Big tech data centers spark worry over scarce western water. *Associated Press.* https://apnews.com/article/technology-business-environment-and-nature-oregon-united-states-2385c62f1a87030d344261ef9c76ccda

Semuels, A. (2016, March 18). The role of highways in American poverty. *The Atlantic.* https://www.theatlantic.com/business/archive/2016/03/role-of-highways-in-american-poverty/474282/

Severin, A., & Chataway, J. (2021). Overburdening of peer reviewers: A multi-stakeholder perspective on causes and effects. *Learned Publishing*, 34(4), 537-546. https://doi.org/10.1002/leap.1392

Shapin, S. (1995). *A social history of truth: Civility and science in seventeenth-century England.* University of Chicago Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump.* Princeton University Press.

Shelton, S. A., & Brooks, T. (2019). "We need to get these scores up": A narrative examination of the challenges of teaching literature in the age of standardized testing. *Journal of Language and Literacy Education*, 15(2), n2. https://eric.ed.gov/?id=EJ1235207

Shill, G. (2020). Should law subsidize driving? *NYU Law Review*, 95, 498-579.

Siddik, M.A.B., Shehabi, A., and Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters, 16.* https://iopscience.iop.org/article/10.1088/1748-9326/abfba1/pdf

Simon, C. M., L'Heureux, J., Murray, J. C., Winokur, P., Weiner, G., Newbury, E., Shinkunas, L., & Zimmerman, B. (2011). Active choice but not too active: Public perspectives on biobank consent models. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 13(9), 821–831. https://doi.org/10.1097/GIM.0b013e31821d2f88

Simonite, T. (2021). What really happened when Google ousted Timnit Gebru? *Wired.* https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/.

Singer, J. (2022). *There are no accidents: The deadly rise of injury and disaster—who profits and who pays the price.* Simon and Schuster.

Singh, R. & Jackson, S. (2021). Seeing like an infrastructure: Low-resolution citizens and the Aadhaar identification project. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-26. https://doi.org/10.1145/3476056

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25), 2477-2478. https://www.nejm.org/doi/full/10.1056/nejmc2029240

Solaiman, I., & Dennison, C. (2021). Process for adapting language models to society (palms) with values-targeted datasets. *Advances in Neural Information Processing Systems*, 34.

Sovacool, B.K. (2021, March). When subterranean slavery supports sustainability transitions? power, patriarchy, and child labor in artisanal Congolese cobalt mining. *The Extractive Industries & Society*, 8(1), 271-293. https://www.sciencedirect.com/science/article/pii/S2214790X20303154

Spice, A. (2018). Fighting invasive infrastructures, *Environment and Society*, 9(1), 40-56. https://doi.org/10.3167/ares.2018.090104

Spinney, Laura (2022, January 9). Are we witnessing the dawn of post-theory science? *The Guardian.* https://www.theguardian.com/technology/2022/jan/09/are-we-witnessing-the-dawn-of-post-theory-science?CMP=Share_iOSApp_Other

Stanford CRFM. (n.d.) *Developing and understanding responsible foundation models.* Retrieved March 3, 2022 from https://crfm.stanford.edu/

Stanford HAI. (n.d.) *Corporate members program.* Retrieved March 3, 2022 from https://hai.stanford.edu/about/corporate-members-program

Stangeland, C. (2016, December). *Fracking: Unintended consequences for local communities.* Homeland Security Affairs. https://www.hsaj.org/articles/13753

Starr, P. (1982). *The social transformation of American medicine.* Basic Books.

Stern, J. (2021, May 28). Pipeline of violence: The oil industry and missing and murdered Indigenous women. *Immigration and Human Rights Law Review.* https://lawblogs.uc.edu/ihrlr/2021/05/28/pipeline-of-violence-the-oil-industry-and-missing-and-murdered-indigenous-women/#post-274-footnote-ref-6

Stevenson, F. A., Kerr, C., Murray, E., & Nazareth, I. (2007). Information from the Internet and the doctor-patient relationship: the patient perspective–a qualitative study. *BMC family practice*, 8(1), 1-8. https://doi.org/10.1186/1471-2296-8-47

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580. https://doi.org/10.1016/j.respol.2013.05.008

Stirling, A. (2008). 'Opening up' and 'closing down': Power, participation, and pluralism in the social appraisal of technology. *Science, Technology, and Human Values*, 33(2), 262-294. https://doi.org/10.1177%2F0162243907311265

Stix, C. (2020). (C. Brasoveanu, Ed.). *European Commission.* https://ec.europa.eu/jrc/communities/sites/jrccties/files/reportontheeuropeanailandscapeworkshop.pdf

