

A Framework for Centralizing forces in AI

The Core Idea

This article is mainly about Large Language Models (LLMs), which are revolutionizing access to information and data, while the development and deployment of LLMs are predominantly controlled by major tech firms, leading to a centralization of power and technology, which are often inaccessible to the general public. This paper explores the implications of this centralization and examines how it contrasts with the decentralizing potential of the internet, which broadly empowers users by providing easier access to information. We propose a survey to assess how LLMs impact user empowerment and learning efficiency, which might work in a decentralizing way. This in turn can lead to bottom-up democratization of organizations. Through this analysis, we aim to highlight considerations for a more inclusive approach to LLM technology, ensuring that its benefits can be more equitably distributed across society.

Description of your threat scenario

We investigate the threat that LLMs can have on centralizing particular structures in society, and its effect of democratization of organizations.

Description of your demonstration

This is theoretical work. We have included a research design for a survey.

Description of your extrapolation into the future

After taking the fundamental concept of centralization and analyzing how LLMs have an effect on this process, we can extrapolate this process into the future.

Description of mitigation strategies

The outcome of the survey can be used to improve user empowerment metrics for LLMs. Improving user empowerment is a bottom-up way to mitigate threats to democracy.

Appendix

Introduction

Large Language Models represent a significant advancement in artificial intelligence, offering capabilities that span complex query answering to generating human-like textual content (Team EMB. 2024). Despite their potential, the accessibility of LLMs remains limited, largely due to the extensive computational resources, vast and diverse datasets, and specialized knowledge required for training these models (Aruna Pattam, 2024). This paper discusses the barriers that prevent individuals from self-training LLMs, highlighting the resulting trend towards centralization where only well-resourced entities control these powerful tools. We examine the implications of this centralization and propose considerations for fostering a more inclusive LLM technology landscape.

Democracy and Centralization

Centralization refers to the concentration of control of an activity or organization under a single authority (Oxford Languages, 2024). Decentralization means the transfer of control of an activity or organization to several local offices or authorities rather than one single one (Oxford Languages, 2024). Widely individually accessible and used LLMs, are developed by major firms like Microsoft and OpenAI. They are centralized models which means they are designed, trained, and owned by these companies. This centralization of development and ownership can increase these companies' power over the public (Andreas J, 2023).

Democracy refers to control of an organization or group by the majority of its members (Oxford Languages, 2024). LLMs lower the barriers for people getting knowledge and information and provide a low-cost and efficient way for learning. This leads to more people learning things, which distributes knowledge and thus empowerment to make decisions, thus being a decentralizing force. A decentralization leads to more people having power in an organization, thus being more democratic.

The centralization of Large Language Models (LLMs) appears inevitable due to the concentrated ownership of computational resources and advanced algorithms. This centralization poses challenges for democratizing access to LLMs. Currently, about 60% of the global population has internet access, a prerequisite for utilizing LLMs (Roser, Ritchie, & Ortiz-Ospina, 2023). However, effectively using LLMs also requires specific technical knowledge and access to well-curated data, which are not widely available. Although large tech companies offer user-friendly LLM tools, these are not equivalent to full access or control over LLM capabilities. For a democratic approach to LLMs, the criteria for data input and model training must be transparent and adhere to democratic principles, potentially enabling users to receive unbiased information. For the sake of quantifiability, we will focus on user empowerment for now.

LLMs and User empowerment

The internet had a lot of potential to decentralize the source of information. Anyone with an internet connection can now learn things on Wikipedia and open courseware (Ritchie, Mathieu, Roser, & Ortiz-Ospina, 2023). This allows people who do not live close to learning institutions or who do not have the money, to enroll in them to still learn. This potentially brings the tools of knowledge in the hands of more people. One can imagine that LLMs can have an analogous effect. LLMs can explain concepts and can drag people through the learning process in a more tailored way than google or Wikipedia can. But do LLMs benefit users equally? If

the better models are behind a paywall, the people with more resources can benefit more from them. And to take ChatGPT as an example: it takes some education to optimally use it. When asking it to write code for example, it is too easy to just let it run, and if the program gives bugs, paste the error messages into the interface. It is too easy to stop thinking, and consequently stop learning.

