

prompting the past

frédéric clavert c2dh, uni.lu frederic.clavert@uni.lu

Short paper proposal for DH2024. Also available on Github, with a repository that includes a code notebook.



Figure 1: “a portrait of Mr Bean as Napoléon Bonaparte” – obviously a more accurate representation of Napoleon than the Ridley Scott movie. Source

introduction

For several years now, images generated by artificial intelligence systems such as midjourney, DALL·E or StableDiffusion, to give just a few examples, have become commonplace. Like text-generating systems such as GPT, LLAMA or BLOOM, they are based on interaction with users, which results in the writing of prompts. The aim of our current research is to examine the various means at our disposal for transforming not only generated images and texts but also prompts into primary sources for historians. This short paper will focus on prompts.

state of the art

In the fields of digital memory studies (Hoskins (2018)) and digital history, Wulf Kansteiner investigates what a specifically trained generative AI for historians

could be (Kansteiner (2022)) but without really looking at primary sources produced by these systems (the images and texts generated and the prompts used, in this case, by historians). The possibilities and risks of AI, particularly generative AI and especially in Holocaust studies, were analysed in an “open forum” of *Eastern European Holocaust Studies* (Makhortykh (2023)). In particular, the (low) adequacy of the responses of some of the generative AIs to Holocaust research was analysed in detail (Makhortykh, Vziatysheva, and Sydorova (2023)). Online software can be tested to understand what generative AI systems “know” about history (Hutchinson (n.d.)). Changes – positive or not – in the memorialisation processes have also been analysed (Makhortykh et al. (2023)), rather from the infrastructure angle. But as far as we know, there is little research on prompts as a source for history or for the study of collective memory. Those articles are also dealing with ethics, privacy and biases (including biased training dataset)¹. Within a couple of years, the digital history and memory studies literature on AI has hence and is still expanding. Nevertheless, and to our knowledge, few researchers have put prompts at the core of their research.

what could we learn from prompts?

Questioning the past is the core activity of historians, their basic epistemological operation in the sense that we ask questions to start the process of elaborating new knowledge about the past. The fact that easy to use tools (image or text generative systems) are based on prompts, which are often explicit or implicit questions, should get our attention. Those systems are incentives to question our world and the world that was, all the more that, as “stochastic parrots” (Bender et al. (2021)), they are a-epistemological: there is no notion of truth, lie or knowledge in the way those systems are working.

If we consider what those systems are based on (code, training dataset for instance), what is required to generate an answer (prompts) or their outputs (images, texts, etc.), then they are generating numerous primary sources – artefacts that tell us a lot about the societies of the (near) past. One of those primary sources are prompts. We consider here prompts relating to the past as open doors to users’ imagination about the past on one side, but also as a way to negotiate with the “machine” (when LLMs or image generative systems are not delivering what the user is expecting) on the other side.

harvesting data

The first methodological barrier to a research project on prompts related to the past is the making of a corpus. Where to find prompts? There are several ways we could assemble a database of prompts. There are several paths that could be followed to constitute a sound corpus. A first path would be to use dedicated

¹For an extended bibliography on AI and collective memory:
https://www.zotero.org/groups/4874770/ai_collective_memory.

search engines, such as Lexica (an example among others). Those search engines poses several problems: either they are basic keywords based search engine and collecting prompts would imply a database of past-related keywords or they are “semantic” (the case of lexica) and becomes hence black boxes. A second path would be to directly do what those search engines are doing: collecting data on discord.

Discord has indeed become a keypoint for communities to share their uses of naming several online database of prompts. Example of lexica <https://lexica.art/prompt/df06475d-3862-4fac-936e-86657566f524>

Lexica.art created this search engine by harvesting generated images published on the Discord Stable diffusion server. Ther's a ‘semantic’-ish system associated to it (they speak of ‘image relevancy’), as this image and its prompt were found through the keyword ‘Napoléon’.

My problem here is that it's not that easy to get a balanced corpus of prompts. I'm working on that, and can for now name only a few leads.

Of course, all ethics and privacy issues are still to be managed.

