デジタルヒューマニティ部門

“Identifying and considering the social implications of private information in image datasets”

Amelia KATIRAI（Research Center on Ethical, Legal, and Social Issues），
Noa GARCIA DOCAMPO（Institute for Datability Science），Kazuki IDE（Center for Infectious Disease Education and Research），Yankun WU（Intelligence and Sensing Lab），Yuta NAKASHIMA (Institute for Datability Science), Atsuo KISHIMOTO（Research Center on Ethical, Legal, and Social Issues）

1. Research background

The development of image generation models and their free availability to members of the public has made use of these technologies widespread, before their social and ethical issues have been adequately considered. Data for image generation models have been shown to contain medical images, in addition to other sensitive imagery, posing a privacy risk (1). Much of this data is scraped from online sources, including social media and other platforms where individuals may share their private images, without awareness that these images may be scraped and used for other purposes. Furthermore, when image generation models are trained, they risk outputting this private information to users of the models (2). However, there has been insufficient attention to image generation models, despite their particular sensitivity.

2. Research aims

The overarching aim of this project was to investigate privacy issues in image generation models, with three sub-aims:

1) To verify whether image-based private information exists in the datasets used for major image generators;
2) To identify the social and ethical implications of these issues;
3) To investigate the perspectives of a survey sample of members of the public on these issues.

3. Research methods

To achieve the aims of this investigation, a case study approach was used, and medical images such as pregnancy ultrasounds were taken as the primary focus. Ultrasounds were selected for two reasons. First, as medical images, they typically contain private and sensitive information about the pregnant person and the health of the fetus. Furthermore, this is often coupled with non-medical identifying information about the pregnant person, such their name or the medical institution where the ultrasound is created. Second, due to their dual medical and social significance, ultrasounds are frequently shared online, and often publicly, meaning that there is potential for them to be scraped into datasets for image generators.

To meet each of the sub-aims of the project, the project was conducted in 3 phases, with each phase corresponding to a sub-aim. Each phase will be described below.

Phase 1

The LAION-400M dataset was analyzed to identify pregnancy ultrasound images, their frequency, quality and context, and the

presence of potentially privacy-sensitive data within the images.

Phase 2

Two reviews of the literature were conducted, one investigating social scientific approaches to privacy, and a second review investigating the social and medical significance of ultrasounds.

Phase 3

An online survey with a sample size of 1,000 participants was conducted. The survey content was decided by the research team, and the actual operation of the survey and the subsequent analysis were outsourced to established Japanese organizations specializing in online surveys and data analysis. The aim of the survey was to understand four points: 1) the extent to which people utilize social media to share images, including pregnancy ultrasounds; 2) the extent to which people undertake particular privacy protective measures; 3) the extent to which people utilize image generation models, and understand their mechanisms; 4) the extent to which people accept the use of images for training image generation models.

4. 研究成果 [1]

The findings of the research are presented below, organized by aim.

[1] Research results are presented here in an abridged form in consideration of future dissemination plans.

Phase 1

The investigation of the LAION-400M dataset led to the identification of real-world pregnancy ultrasounds within the dataset. Furthermore, from these ultrasounds, four types of personal information were identified as included in these images: name, location, date/time, and phone numbers. The identified ultrasounds were at times a smaller part of larger images such as baby shower invitations or images depicting expecting parents. Given that pregnancy ultrasounds are a type of medical data which can be highly sensitive as they may reveal information about a fetus' or person's health, the presence of these images coupled with identifying information in the dataset, and without adequate consent, is highly problematic. The findings of this study led to recommendations for datasets, including recommendations for effective data privacy and consent measures.

Phase 2

The review of the literature showed that pregnancy ultrasounds carry both medical and social significance, as they may be perceived to be a part of antenatal care, but also serve social functions such as when images are saved for posterity (3). This increasingly occurs online, as images are shared through online platforms, representing an "online birth" prior to the "physical birth" (4). Although such images are primarily shared with family or friends, research indicates that they are also shared publicly, including through platforms such as Instagram (5,6). This means they may be scraped into datasets for image generation.

