The Subject(s) of Representational Harm

NASANBAYAR ULZII-ORSHIKH, University of Michigan, USA
MARK ACKERMAN, University of Michigan, USA
JUSTINE ZHANG, University of Michigan, USA

1 THE SUBJECT OF REPRESENTATIONAL HARM

This position paper wishes to examine what is meant by "representational harm" in algorithmic systems [6]. The notion of data representativeness we are arguing for here is that of "representational capacities"—ones that are realized, crucially, in acts of cultural articulation [3]. Through a few empirical examples, we aim to show that cultural referents have a visual "plasticity" [2] due to the partiality of the data collected about them and that the harm-bearing T2I representations themselves can instead be rethought of as involved in subjects' syncretic cultural reconfigurations.

A core issue posed by biased data is representational harms, which, canonically, refers to the unfavorable depictions that text-to-image (T2I) systems can generate of marginalized populations, affecting users' beliefs with negative stereotypes and thus reproducing social hierarchies [1, 5]. We agree these occur, but we would like to complicate the story, allowing further refinements about representational harms and representativeness. In our discourse analysis of 142 genAI images shared by Mongolians on Meta, we saw the inherent harmfulness of stereotypical representations becomes less urgent when these systems are taken up by cultural subjects beyond laboratory settings. One of our participants, for example, stated about T2I depictions of Mongolian historical figures: "It's shameful for some fellows, [to act] as if they have seen what they [Mongolian khans] looked like in their time, held their flags, and fought in their wars, to feign knowledge of how something was or wasn't. Let them [users of T2I systems] produce ... various imaginations using the innovations of technology. Why not, what's wrong with it? It feels rather interesting." In this case, the subject deprioritizes the urge to normatively discern whether an AI representation is faithful to the actual khans (or stereotypical of them); instead, they actively encourage imaginative genAI usages based on the khans' portraits as "source." Equally importantly, in pointing to the impossibility of the critics having experienced the events themselves, the participant highlights the visual "plasticity" of cultural referents like the khans: partial knowledge of historical figures allows for continuous political manipulations of their visual representations [2]. This quote demonstrates that, within the provisional schematic of data and referent, where the former is assumed to be representative of the latter, the partiality of the collected data imposes an upper bound on how representative the images can actually be about cultural referents. Instead, the visual plasticity positions both the "source portraits" and the genAI "imaginations" on equal footing in terms of their relation to their referent. Both are active renditions of the referent – or, simply, cultural artifacts – produced by subjects for personally meaningful purposes. Conceptualizing the interactions with T2I systems primarily as making of cultural artifacts, rather than direct subjection to representational harms, allows us to widen the social and cultural practices that T2I systems can be thought to mediate.

2 TOWARDS ARTICULATION

One of such practices, mediated by T2I systems, is articulation. Through the following example we will see how a cultural community selectively and syncretically joins inner elements with external ones to reconfigure itself as an

Authors' addresses: Nasanbayar Ulzii-Orshikh, nulziior@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Mark Ackerman, ackerm@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Justine Zhang, tisjune@umich.edu, University of Michigan, Ann Arbor, Michigan, USA.

"articulated ensemble" rather than an "organic structure," whose minor modification would logically imply its end [3]. A point of contestation among our participants was the genAI-made images of alternative Olympic uniforms for 2024, which many received positively in light of the state-sponsored ones, made by native Mongolian designers, that they deemed to be "Chinese-looking." The participants found themselves constrained in two cultural dimensions: first, the need to adhere (for themselves) to 'original' Mongolian attributes that historically helped distinguish them from their Manchurian rulers; second, the concern about the legibility of the uniforms as Mongolian for the global Other. Presented with a genAI-made uniform that featured symbols perceived as 'originally' Mongolian, several participants strategically preferred to remove them, so that the uniform is legible, paradoxically, *as* Mongolian for the global audience: "The intricate symbols make us look like we're Chinese, stop it, foreigners wouldn't even understand their significances." Another participant added: "It's better only if a national thing [uniform] remains simple[;] when it becomes overly fancy, extravagant, a Southern [China and broadly East Asia] style enters and it [the uniform] makes no sense, it seems." These participants preferred to exclude the intricate ornaments in the interest of "simpl[icity]," for it would help distinguish them in the eyes of the foreign Olympics watchers. Intentionally withholding elements that are considered to be indeed 'originally' (or even essentially) Mongolian in order to construct a cleaner legibility for the global Other is an example of articulation [3].

While the current literature on T2I systems' representational harms predominantly focuses on researcher-to-system direct audits, the broader algorithmic literature does offer several subject-centric conceptions of harm. For instance, Qadri et al. [7] demonstrate that T2I systems can impose a "regime of representation" that enacts cultural erasure upon marginalized communities. van Nuenen et al. [8] highlight the rich tension from the "representational options" of a platformic system constantly interrupting the otherwise dynamically reconstituting identities of intersectional users. Karizat et al. [4] show the ways social media users can find themselves always-already outside the normative values that algorithms entail. Thus, by posing articulation as a T2I system-appropriate analytic that foregrounds the subject's tactical moves within the dominant empirical vignette of its representational harms, we hope to help develop a subject-centric language useful for T2I systems and the syncretic representational acts they mediate. As part of this effort, we juxtapose the normative questions canonically posed to the T2I system representations with the equally, if not more, urgent need to normatively discern the structures of cultural subjectification that our participants attempt to surmount through their selective usage of these representations. We look forward to engaging in conversations about tactical algorithmic assemblages and data annotation mechanisms.

REFERENCES

[1] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5454–5476. https://doi.org/10.18653/v1/2020.acl-main.485

[2] Isabelle Charleux. 2008. Chinggis Khan: Ancestor, Buddha or Shaman? On the uses and abuses of the portrait of Chinggis Khan. *Mongolian Studies* 30/31 (2008), 207–258.

[3] James Clifford. 2001. Indigenous Articulations. *The Contemporary Pacific* 13, 2 (Sept. 2001), 467–490. https://doi.org/10.1353/cp.2001.0046

[4] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 305 (Oct. 2021), 44 pages. https://doi.org/10.1145/3476046

[5] Jared Katzman, Angelina Wang, Morgan Scheuerman, Su Lin Blodgett, Kristen Laird, Hanna Wallach, and Solon Barocas. 2023. Taxonomizing and Measuring Representational Harms: A Look at Image Tagging. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 12 (June 2023), 14277–14285. https://doi.org/10.1609/aaai.v37i12.26670

[6] Alexandra Sasha Luccioni, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. 2023. Stable bias: evaluating societal representations in diffusion models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 2458, 14 pages.

[7] Rida Qadri, Renee Shelby, Cynthia L. Bennett, and Remi Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 506–517. https://doi.org/10.1145/3593013.3594016

[8] Tom van Nuenen, Jose Such, and Mark Cote. 2022. Intersectional Experiences of Unfair Treatment Caused by Automated Computational Systems. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 445 (Nov. 2022), 30 pages. https://doi.org/10.1145/3555546