Imagined Recollections: Exploring Memories of Forced Labor through AI-Generated Imagery

Hasan Selim Yagci
selim.yagci@tum.de

Abstract

With the recent developments in generative forms of Artificial Intelligence (AI), it is vital to rethink our interaction with history, atrocities and memorialization. The digital media has already transformed the accessibility of memories through web archives of images, interviews and evidences. In this project, using the interview data available on digital Forced Labor archives, we investigated the capabilities of diffusion models to generate memory guided visual representations. Despite the limits of the model, the results showed promising results which potentially can help past atrocities be more accessible to young people and consequently increase awareness towards current labor exploitation and other rights and justice issues.

1 Introduction

Remembering past atrocities constitutes a significant part of public memory and aims at truthful representation, respecting victims' sufferings, seeking justice, healing the wounds of atrocity and preventing the recurrence (Gill, 2023). Throughout recent history, we have witnessed contrasting outcomes from the attempts of shaping the public memory in the context of memory politics. In the case of Japan and South Korea, there have been conflicting narratives about forced labor victims of South Korea, revealing Japan's denialistic propaganda (Sîntionean, 2020). Whereas in Germany, despite the public spaces filled with sites of memory, it has not stopped the rise of far-right, anti-semitism and hate (Evans and Braslavsky, 2019). Both examples might indicate that creating a collective memory is complex and as the world continues to struggle with contemporary atrocities, the importance of rethinking the form and content of memorialization practices for past atrocities is becoming more evident.

Reflecting on current and past practices, we can see the emergence of digital media has transformed the way memories are formed and shared (Makhortykh1 et al., 2023). Public perception of mass atrocities shaped largely by visual media and especially digital visual history concept has been shifting the top down approaches towards more creative ones by enabling interactivity and accessibility (Ebbrecht-Hartmann et al., 2023). One example of transformed practices is the *#everynamecounts* memorial in Berlin, initiated by Arsolen Archives. The project is about digitizing the names of victims and survivors of National Socialism and enabling thousands of volunteers to actively participate to create a virtual collective memory (eve, 2024).

Besides the general digital transformation, the recent developments in AI and generative models has already started influencing the people working on Holocaust heritage and memory. The AI exhibition installed in Ashkelon Palace of Culture is one of the recent examples (Riba, 2023). Although the effort of people behind the exhibition trying to make Holocaust accessible to young audiences has been criticized about sensitive depiction, it has raised the question about the potential of AI to revolutionize the memorialization practices while preserving its ethical and responsible usage (Makhortykh et al., 2023).

In the light of this question, the main objective of our project is investigating the capabilities of pretrained generative models over downstream tasks of generating images representing memories of forced labor victims. The overview of the experiment pipeline can be seen in Figure 1. The rest of the report is outlined as follows: In Section 2, the methodological processes are explained in detail giving the information about the data and its source, hyperparameter specification of the model and the further improvements. In Section 3, the findings are presented with the subsequent takeaways. The possible future steps that could be taken to improve the project is mentioned in Section 4. Finally, the report is concluded in Section 5.

