Causality in the time of LLMs: Round table discussion results of CLeaR 2023

Cheng Zhang, Dominik Janzing, Mihaela van der Schaar, Francesco Locatello, Peter Spirtes, Kun Zhang, Bernhard Schölkopf, Caroline Uhler and CleaR 2023 attendees in this session

Abstract

The field of machine learning and AI has witnessed remarkable breakthroughs with the emergence of LLMs, which have also sparked a lively debate in the causal community. As researchers in this field, we are interested in exploring how LLMs relate to causality research, and how we can leverage the technology to advance it. In the second conference of Causal Learning and Reasoning (CLeaR), 2023, we held a round table discussion to gather and integrate the diverse perspectives of the CLeaR community on this topic.

There is a general consensus that LLMs are not yet capable of causal reasoning at the current stage but has a lot of potential with public available information by CLeaR 2023. Enhancing causal machine learning is vital not only for its own sake but also to help LLMs improve their performance, especially regarding trustworthiness. In this document, we present both the summary and the raw outcome of the round table discussion. We acknowledge that with the progress of both fields, the opportunities and impact may rapidly change. We will repeat the same exercise in CLeaR 2024 to document the evolution.

1. The round-table discussion set up

In CleaR 2023, we have organized 1 hour round table discussion. People are randomly grouped into 5 to 10 people groups to discuss the following two questions.

- How do advances in LLMs disrupt causal research?
- How can we as a community utilize and contribute to the advances in LLMs?

Each group is required to submit their answer independently through a form as shown in 1. We emphasize that we welcome diverse opinions, and people are not required to agree on the same answer within the group. We aim to capture the community's thoughts regarding these two questions. The organizers have noticed that people have quite different experiences with LLMs. Some are using LLMs on an almost daily basis, while others have hardly tried them. We did not give people extra time to experiment with LLMs but rather asked them to express their current views.

2. Insights

We observe very diverse opinions regarding both questions from the raw answers shown in section 4. And many answers have pointed to great research opportunities. We highlighted some of these insights below and recommend the reader to check section 4 for details.

How do advances in LLMs disrupt causal research? Some of the benefits of LLM are that they can help causal algorithms by extracting useful information from various domains (group 1, 6), summarizing human opinions and common sense (group 6). They can also serve as a versatile tool for data processing, information parsing, and model development (group 1, 2, 3, 4 and 9).

However, there are also some challenges. One is that causality aims to provide new insights and overcome the limitations of human intuition. The current LLMs, which rely on human knowledge, may not be enough for this purpose. (group 5, 6, 8) Another is that the risk of generating misleading ”causal” answers is high and harmful (group 4, 8). Finally, there is also a concern about the potential negative impact on other research directions due to the popularity of LLMs and the fact that causality as a research field is still developing (group 8).

This also creates some opportunities. It motivates the community to rethink how to evaluate models (group 9) and design evaluation methods for non-traditional causal methods. It also brings more attention to the text-based causal research (group 3, 7).

How can we as a community utilize and contribute to the advances in LLMs? Causality can provide many opportunities for enhancing LLMs, and many experts concur on this. Some of the main opportunities are: First, causality can make LLMs more trustworthy in general (group, 1,2,3,7), by making the results more explainable, understand the underlying mechanism of the model, improve the robust of the model. Secondly, it can also deal with the problem of selection bias in the training data. Finally, as LLMs currently not sufficient for causal reasoning, there is a lot of potential for adding such capabilities. This can be done either by incorporating external causal components (group 4, 7, 9) or by taking into account new experimental data (group 8) and feedback (group 4, 9).

3. Related Work

Other related work regarding how LLMs disrupt causal research include Willig et al. (2022); Long et al. (2023); Zhang et al. (2023); Kıcıman et al. (2023). There is a positive correlation of the sentiment and the publication date of these papers. Beyond these LLMs focused papers, there is large body of research work considering generic deep learning and causality, for example, Schölkopf et al. (2021); Jin et al. (2022); Zhou et al. (2023); Berrevoets et al. (2023). Some of these insights are valid considering LLMs as well.

4. Input from different groups in CleaR 2023

In this section, we try to capture the raw input from each group from the submitted form.