Stokstad, E. (2019). Major U.K. genetics lab accused of misusing African DNA. *Science*. https://doi.org/10.1126/science.aba0343

Strong, J., Ryan-Mosley, T., Cillekens, E., Hao, K., Reilly, M., and Lichfield, G. (2020, December 2). Podcast: Facial recognition is quietly being used to control access to housing and social services. *MIT Technology Review*. https://www.technologyreview.com/2020/12/02/1012901/no-face-no-service/

Stumpf, J. (2018). *How to understand urban blight in America's neighborhoods and work to eliminate it.* Dickinson Magazine. https://www.dickinson.edu/news/article/3461/how_to_understand_urban_blight_in_america_s_neighborhoods_and_work_to_eliminate_it

Summers-Trio, P., Hayes-Conroy, A., Singer, B., & Horwitz, R. I. (2019). Biology, biography, and the translational gap. *Science Translational Medicine*, 11(479). https://doi.org/10.1126/scitranslmed.aat7027

Swanson, K. W. (2009). The emergence of the professional patent practitioner. *Technology and Culture,* 50(3), 519-548. https://www.jstor.org/stable/40345727

Swift, A. (2019). Soviet-American Rivalry at Expo '58. In Molella, A. P., & Knowles, S. G. (Eds.). (2019). *World's Fairs in the Cold War: Science, technology, and the culture of progress.* (pp. 27-45) University of Pittsburgh Press.

Syverson, B., (2020, October 19). The rules of brainstorming change when artificial intelligence gets involved. Here's How. *IDEO*. https://www.ideo.com/blog/the-rules-of-brainstorming-change-when-ai-gets-involved-heres-how?utm_content=143682113&utm_medium=social&utm_source=twitter&hss_channel=tw-23462787

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *arXiv*. https://arxiv.org/abs/2102.02503

Tan, S.S.L., & Goonawardene, N. (2017). Internet health information seeking and the patient-physician relationship: a systematic review. *Journal of Medical Internet Research*, 19(1), https://doi.org/10.2196/jmir.5729

Terry, S. F., Terry, P. F., Rauen, K. A., Uitto, J., & Bercovitch, L. G. (2007). Advocacy groups as research organizations: the PXE International example. *Nature Reviews Genetics*, 8(2), 157-164. https://doi.org/10.1038/nrg1991

The British Psychological Society. (2017). *Working with interpreters: Guidelines for psychologists.* https://www.bps.org.uk/news-and-policy/working-interpreters-guidelines-psychologists

The City of Ann Arbor. (n.d.) *Drinking water*. Retrieved March 13, 2022 from https://www.a2gov.org/departments/systems-planning/planning-areas/water-resources/Pages/Drinking-Water.aspx

The Eye. (2020). *Enter the Eye: An open directory data archive.* https://the-eye.eu/

The International HapMap Consortium. (2004). Integrating ethics and science in the International HapMap Project. *Nature Reviews Genetics*, 5(6), 467–475. https://doi.org/10.1038/nrg1351

The Parliament of the Commonwealth of Australia. (2021). *Security legislation amendment (critical infrastructure) bill no. , 2021: A bill for an act to amend legislation relating to critical infrastructure, and for other purposes.* House of Representatives. https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/r6657_aspassed/toc_pdf/20182b01.pdf;fileType=application%2Fpdf

Tucker, B. P. (1998). Deaf culture, cochlear implants, and elective disability. *Hastings Center Report,* 28(4), 6-14. https://doi.org/10.2307/3528607

Turing, A. M. (1950). Computing Machinery and Intelligence. *Creative Computing*, 6(1), 44-53.

U.S. Copyright Office Review Board. (2022, February 14). *Re: Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise* (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071) [Letter]. https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf

U.S. Geological Survey. (2018, June 8). *Mining and water quality.* USGS Water Science School. https://www.usgs.gov/special-topics/water-science-school/science/mining-and-water-quality

United Food and Commercial Workers International Union. (2020, August 24). *UFCW Statement on Amazon Cashierless Grocery Store Opening.* UFCW. https://www.ufcw.org/press-releases/cashier/

United Nations Environment Programme. (2011). *Environmental assessment of Ogoniland*. https://postconflict.unep.ch/publications/OEA/UNEP_OEA.pdf

Van Noorden, R. (2021, February 3). Scientists call for fully open sharing of coronavirus genome data. *Nature*, 590(7845),195-196. doi: https://doi.org/10.1038/d41586-021-00305-7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems* (pp. 5998-6008). https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vats, A. (2020). *The color of creatorship: Intellectual property, race, and the making of Americans.* Stanford University Press.