- contacting AI-systems developers and firms
 - craiyon: did not keep time-date informations up to now;
 - no way openAI is accepting – and not sure I want to work with them;
 - Stable Diffusion
 - For midjourney, I need to investigate more, but can only be done with a paid subscription and not sure gives the right to
- prompts search engines (lexica.art, PromptHero, craiyon, etc.)
 - limitation : date time informations
- Discord servers
 - It's how search engines built their corpus => there could be enough information.
- reddit
 - I don't know if it's a good tool / but easy to scrap data there
- Twitter
 - hashtags where people just tweet prompts + image.
 - Elon Musk's problem: might not be possible nor pertinent.

That's for the source. The second point is, well, what do we mean by ‘past’?

So I do not have this corpus yet, but started experiments. And reading those « experiments ».

analyzing data

distant reading prompts

Here's a sort of topic modelling (not topic modelling but it assumes same function) of a corpus of 1908 prompts.

Iramuteq is basically clustering prompts based on collocations of words. The keywords you see are words that are the most representative of those clusters and allow for an interpretation of the clusters. There are other, thanks to this software, ways to help the user interpret the clustering (getting the most representative prompts of each clusters for instance).

The prompts here are containing 'european union' - I wanted other keywords for this presentation, but the scripts is not working anymore so I could not collect data specifically for this presentation.

- it's not only about the past, but quite a lot about
- there are clusters about styles – note that a lot are 'historical' (soviet propaganda)
- there are elements of recent politics (nigel farage / marine)
- europe as something linked to a period of time: Middel Age (heraldic) / 17th Century (Rembrandt)
- notions quite linked to european history: 'empire'
- the question of the link with news is quite important
- war is very present => because Europe, but it seems that there is a global association of history and war

My hypothesis was that prompts are a door to people's imagination about the past. It's now that I will comment and criticize this hypothesis.

Discussing this hypothesis is also switching from DH methods to digital sociology methods: mixing data with qualitative analysis, interviews, etc.

If we stick to distant reading of prompts, we do not have the whole picture, and are missing what's happening between the users and the interface.

oral history of prompting?

- the aim here of the user was to get a nazi salute – and it's not working
 - think about the 'Napoléon rides a shark' image that you saw before: it's the same, it doesn't work
- It's not working for different reasons:
 - well, this system does not know what a nazi salute is – images here are just pixels which succession is statistically pertinent
 - there might be limitations implemented in the text-to-image system (not likely with stable diffusion)

What does a user who did not get the results it wanted?

- as it fails, users might have strategies to get what they want
 - strategy based on changing prompts,
 - means that prompts, not only images (or generated text), are the result of the settings of the machine, or at least partly, and of a man-machine interaction.

should we create our own prompt-generating-systems?

- we need prompts with proper metadata and that relate to the past
 - one of the way to get that is to set up our own image / text generation system. With open source systems, it could be possible.
 - we could train this system on ‘historical data’.
 - many problems to solve – including, of course ethics.
- we need to do scalable reading
 - but also to mix methods, not only DH but also (digital) sociology: interviews and surveys.

Here again, might cause ethical problems.

Conclusion

References

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT ’21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Hoskins, Andrew, ed. 2018. *Digital Memory Studies: Media Pasts in Transition*. New York: Routledge.
- Hutchinson, Daniel. n.d. “What Do AIs ‘Know’ About History? A Digital History Experiment.” Accessed October 5, 2023.
- Kansteiner, Wulf. 2022. “Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?” *History and Theory* 61 (4): 119–33. <https://doi.org/10.1111/hith.12282>.
- Makhortykh, Mykola. 2023. “Open Forum: Possibilities and Risks of Artificial Intelligence for Holocaust Memory.” *Eastern European Holocaust Studies* 0 (0). <https://doi.org/10.1515/eehs-2023-0053>.
- Makhortykh, Mykola, Victoria Vziatysheva, and Maryna Sydorova. 2023. “Generative AI and Contestation and Instrumentalization of Memory About the Holocaust in Ukraine.” *Eastern European Holocaust Studies*, November. <https://doi.org/10.1515/eehs-2023-0054>.
- Makhortykh, Mykola, Eve M. Zucker, David J. Simon, Daniel Bultmann, and Roberto Ulloa. 2023. “Shall Androids Dream of Genocides? How Generative

AI Can Change the Future of Memorialization of Mass Atrocities.” *Discover Artificial Intelligence* 3 (1): 28. <https://doi.org/10.1007/s44163-023-00072-6>.