This might lead to 2 distinct groups of users: one group that fills in prompts thoughtlessly and derives less value from using ChatGPT than when they would not have used it. And there might be a group that knows how to leverage the power of ChatGPT and uses it to learn more than they would have otherwise done. And there might be people who sometimes use ChatGPT mindlessly, and sometimes know how to use it (depending on the specific field of their inquiry). This might lead to a divide. We have also seen this with the use of the internet in general: where some people have fallen prey to spending way too much time on social media and googling things mindlessly, whereas others have used it to learn more than was possible ever before (Dutton, 2013). Just by making the space of possibilities bigger (you can learn much less, as well as much more, and everything in between), the possibility arises for an increase in inequality of knowledge. To test how empowering LLMs can be for learning, we propose a survey design.

Due to the short duration of the hackathon, we cannot get enough data, so this is just a proposal. The goal is to find out for what percentage of the population different LLMs are empowering to which degrees. We want to see how this is different for different age groups. By analyzing how different users are empowered by different LLMs, we can estimate which LLMs work to give the most value to people, and empower them with the powerful tool that is information. Here we define empowerment as the process of becoming stronger and more confident, especially in controlling one's life and claiming one's right (Oxford Languages, 2024).

Survey design

Target group: Everyone who uses LLMs.

Screening Questions

Are you 18 years or older? If not, thank you for your time, we are searching for people who are 18 years or older.

Are you using LLMs (like ChatGPT)? Yes/No

What is your age group? (18-25, 26-40, 40-60, 60-80, 80+).

Which LLM are you using most regularly?

For this LLM, please answer the following questions:

Quantitative questions

How much do you use it on average (hours a week)?

How much does your productivity increase when using this LLM on average? (0-100 percent).

Describe your increase/decrease in personal empowerment by the LLM, on a scale of -10 to 10? (-10 = very disempowered, 10 = very empowered).

Qualitative questions

Explain how your productivity increases or decreases when using this LLM.

Explain how your empowerment increases or decreases when using this LLM.

For the main areas that you use LLMs for, explain how it improves your productivity, from -10 to 10: with -10 meaning that it decreases your productivity dramatically, and a 10 that dramatically improves productivity. Give a separate explanation and rating for each area.

Conclusions

There are many forces that the LLM revolution brings with it that either centralize or decentralize specific structures in society. We decided to look at one of these, and write a research design proposal that can be readily executed. This survey can be distributed and can give insight into how different LLMs can lead to user empowerment. By analyzing how different users are empowered by different LLMs, we can estimate which LLMs work to give the most value to people, and empower them with the powerful tool that is information, giving more people more agency in the organizations they are part of. This is the core of bottom-up democratization.

Sources

Andreas J. (2023). Journal. <https://journals.sagepub.com/doi/pdf/10.1177/20563051231186353>

Aruna Pattam. (2024). Internet. Retrieved from <https://medium.com/ai-horizons/navigating-the-world-of-language-models-large-vs-small-models-b12077872db2>

Dutton, W. H. (2013). The Oxford Handbook of Internet Studies, Ch. 7. OUP Oxford.

Oxford Languages. (2024). Oxford languages. Oxford Languages; Oxford University Press. <https://languages.oup.com/google-dictionary-en/>

Ritchie, H., Mathieu, E., Roser, M., & Ortiz-Ospina, E. (2023). Internet. OurWorldInData.org. Retrieved from <https://ourworldindata.org/internet>

Team EMB. (2024). Internet. Emb.global. Retrieved from <https://blog.emb.global/what-are-large-language-model/>