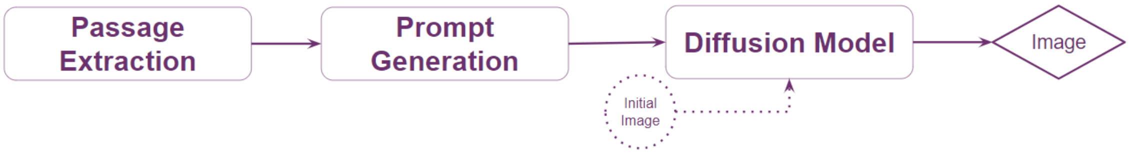


Figure 1: Overview of experiment pipeline

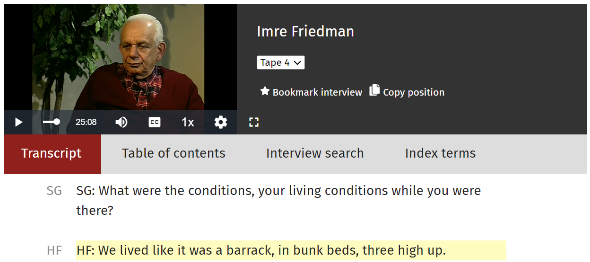


Figure 2: Example of passage extraction, using the keyword *barrack*

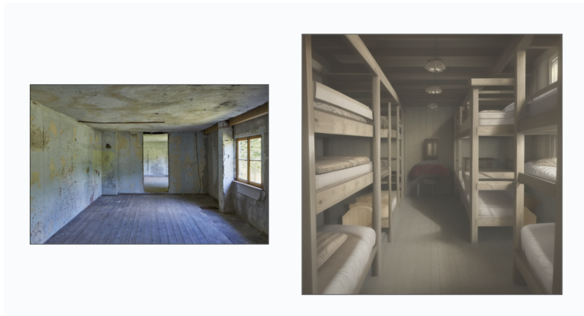


Figure 3: Image-to-image model output on the right, guided by new photograph from a barrack and prompt *"we lived like it was a barrack, in bunk beds, three high up"*

2 Methodology

The methodology is divided into two main processes for the experimental setup: the information about text retrieval process (section 2.1) and the detailed description of image generation process (section 2.2).

2.1 Retrieving Text from the Interview Archive

In order to capture the narratives that are going to be represented in visual domain, testimonies of forced labor victims are retrieved from the "Forced labor 1939-1945" digital archive (for, 2009).

The archive was created in 2007 as a part of the "Documentation of Life Story Interviews with Former Slave and Forced Laborers" project. It includes audio and video interviews of over 500 survivors and their transcripts. Platform allows users to search for keywords and locations.

Using search results of keywords that may give visual information (i.e. *wall, barrack, airplane, bombs*) or words related to location/clothing (i.e. *behind, factory, uniform, trousers*), passages are extracted (see Figure 2).

2.2 Generating Images guided with Text Prompts

The experiments are split into two main approaches, text-to-image and image-to-image generation. For both approaches, Stable Diffusion model is used for inference.

2.2.1 Diffusion Model

Stable Diffusion is a type of latent diffusion model. The model is trained to construct an image starting from an initial noise (*random latent image representations*) by removing noise every step. The input prompt gets mapped by text encoder and model conditions its output on resulting text embeddings. In general, results get better the more steps the model iterates (Rombach et al., 2022). The base model has around 1B parameters which makes fine tuning computationally expensive. Thus for the downstream experiments, the model is used only for inference with pretrained weights. On the other hand, some of the hyperparameters of the model are adjusted to optimize, including the following: steps count, learning rate, batch size, guidance scale, scheduler and sampling method. Size of the output image set to 512x512 pixels and seed is used as control parameter, to keep the style similar among different prompts.

In the image-to-image case, instead of random noise, the diffusion process starts with the initial image (and added noise) passed to the model. Keeping the guidance scale and steps count lower, the model preserves the features of the initial image more (see Figure 3).

2.2.2 Prompt Engineering

Since the passages themselves as a plain input to the diffusion model will not give the best results,

Figure 4: Example output using tokenized prompt with display features

modifications are applied. The text prompt can be separated into two main components, the context and the features of display (Witteveen and Andrews, 2022).

- In the context component, two different structures are followed: the contextual information as a plain sentence (i.e. *we lived like it was a barrack, in bunk beds, three high up*) or a tokenized version (i.e. *photo of forced laborers, blindfolded, by the concrete wall, Germany, 1940*, for the example output see Figure 4).

- Whereas the display features component contains attributes of style, photographic element, lighting and etc. (i.e. *realistic, soft lighting, drone view, 4k*)

2.3 Further Improvements

In order to enhance the quality of the generated images several techniques are utilized.

