EXECUTIVE SUMMARY

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

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.

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.

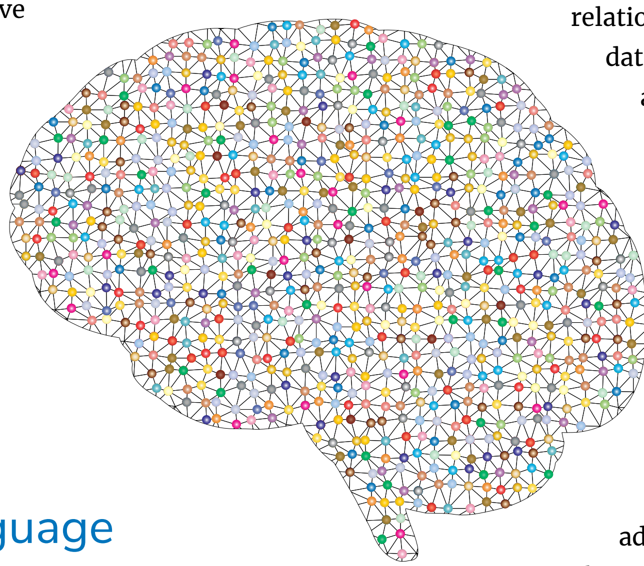
However, LLMs have already generated serious concerns. Because they are trained on text from old books and webpages, LLMs reproduce historical biases and hateful speech towards marginalized communities. They also require enormous amounts of energy and computing power, and thus are likely to accelerate 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 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

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.

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

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.

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.

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'

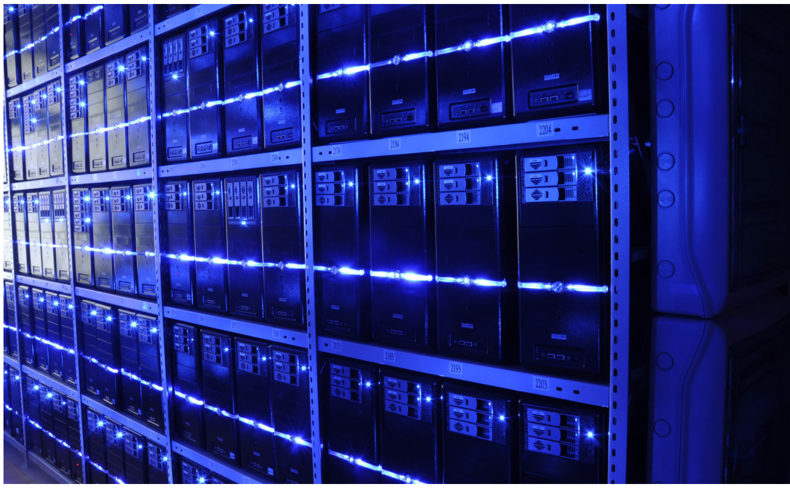
PDF: Click to return to top

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 transformation 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.



Credit: Baltic Servers

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

PDF: Click to return to top

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.

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.

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.

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.

VIEW THE FULL REPORT

myumi.ch/LLMReport

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.

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