

Stop the Nonconsensual Use of Nude Images in Research

Princessa Cintaqia,¹ Arshia Arya,² Elissa M. Redmiles,³ Deepak Kumar,²
Allison McDonald*,¹ Lucy Qin*³

¹Boston University,

²University of California San Diego,

³Georgetown University

{cintaqia, amcdon}@bu.edu, {aarshia, kumarde}@ucsd.edu, {elissa.redmiles, lucy.qin}@georgetown.edu

In order to train, test, and evaluate nudity detection models, machine learning researchers typically rely on nude images scraped from the Internet. We find that some researchers collect and distribute this content without proper guidelines and consent, exacerbating harm against the subjects depicted. **The collection and distribution of nonconsensually collected nude images by researchers perpetuates image-based sexual abuse (IBSA).** IBSA is a broad category of harm that relates to the nonconsensual creation (e.g., “upskirting,” “deepfakes”) or distribution of intimate content, as well as threats to cause these harms (Henry et al. 2022). IBSA is illegal in many countries and can lead to consequences similar to other forms of sexual violence (McGlynn et al. 2021; Bates 2017). **We conducted a systematic review of 153 papers** using relevant keywords and excluded any that did not meet our inclusion criteria: (1) be published at a CS venue, and (2) use a dataset containing real (non-generated) nude images. Our team then manually annotated the final set of 153 papers using questions to capture key metrics (e.g., number of nude images collected) and questions to explore a paper’s framing and research practices. We find that ethical norms around the usage of nude images are sparse, leading to a litany of problematic practices, like distributing and publishing nude images with uncensored faces, and intentionally collecting and sharing abusive content.

Researchers engage in IBSA: Images are distributed without consent via publication, annotation, and open science. One of the primary challenges that victim-survivors of IBSA face is stopping further dissemination of nonconsensually created and/or shared content (Huber and Ward 2024). Distribution of nude images as examples contributes to this challenge. More than a half of papers, including those published in top-tier venues, embedded example nude images (sometimes with faces uncensored).

Images are further disseminated through annotation. Although a quarter of papers described the need for manual labeling (including annotating “centers of private parts” and specific sexual acts such as “Vaginal Penetration, Anal Penetration, BDSM, Bestiality”), they were opaque about who performed annotations and provided no detail regarding protections against dissemination by third parties.

Finally, 5 papers made their created datasets of nude images publicly available. Although there are norms within the ML community towards sharing datasets, public accessibility is not appropriate in this context.

Researchers amplify harm by collecting and using abusive content. Almost two-thirds of papers created new datasets. A fifth of those did not mention a data source. Others collected data from Reddit, Google, “public social networks”, and 4chan, all platforms in which nonconsensual intimate content is often distributed (Huber and Ward 2024). In some instances, researchers deliberately included abusive content by collecting images from subreddits that have since been banned (“r/StruggleFucking” and “r/rape_roleplay”). In another example, researchers used images described as being extracted from “hidden or self cameras.” A few papers used vague data collection criteria that could include images of minors (e.g., collecting images of “girls, boys, teenager” in conjunction with keywords related to sexual acts).

Scraping nude images violates fundamental rights. We did not find any papers that discussed the (lack of) consent from image subjects. Regardless of whether the content being scraped was originally created and/or distributed consensually, it is being collected nonconsensually by researchers for use cases that are unbeknownst to the data subjects.

We need ethical standards for the handling of nude data. We observed virtually no difference between how researchers handled nude images compared to other datasets. Despite dealing with highly sensitive data, most papers make no effort to describe protections against accidental data leakage or deletion plans for the nude images they used. As mentioned, many papers themselves further disseminate nonconsensual images, and offer no—or very weak—justification for the harms of the research process.

Researchers *should* work to build systems that detect, take down, and prevent the spread of abusive content, while simultaneously recognizing that not all nude content is abusive. However, if we accept that there exist tasks for which highly sensitive and potentially harmful data are necessary, it is imperative that we also forge pathways to the ethical creation and handling of such data. We call for a new model of data governance in this domain that centers ongoing, informed consent. Specifically, we propose a participatorily-governed data trust of nude and sexual images that are consensually collected and used for research purposes.

*Equal advising contribution

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Ethical Statement

The authors of this project were all made aware of the sensitive nature of this work prior to starting the project. Support resources, such as free sessions with a licensed therapist, were available to authors who might encounter abuse content while annotating papers. Furthermore, due to concerns about the inclusion of sample images in papers and some of the research practices we describe in our findings, we have taken measures to reduce the discoverability of the papers within our dataset. We intentionally do not share links to public datasets we have found in the process of analyzing our dataset (though we make exceptions for large, well-known datasets, such as NudeNet). We are in the process of responsibly disclosing our findings to relevant publishers (e.g., IEEE, ACM) and hosting platforms for available datasets. In particular, our analysis demonstrated that there is a large quantity of published papers that contain uncensored nude images (including faces) and known abuse images. These images should be removed from research repositories.

The full version of this paper is available at:

<https://bit.ly/nonconsensual-data-research-full>

References

- Bates, S. 2017. Revenge Porn and Mental Health: A Qualitative Analysis of the Mental Health Effects of Revenge Porn on Female Survivors. *Feminist Criminology*, 12(1): 22–42.
- Henry, N.; McGlynn, C.; Flynn, A.; Johnson, K.; Powell, A.; and Scott, A. J. 2022. *Image-based sexual abuse: A study on the causes and consequences of non-consensual nude or sexual imagery*. Abingdon, Oxon: Routledge. ISBN 978-0-367-52440-1.
- Huber, A. R.; and Ward, Z. 2024. Non-consensual intimate image distribution: Nature, removal, and implications for the Online Safety Act. *European Journal of Criminology*, 14773708241255821.
- McGlynn, C.; Johnson, K.; Rackley, E.; Henry, N.; Gavey, N.; Flynn, A.; and Powell, A. 2021. ‘It’s Torture for the Soul’: The Harms of Image-Based Sexual Abuse. *Social & Legal Studies*, 30(4): 541–562. eprint: <https://doi.org/10.1177/0964663920947791>.