- **Deterministic Generation** The main idea of this method is to generate several outputs every inference step and to select the best one to improve with a detailed prompt in the next step.

- **Negative Prompts** Another guiding mechanism of Stable Diffusion model is defining prompts that are going to be excluded while generating the image. This method is useful to prevent unnatural things from appearing in the generated images.

- **Prompt Weighing** In order to emphasize significant parts of the text prompt, we can weigh the prompts using the Compel library (HuggingFace, 2024).

3 Discussion

Overall, the resulting images showed promising results with regard to representing memories of forced labor victims visually (see Appendix A). Even though sensitivity and fact based generation principles are injected through the guidance of prompts and negative prompts, the limit of the model in use (see Section 2.2.1) affects the generation quality. Other takeaways from the experiments can be listed as follows:

- Available diffusion models are largely pretrained on data irrelevant to specific history and specific place of the events, which is causing distortion in the factual representation.

- Lack of *empathy* in models can still lead to feeling-ignoring content despite the controlled prompts.

- Finetuning diffusion models is computationally expensive and recent developments are not enough to make the training process more sustainable and source-efficient.

- Experiments show even with the limited capability, generative models can be instrumentalized to promote false information concerning atrocities and denialistic propaganda by distorting representations. Although wrong mediatization of memories is still a problem regardless of Generative AI, it could become more severe and end up demolishing the photographic truth.

- Transforming memory practices with AI is still a promising research area. Especially in the scope of forced labor, models' visual and interactive power can help people reflect on current justice issues concerning labor exploitation.

- Research shows western engines prioritized images showing deportation and liberation of the camps, however, in Russian the results are more graphic content (Makhortykh et al., 2021). Similarly, since the diffusion models can be trained on different language corpora, their capability to represent differences between narratives of different cultures and memory politics could be a very important measurement for the social studies research (i.e. Japan and Germany cases in Section 1).

4 Future Work

Training Diffusion Models with Reinforcement Learning One of the further work in the objective of this project is improving the aesthetic quality and prompt-image alignment using Reinforcement Learning approaches (Black et al., 2023). In this approach, the diffusion models are finetuned on three downstream tasks using the feedback of a large vision-language model (VLM): image compressibility, human-perceived aesthetic quality, and prompt-image alignment. Then, the similarity between the feedback of VLM and initial prompt of the generated image is measured using a transformer based model (i.e. BERTScore). The resulting similarity score gets utilized as the reward function for the reinforcement learning objective to improve performance of the diffusion model in the loop (i.e. applying *Denoising Diffusion Policy Optimization*).

Automating Data Retrieval and Prompt Construction During the experiments, collecting the data and constructing the prompts are done manually. In order to create an automated pipeline (i.e. to make it available for user interaction), one can integrate intermediate models. Using a text summarization model and a Part-of-Speech Tagger (Chiche and Yitagesu, 2022), relevant passages from the large transcript documents can be extracted. Then, training an autoregressive language model (like GPT4) prompts can be constructed with user specified style and attributes as in Section 2.2.2.

5 Conclusion

The emergence of generative models in the field of memory politics, visualization of history and memorialization practices is inevitable considering the recent rise in the technology. In the scope of this project, we focused on forced labor and memorialization practices and tried to investigate diffusion models' power to represent the past. More experiments should be conducted not only for assessing the capability of models generating realistic images but also for detection of generated content. Despite the shortcomings of models to create authentic historical images, the ethical and responsible development should be promoted against the possible instrumental uses of the past.

References

2009. Forced Labor 1939-1945: Memory and History. https://www.zwangsarbeit-archiv.de/. Accessed: 2024-03-01.

2024. everynamecounts: Digitales Denkmal für NS-Opfer! https://everynamecounts.arolsen-archives.org/. Accessed: 2024-03-01.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. 2023. Training diffusion models with reinforcement learning.

Alebachew Chiche and Betselot Yitagesu. 2022. Part of speech tagging: A systematic review of deep learning and machine learning approaches. *Journal of Big Data*, 9(1).