4.1. Group 1

- **Name:** Karel D’Oosterlinck, Paloma Rabaey, William Orchard, Michael Knaus, Luis Enrique Sucar, Alexander Marx

- **How do advances in LLMs disrupt causal research?**
  - LLMS can act as research assistants: - Extract structure from unstructured data to input in causal algorithms - Help build clinical trials by searching for relevant literature and suggesting e.g. confounders to account for based on existing research. - Help summarize and suggest research questions.

- LLMs could one day act as causal engines: - Actually write code, manipulate data in order to make causal discoveries. Not sure here if the LLM will actually be better than the existing causal algorithms. - Hopefully LLMs can get some commonsense understanding about a domain and apply that to specific settings.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  - The causal community can contribute by developing new explanation methods. Behavioral studies of LLMs will not tell us everything. We need methods to understand model internal representations and how models behave under interventions. This will give us more insight into what mechanisms are learned by LLMS. - We need to think about how we can inject existing causal knowledge into LLMs. Current models learn almost everything in a self-supervised fashion, even on domains where we already know the causal structure. Models could be more robust and more efficient if they incorporate the known causal models that we humans have already discovered and studied.

4.2. Group 2

- **Name:** Yuan Xue, Azlaan Mustafa Samad, Jesse Krijthe, Luca Castri, Andreas Sauter, Wei Zhang, Johanna Schrader
- **How do advances in LLMs disrupt causal research?**
  As a tool, how could we use it to do research, like writing scripts. In long-term, GPT model could ba large enough to catch causality.
- **How can we as a community utilize and contribute to the advances in LLMs?**
  Making the auto regressive or attention machnism causal. Using causal tools for explainable LLMs.

4.3. Group 3

- **Name:** Yanai Elazar, Kirtan Padh, Francesco Locatello, Sara Magliacane, Atticus Geiger, Connor Jerzek, Tobias Sikosek, Yasser Taha
- **How do advances in LLMs disrupt causal research?**
  One disruption is that it makes text-based causal inference more prominent in causal research, and LLMs can be used to generate text-based counterfactuals. Another one is that LLMs can be used as a neat way to merge different datasets that can be then exploited by causal algorithms, improving what you can learn single dataset.
- **How can we as a community utilize and contribute to the advances in LLMs?**
  Causality can improve interpretability, fairness, and robustness of LLMs, suggest new training regimes to address weak points and help LLMs to deal with selection bias. Causality can help also understand the causal relationship between training data and the (emergent) behavior model. Finally, causality research can also help infuse causal reasoning in LLMs.

4.4. Group 4

- **Name:** Marco Simnacher, David Kaltenpoth, Marco Heuvelman, Maurice Frank, Vivian Nastl

- **How do advances in LLMs disrupt causal research?**
  - Does the success of LLMs prove that scaling simple and predictive models trained on gigantic amounts of data can render explicit causal modeling e.g. for reaching human level intelligence unnecessary?

  - Many causality papers are about telling convincing "stories" supported by appropriate evidence. The ability of LLMs to also tell convincing stories that appear to be causal is potentially misleading/harmful.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  Utilizing LLMs: - data annotation/analysis/interpretation (to varying degrees of trustworthiness) - counterfactual chat data generation (reset and take different conversation "paths") - just use it as a text tool (rewriting text etc)

  Contributions to LLM: - e.g., doWhy Plugin such that LLMs have access to causal tools - quantifying causal capabilities of LLMs (or show that there is no rigorous causality) - add causal training/feedback

4.5. Group 5

- **Name:** Sanghack Lee, Jonas, Tomonori Izumitani, Jakob Zeitler

- **How do advances in LLMs disrupt causal research?**
  Only if they can surpass the limitations of human causal reasoning such as confounder or mediator identification Etc.

  Only for small tasks like creating graphs

  Not in near future.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  Feed it more causal data, like graphs etc so it's a better causal search engine for confounder etc

  Learn causal relationships beyond correlation

  Causal constraints in the transformer? Doesn't need to be perfect

4.6. Group 6

- **Name:** Bijan Mazaheri, Phillip Faller, Yuejiang Liu, Riccardo Massidda, Kseniia Soloveva, Georg Manden, Luco Torresi

- **How do advances in LLMs disrupt causal research?**
  A strength of the field of Causal Inference is its ability to give new insights that contradict human intuition. In this sense, LLMs are limited as summary of human opinions and are susceptible to many of the same pitfalls.