This points to the sensitivity of ultrasounds from a privacy perspective. However, the review of the literature on the concept of privacy revealed that privacy is a frequently contested concept, even among experts. As a result, many privacy scholars argue for an anti-reductionist or pluralistic understanding of privacy, acknowledging that privacy is inextricably linked with contextual, social, and personal factors (7). For this study, Helen Nissenbaum's theory of contextual integrity was applied, which emphasizes the importance of norms and expectations for how information will be used within particular contexts, and how extraction from these contexts and the breach of these norms and expectations can be a breach of privacy (8,9).

Phase 3

The final phase of the project, the online survey, showed that participants had posted potentially privacy-sensitive images—including images of people posted without consent, or medical images—to social media. Moreover, although small in number, some participants had indeed posted pregnancy ultrasounds to social media, in some of these cases making them publicly visible. The majority of participants indicated that publicly available images on social media should only be used for training only with conditions such as without violation of copyright or with consent, or should not be used at all.

Thus, the three phases of this project confirm the presence of unaddressed issues related to privacy in image generation datasets. This research highlights the need for both further interdisciplinary research in this area on the one hand, and urgent improvements in practice around image generation and the associated datasets on the other. Future dissemination of the research findings is planned.

References

1. Katirai A, Garcia N, Ide K, Nakashima Y, Kishimoto A. Situating the social issues of image generation models in the model life cycle: a sociotechnical approach. AI Ethics. 2024;1–18.
2. Carlini N, Hayes J, Nasr M, Jagielski M, Sehwag V, Tramer F, et al. Extracting training data from diffusion models. In: 32nd USENIX Security Symposium (USENIX Security 23). 2023. p. 5253–70.
3. Øyen L, Aune I. Viewing the unborn child–pregnant women's expectations, attitudes and experiences regarding fetal ultrasound examination. Sex Reprod Healthc. 2016;7:8–13.
4. Johnson SA. "Maternal devices", social media and the self-management of pregnancy, mothering and child health. Societies. 2014;4(2):330–50.
5. Harpel T. Pregnant women sharing pregnancy-related information on Facebook: web-based survey study. J Med Internet Res. 2018;20(3):e115.
6. Leaver T, Highfield T. Visualising the ends of identity: pre-birth and post-death on Instagram. Inf Commun Soc. 2018;21(1):30–45.
7. Trepte S, Masur PK. Definitions of Privacy. In: The Routledge Handbook of Privacy and Social Media. Routledge; 2023. p. 3–15.
8. Malkin N. Contextual integrity, explained: A more usable privacy definition. IEEE Secur Priv. 2022;21(1):58–65.

9. Nissenbaum H. Privacy as contextual integrity. Wash Rev. 2004;79:119.

Research output

Conference Paper

① Rawisara Lohanimit, Yankun Wu, Amelia Katirai, Yuta Nakashima, Noa Garcia, 2025. “Privacy in Image Datasets: A Case Study on Pregnancy Ultrasounds.” Eighth AAAI/ACM Conference on AI, Ethics, and Society, 2025 [Conference Paper; Submitted]

Presentation

② Rawisara Lohanimit, Yankun Wu, Yuta Nakashima, Noa Garcia, 2024. “Personal Information in Image Datasets.” Spain-Japan Itinerant Symposium in Neurosciences, Neurorights, and Artificial Intelligence. [Poster Presentation; Accepted, Completed]

Workshop Proposal

③ Noa Garcia, Amelia Katirai, Kento Masui, Mayu Otani, Yankun Wu, 2025. 2nd Workshop on Critical Evaluation of Generative Models and their Impact on Society. International Conference on Computer Vision 2025 [Workshop Proposal; Accepted]