Tobias Ebbrecht-Hartmann, Noga Stiassny, and Lital Henig. 2023. Digital visual history: Historiographic curation using digital technologies. *Rethinking History*, 27(2):159–186.

Jennifer Evans and Paula-Irene Braslavsky. 2019. Can remembering past atrocities prevent future ones? Accessed: 2024-03-01.

Scherto Gill. 2023. Ethics of remembering. *Revista Latinoamericana de Estudios sobre Cuerpos, Emociones y Sociedad*.

HuggingFace. 2024. HuggingFace Diffusers. https://huggingface.co/docs/diffusers/index. Accessed: 2024-03-01.

Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa. 2021. Hey, google, is it what the holocaust looked like? *First Monday*.

Mykola Makhortykh, Victoria Vziatysheva, and Maryna Sydorova. 2023. Generative ai and contestation and instrumentalization of memory about the holocaust in ukraine. *Eastern European Holocaust Studies*, 1(2):349–355.

Mykola Makhortykh1, 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*.

Naama Riba. 2023. These Holocaust AI Generated Images Distort History. https://www.haaretz.com/israel-news/2023-02-01/ty-article-magazine/.premium/new-israeli-exhibit-showcases-ai-generated-holocaust-images/00000186-0c8d-dc4a-a3be-efbf01e20000. Accessed: 2024-03-01.

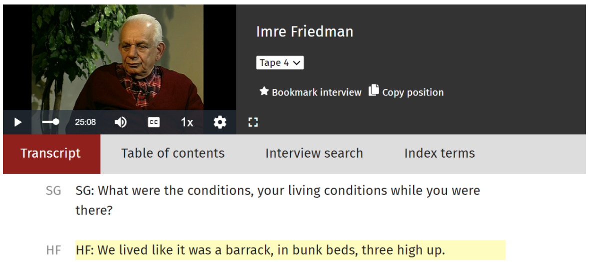
Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models.

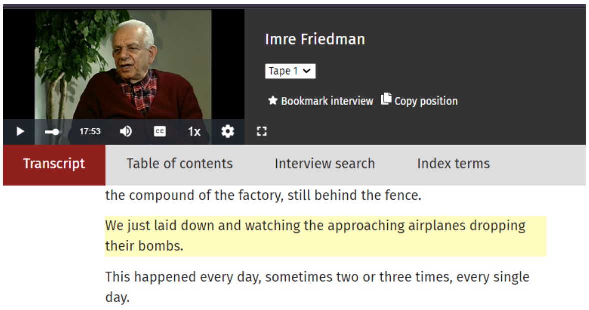
Codruta Sîntionean. 2020. The role of historical memory in japan - south korea relations. *European Journal of Interdisciplinary Studies*.

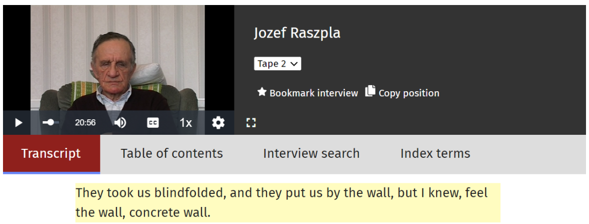
Sam Witteveen and Martin Andrews. 2022. Investigating prompt engineering in diffusion models.

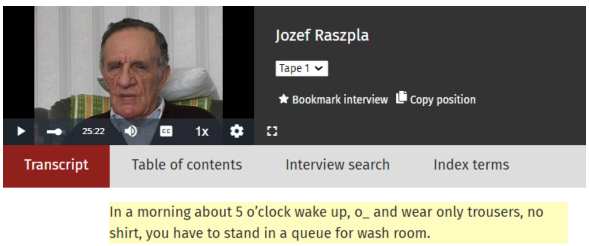
A Generated Images

A.1 Example prompts from forced labor archive









A.2 Example generated images guided with extracted prompts