  On the other hand, LLMs have a strong ability to summarize and synthesize domain knowledge. This may help in the determination of DAG structures which are only partially known by specific humans, but fully known by a community. For example, we could train an LLM on the entirety of bioRxiv to try to summarize the causal opinions of the field.

  We are also interested in whether LLMs could help generate hypothesis for research.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  We believe the field of causality could be useful in helping teach LLMs insights from our field. Causality experts can help determine example prompts for LLMs to teach them how to reason causally.

4.7. Group 7

- **Name:** Wojeciech Niemiro, Dominik Janzing, Cecilia Casolo, Zhufeng Li, Lukasz Rajkowski, Mate Dravucz, Fabio Zennaro

- **How do advances in LLMs disrupt causal research?**
  1. Discover the property of words, but not the property of the language 2. Our causal reasoning is already reflected in language 3. LLM lacks statistical basic 4. Assessing the capability of causal reasoning in chatJPT and LLMs in general make us rethink what causality really means.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  1. Help to understand the limit of causal reasoning in LLM, 2. Incorporate causal library and run the code to conclude, 3. Lack of domain shifting capabilities. 4. Causality is more concern with experiments (intervention), LLM is still very behind. 5. We can help to formulize the causality concept to LLM.

4.8. Group 8

- **Name:** Ricardo Cannizzaro, Mario Figueiredo, Alicia Curth, Eli Y. Kling, Sergey Plis, Thomas Trappenberg, Oscar Clivio, Rhys Howard

- **How do advances in LLMs disrupt causal research?**
  • It takes the oxygen out of the academic discussion. All funding and attention is being diverted to LLMs. This is, currently, detracting from research efforts. This may change once people think LLMs are solved. • We will lose people as they follow the hype. • This follows the tech hype-curve that we've seen many times before. Neural Networks, self-driving cars, crypto. LLM is just reaching its hype-curve now. Causality is still quite new; there is good foundational theory but we're still as a community trying to figure out how we can make it useful; what is the application? As we try to work this out, we'll gain more momentum, this will continue until we hit some critical mass and then enter the rapid increase of the hype curve where we will see a lot more attention. It is also more accessible to the general public, which is helped in its popularisation. • External factors influence researchers' interest in causality: big tech are hiring researchers and interns that are doing Deep Learning, so everyone wants to do that. Also, the funding is going towards DL so that's where the research efforts are going. • LLMs are not being genuinely marketed to mass-media; tech companies are pretending they are more capable than they actually are. So, the average person is misunderstanding ChatGPT, and treat is as a human-level intelligence. This has led to people over-trusting ChatGPT, and this has led to tragic consequences, such as a person dying by suicide after developing a close relationship over a significant time and having the ChatGPT recommend that course of action. • Which are the industries that first adopt disruptive technology? o Gaming, entertainment industry, porn

- **How can we as a community utilize and contribute to the advances in LLMs?**
  • LLMs - Context: o LLMs are purely associational. There are no causal models. o Human feedback reinforcement learning o It appears causal but there is no reasoning happening. o LLMs are predictive models – they just predict which words come next, and are trying to give responses that closely approximate human responses. There was agreement on this, but some felt that it was a reductive statement, and thought that ChatGPT is more than just a relational database. • Can causal information be added to the training set to boost responses? • Can we add causal reasoning into the LLM architecture? • Can a causal reasoning capability emerge in the existing LLM architectures? • It would be helpful to integrate meta-causal graph analysis. Maybe we can contribute this. • Can we build a system that combines knowledge graphs with LLMs (eg ConceptNet) • Utilise: Could we potentially use LLMs to suggest variables of interest to consider in learning causal models? • LLMs can be thought of a new type of associative database. We don't know the query language, because it emerged. But it cannot reason – so we should plug it into a tool to do different functionality (eg causal reasoning, symbolic reasoning, using knowledge graphs) • This is an amazing knowledge database; we should be able to build things on top of LLMs to get it to do something more useful. • Can we implement a way to add new causal knowledge from scientific studies, papers, etc, and uniquely treat it as causal knowledge that can be used for causal inference? • We could contribute by building plugins that provide causal reasoning capability to LLMs. • Chat GPT does not ask questions. This is because ChatGPT is not very good at counterfactuals. We need counterfactuals to imagine which interventions might be useful to explore. ChatGPT does not ask questions – it lacks the counterfactual / causality to know which questions would be helpful. Can we contribute counterfactual capabilities to LLMs?