Vézina, B., & Moran, B. (2020, August 10). Artificial Intelligence and creativity: Why we're against copyright protection for AI-generated output. *Creative Commons Blog*. https://creativecommons.org/2020/08/10/no-copyright-protection-for-ai-generated-output/.

Viable. (n.d.) https://askviable.com/

Vincent, J. (2021, May 25). Microsoft has built an AI-powered autocomplete for code using GPT-3. *The Verge*. https://www.theverge.com/2021/5/25/22451144/microsoft-gpt-3-openai-coding-autocomplete-powerapps-power-fx

Wang, Z., Rodriguez Morales, M.M., Husak, K. Kleinman, M., & Parthasarathy, S. (2021). *In Communities We Trust: Institutional Failures and Sustained Solutions for Vaccine Hesitancy.* https://stpp.fordschool.umich.edu/research/research-report/communities-we-trust-institutional-failures-and-sustained-solutions

Welbl, J., Glaese, A., Uesato, J., Dathathri, S., Mellor, J., Hendricks, L. A., Anderson, K, Kohli, P., Coppin, B., and Huang, P. S. (2021). Challenges in detoxifying language models. *arXiv*. https://arxiv.org/abs/2109.07445

West, D. (2019). *The Future of Work: Robots, AI, and Automation.* Brookings Institution Press.

White, R. F. (2007). Institutional Review Board mission creep: The common rule, social science, and the nanny state. *The Independent Review*, 11(4), 547–564. http://www.jstor.org/stable/24562415

Whyte, K.P. (2017, February 28). The Dakota Access Pipeline, environmental injustice, and U.S. colonialism. *Red Ink: An International Journal of Indigenous Literature, Arts, & Humanities*. 19(1), https://ssrn.com/abstract=2925513.

Willis, D. E., Andersen, J. A., Bryant-Moore, K., Selig, J. P., Long, C. R., Felix, H. C., ... & McElfish, P. A. (2021). COVID-19 vaccine hesitancy: Race/ethnicity, trust, and fear. *Clinical and translational science*, 14(6), 2200-2207. https://doi.org/10.1111/cts.13077

Working Group on Mining and Human Rights in Latin America. (2014, May 22). *The impact of Canadian mining in Latin America and Canada's responsibility - Executive summary of the report submitted to the Inter-American Commission on Human Rights.* Due Process of Law Foundation. https://www.dplf.org/en/resources/impact-canadian-mining-latin-america-and-canadas-responsibility-executive-summary

Wyden, R. (2022). *Wyden, Booker and Clarke Introduce Algorithmic Accountability*

Act of 2022 To Require New Transparency And Accountability For Automated Decision Systems. Press Release.

Zabel, J. (2019). The Killer Inside Us: Law, Ethics, and the Forensic Use of Family Genetics. *Berkeley Journal of Criminal Law*, 24(2). https://doi.org/10.15779/Z385D8NF7

Zeavin, H. (2021). *The distance cure: A history of teletherapy*. MIT Press.

Zhang, K. (2018, December 13). How big data has created a big crisis in science. *The Conversation.* https://theconversation.com/how-big-data-has-created-a-big-crisis-in-science-102835

Zhavoronkov, A. (2021, July 19). Wu Dao 2.0 - bigger, stronger, faster AI from China. *Forbes*. https://www.forbes.com/sites/alexzhavoronkov/2021/07/19/wu-dao-20bigger-stronger-faster-ai-from-china/?sh=3020dbf16fb2

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *Proceedings of the IEEE international conference on computer vision* (pp. 19-27).

Zou, Y., Mhaidli, A. H., McCall, A., & Schaub, F. (2018). "I've got nothing to lose": Consumers' risk perceptions and protective actions after the Equifax data breach. *Fourteenth Symposium on Usable Privacy and Security*, 197-216. https://www.usenix.org/conference/soups2018/presentation/zou

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power.* Profile Books.

For Further Information

If you would like additional information about this report, the Technology Assessment Project, or University of Michigan's Science, Technology, and Public Policy Program, you can contact us at **stpp@umich.edu** or **stpp.fordschool.umich.edu**.

LEARN MORE

myumi.ch/LLMReport

GERALD R. FORD SCHOOL OF PUBLIC POLICY
SCIENCE, TECHNOLOGY, AND PUBLIC POLICY
UNIVERSITY OF MICHIGAN

Technology Assessment Project
Science, Technology, and Public Policy Program
Gerald R. Ford School of Public Policy
University of Michigan
735 S. State Street
Ann Arbor, MI 48109

(734) 764-0453
stpp.fordschool.umich.edu
stpp@umich.edu

Brain/circuit icon: Freepik