

Yuqing Zhai

10/23/2024

In 2022, LAION (Large-scale Artificial Intelligence) have released and produced LAION-5B, a training dataset containing links to images scraped from Internet. In 2023, Chief Technologist David Thiel of the Stanford Internet Observatory discovered that LAION-5B contains suspected Children Sexual Abuse Material (CSAM), which make the possession of the dataset illegal and the models trained on it problematic. [1, 2, 3] This essay will first analyze the incident with concerns about **privacy, crime, and intellectual property**, it will then use **act deontological ethic** to argue that **all entities** which host, collect or scrape, distribute, and train models upon these images should be held **morally accountable**.

First, regarding the issue of crime. As stated in 404 Media, US Laws are pretty strict, any possession of CSAM, even if accidentally, is a federal crime. The LAION-5B training dataset is not image themselves but rather links to these images, and “It’s not clear where URL that link to child exploitation images would land under current laws”. [1] **Whether possession of LAION-5B is a crime is legally obscure.** Secondly, regarding the issue of privacy, the original Stanford paper said “implies the possession of thousands of illegal images—not including all of the intimate imagery published and gathered non-consensually”. [4] These intimate imagery could include personal facial information, or sensitive Personal Identifying Information (PII) could well be text embedded in these images. Although it is not a violation of privacy, as these images are publicly accessible, **creating such aggregated dataset might facilitate deanonymization of people.** Thirdly, regarding intellectual property (IP), As stated by Vencu in the LAION discord group, “Since we are not distributing or deriving other images from originals, I do not think the image licensing apply”, [1] **Laws regarding IP do not apply to LAION-5B**, as it only distribute the links to original images. **The use of these images for commercial AI model** is a different problem whether it is a fair use or a violation IP **is still controversial**, as also mentioned by 404 media “Earlier this year, a group of artists filed a class-action lawsuit against Stability AI, DeviantArt, and Midjourney for their use of image generator Stable Diffusion, which uses LAION’s datasets.”. [1]

Act deontological ethic argues that we should analyze individual situation and act according to the most important duty we need to fulfill. There are few entities involved in this incident: (1) the original website who host CSAM, (2) the LAION team (3) the Hugging Face (HF) platform where LAION-5B get released (4) the companies like Google and Stability AI who trained these model based on LAION-5B. **For the original website**, one of its duty is to prevent the distribution of CSAM to protect the victim in actual incidents. They have failed to do so as Thiel mentioned that 1008 CSAM are verified externally, which means that these material have not been taken down by the time the paper was written. [1, 2, 3]. **For the**

LAION team, as noted by common **data science ethics**, they should make sure that data is safe. It is understandable that they cannot remove 100% CSAM, as it seems practically impossible considering the size of the dataset (Ref). Yet as noted by Thiel, “LAION did initially use CLIP to try and filter some things out, but it does not appear that they did that in consultation with any child safety experts originally. It was good that they tried. But the mechanisms they used were just not super impressive”, [1, 4] the LAION team fails to fulfill their responsibility to make sure that data is safe to their best effort. **For Hugging Face Platform**, one of their most important responsibilities is to prevent the distribution of illegal content like CSAM in order to facilitate a safe AI research platform. However, as mentioned by 404 Media, they took over 7 months to address a report of CSAM. [1] This incompetency makes them fail to prevent the distribution of illegal content in a timely manner. **For AI companies**, one of their most important duties is to make sure their model is safe to use, this include not generate CSAM image, especially these could relate to real specific victims. As noted by paper from Stanford, “it [CSAM] likely does still exert influence. The presence of repeated identical instances of CSAM is also problematic, particularly due to its reinforcement of images of specific victims”. [1, 4] The AI companies failed to adhere **data science ethics** by ensuring the validity of their data and foreseeing the severe implication of such “polluted data”, and they failed to fulfill their most important duties. **In conclusion**, we see that all the entity involving the CSAM in the LAION-5B dataset **have failed to fulfill their most important moral responsibility, and should thus be held accountable for it.**

Reference

[1]

Cole, S. (2023, December 20). Largest dataset powering AI images removed after discovery of child sexual abuse material. *404 Media*. <https://www.404media.co/laion-datasets-removed-stanford-csam-child-abuse/>

[2]

David, E. (2023, December 20). AI image training dataset found to include child sexual abuse imagery. *The verge*. <https://www.theverge.com/2023/12/20/24009418/generative-ai-image-laion-csam-google-stability-stanford>

[3]

Vigliarolo, B. (2023, December 20). Study uncovers presence of CSAM in popular AI training dataset. *The register*. https://www.theregister.com/2023/12/20/csam_laion_dataset/

[4]

Thiel, D. (2023). *Identifying and eliminating CSAM in generative ML training data and models*. Stanford Digital Repository. <https://doi.org/10.25740/kh752sm9123>