4.9. Group 9

- **Name:** Henri Arno, Daigo Fuijwara, Fabrizio Russo, Francesco Montagna, Tobias Freidling, Malte Luttermann, Vittorio Del Tatto, Mizu Nishikawa-Toomey, Elisabeth Ailer

- **How do advances in LLMs disrupt causal research?**
  LLM open up possibilities for new research, for instance, the community can think about ways to evaluate the notion of causality in such models. We are not sure it disrupt causal research in a very fundamental way given that it demonstrated to not be able to capture causal knowledge. However, it does represent a tool that could be adapted to support parsing information in various ways (as discussed later). In situations where the risk of coming to a wrong causal conclusion is low, LLMs may potentially be applied more liberally, in particular if it is used to generate hypotheses that can be tested in follow-up studies. Moreover, the application of LLMs to causal research could part of the research focus from the development of causal models and theories to the development of evaluation metrics to assess the outcome of these models.

- **How can we as a community utilize and contribute to the advances in LLMs?**
  Maybe GPT have no or little causal model in itself at recent state. So we can contribute it to install causal knowledge by some causal structures or interventional data. One way it could be useful to causal researchers is to link LLMs to experimental data in some sort of a tuning step. This should make sure that the LLM prioritise knowledge from the tuning sources

so that practictioners can use it to parse large amount of knowledge. Moreover, LLMs can assist in the general research process as a tool for writing academic articles or code as well as recommending (and summarizing) scientific literature. In the latter application, there is a risk of one-sided and biased results, however. As a community, we should also require a transparency in the training procedure and training data employed by LLMs.

References

Jeroen Berrevoets, Krzysztof Kacprzyk, Zhaozhi Qian, and Mihaela van der Schaar. Causal deep learning. *arXiv preprint arXiv:2303.02186*, 2023.

Zhijing Jin, Amir Feder, and Kun Zhang. Causalnlp tutorial: An introduction to causality for natural language processing. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, pages 17–22, 2022.

Emre Kıcıman, Robert Ness, Amit Sharma, and Chenhao Tan. Causal reasoning and large language models: Opening a new frontier for causality. *arXiv preprint arXiv:2305.00050*, 2023.

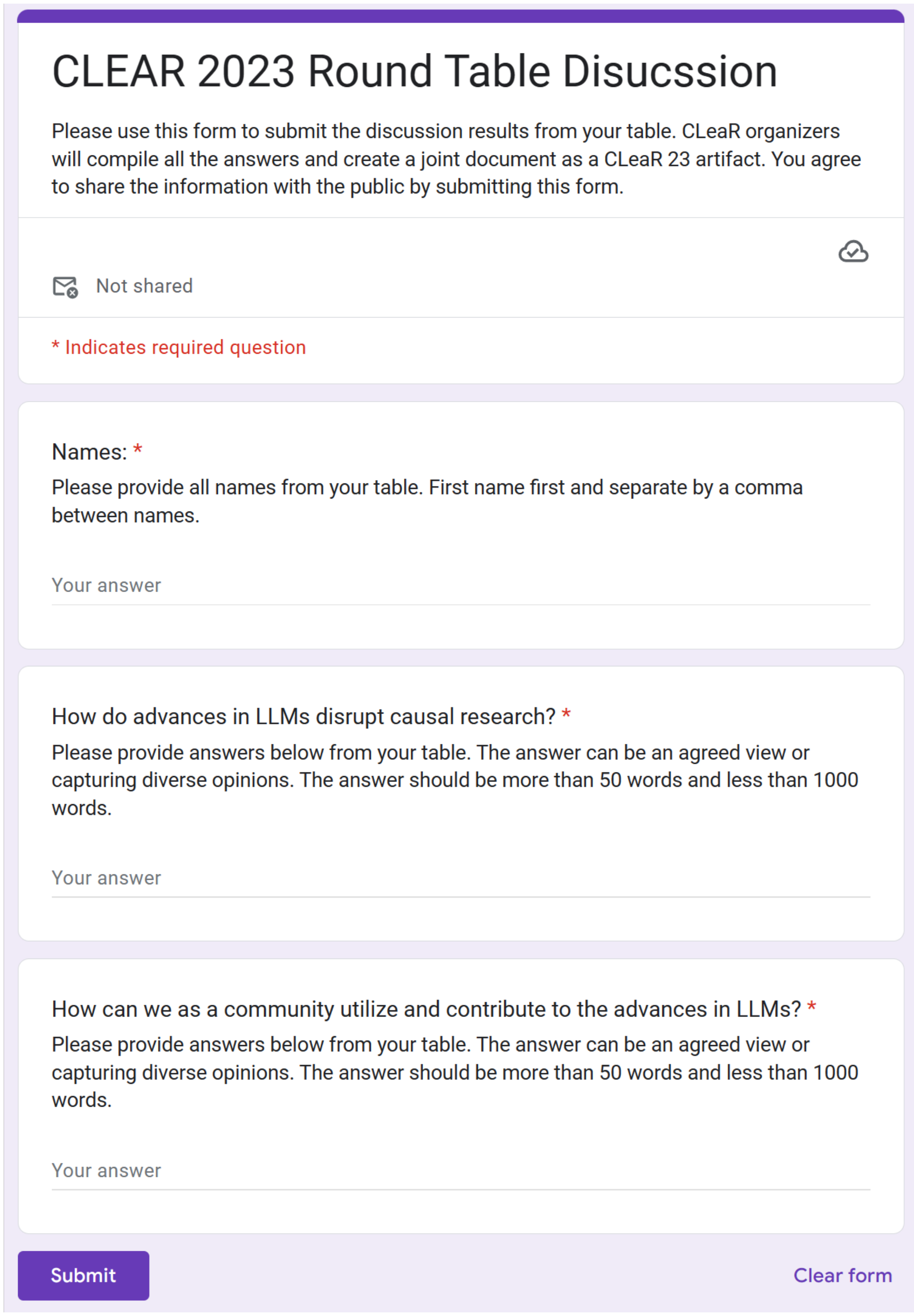
Stephanie Long, Tibor Schuster, Alexandre Piché, ServiceNow Research, et al. Can large language models build causal graphs? *arXiv preprint arXiv:2303.05279*, 2023.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Moritz Willig, Matej Zečević, Devendra Singh Dhami, and Kristian Kersting. Can foundation models talk causality? *arXiv preprint arXiv:2206.10591*, 2022.

Cheng Zhang, Stefan Bauer, Paul Bennett, Jiangfeng Gao, Wenbo Gong, Agrin Hilmkil, Joel Jennings, Chao Ma, Tom Minka, Nick Pawlowski, et al. Understanding causality with large language models: Feasibility and opportunities. *arXiv preprint arXiv:2304.05524*, 2023.

Guanglin Zhou, Lina Yao, Xiwei Xu, Chen Wang, Liming Zhu, and Kun Zhang. On the opportunity of causal deep generative models: A survey and future directions. *arXiv preprint arXiv:2301.12351*, 2023.



CLEAR 2023 Round Table Disucssion

Please use this form to submit the discussion results from your table. CLeaR organizers will compile all the answers and create a joint document as a CLeaR 23 artifact. You agree to share the information with the public by submitting this form.

Not shared

* Indicates required question

Names: *

Please provide all names from your table. First name first and separate by a comma between names.

Your answer

How do advances in LLMs disrupt causal research? *

Please provide answers below from your table. The answer can be an agreed view or capturing diverse opinions. The answer should be more than 50 words and less than 1000 words.

Your answer

How can we as a community utilize and contribute to the advances in LLMs? *

Please provide answers below from your table. The answer can be an agreed view or capturing diverse opinions. The answer should be more than 50 words and less than 1000 words.

Your answer

Submit

Clear form

Figure 1: The form that the attendees are asked to submit as a group at the end of the round table discussion