Evaluating Refusal

Shira Abramovich
McGill University
Montréal, Canada
shira.abramovich@gmail.com

Anna J. Ma
McGill University
Montréal, Canada
anna.ma@mail.mcgill.ca

Abstract

How might we find a place for refusal within the evaluation of Generative AI (GenAI) systems? Current evaluation frameworks justifiably focus on possible uses of models. Given the myriad unsolved issues in GenAI systems and their rapid rise, some developers and potential users are rejecting their use in both public and private settings [22]. Respecting the autonomy of users means respecting their decision *not* to use these technologies. Based on literature on refusal and data ethics, we provide several provocations positing that refusal is a generative act, and advocate the inclusion of refusal in evaluation frameworks.

1 Introduction

In this provocation paper, we propose refusal as a generative response to Generative AI (GenAI) systems, and one that is worth incorporating into evaluation frameworks. As Sara Ahmed asserts [1], refusal is a practice of "saying no without being given the right to say no." Refusal can manifest as an action or an orientation in the world, but it is always mediated by greater systems of power [26]. Inspired by scholarship on refusal in data ethics, we follow the relationship between GenAI evaluations and refusal in four loose themes.

2 Centering Refusal in Evaluation Practices

2.1 Interrogating the Potential for Change and Refusal in Evaluations

Provocation 1: Most evaluations are reformist reforms. Evaluations of GenAI for social impact are *reforms* because they only change a small part of the technology development pipeline. Philosopher André Gorz articulates a difference between "reformist reforms," which prioritize what is practical in an existing system, and "non-reformist reforms," which rearrange structures of power [12]. We propose that evaluations both *are* and *encourage* reformist reforms, as they do not undermine current relations of power. Further, harm-focused social impact evaluation frameworks may act as a way to placate dissent and organizing for real change.

Call to Action 1: Focus on a long-term goal rather than incrementalist improvements. As Ben Green argues [14], long-term goals can help refocus tech work towards non-reformist aims. In the context of evaluation frameworks, this might mean transparency about and disruption of coercive power relations between stakeholders. By adopting a structural perspective and a long-term vision for justice, designers of evaluation frameworks can also avoid what Zong and Matias describe as a coercive presupposition that non-users will automatically become users once placated [26]. To this end, we suggest that evaluations explicitly examine the long-term goals of GenAI systems and foreground the political implications of widespread long-term use on both users and non-users.

2.2 Decentering Technical Expertise as Refusal

Provocation 2: Evaluations further center technical expertise. Requiring technical evaluations for a critique to be seen as legitimate contributes to the silencing of marginalized critique. Chelsea Barabas argues that meaningful critique can only arise when we are able to reorient our critical gaze toward powerful system actors and reframe interventions like evaluations so that they play a *supporting* role in the critique voiced by those impacted [2].

Call to Action 2: Decenter tech and computer science (CS) academia as arbiters of truth. Decentering technologists in the evaluation of algorithmic systems requires building relationships

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

with impacted populations and trusting their critiques without needing the validation of an empirical evaluation. Barabas calls this process "re-centering the margins," and presents it as an important modality of refusal [2]. Barabas also recounts that many technologists who successfully effect social change with harmed groups often use boring, conventional technical methods that are not always valued in CS academic publishing [2]. Decentering CS academia involves venturing beyond disciplinary silos and learning to value and integrate perspectives from the humanities and social sciences that are accompanied with or without academic credentials.

2.3 Acknowledging the Validity of Refusal

Provocation 3: Current evaluations fail to acknowledge the validity of refusal. Evaluations presuppose the continued development of generative systems, thus contributing to a climate in which the use of GenAI technologies is not seen as the political—if forced—choice that it is. As Ruha Benjamin argues [5], "it is coercive to say one has a choice, when one of those choices is automatically penalized". We dub such choices "coerced choices."

Call to Action 3: Support evaluations that encourage real user agency. Evaluation developers must start recognizing that refusal is a productive tool for evaluating GenAI technologies. Zong and Matias, for instance, elucidate autonomy (the capacity to freely make informed choices), time (the timescale in which refusal operates), power (the capacity to produce a change), and cost (the negative ramifications of refusal) as the four constituent elements of refusal from below (i.e., from the margins) [26]. Considering such axes of refusal when building evaluation frameworks can help facilitate the ability of users to refuse GenAI systems.

Ultimately, we urge evaluation developers to create frameworks which consider the implications of widespread GenAI use for both users and non-users. This type of evaluation framework goes beyond the technical properties and performance of the system to consider the social and institutional context of deployment for *both* users and non-users.

2.4 Recognizing Refusal as a Generative Practice

Provocation 4: Evaluations are part of an expansionist tech culture. There is a pervasive view within tech culture that exclusion from technology "always and necessarily involves inequality and deprivation" [25], and that expanded use is *positive*, despite known risks of certain technologies. This mentality informs evaluation frameworks for GenAI that do not consider refusal as a valid avenue to pursue. As a result, many designers and developers in CS fail to consider the refusal of technology as a legitimate act worthy of further inquiry [25].

Call to Action 4: View refusal as generative. We encourage a mindset shift that embraces limits as generative, in order to promote evaluation frameworks that treat limits as productive boundary-setting rather than a problem to be solved. As Benjamin notes, refusal is "seeded with a vision of what can and should be" [5]. Moreover, Zong and Matias argue that acts of refusal can be considered a form of participation in the design process of the socio-technical systems they seek to change [26]. Seeta Gangadharan explains that "when marginalized people refuse technologies, they imagine new ways of being and relating to one another in a technologically mediated society" [9].

Refusal can also initiate behavioral change [26] and generate even broader systemic change by reconfiguring systems of power entirely, as seen in the case of indigenous data sovereignty [21]. Finally, refusal can also motivate software design innovation. In the past, the refusal of corporate information systems has generated grassroots information communication technology infrastructure [15], novel experimentation infrastructure [17], and new digital tools for preserving privacy [6]. As a first step, we recommend that system builders seek understandings of how and why users may refuse GenAI systems, in order to incorporate these perspectives into the design process.

3 Conclusion

In this provocation paper, we have argued that refusal—an act initiated primarily by people with low political power over technology—has generative potential and should be taken seriously in any discussion of evaluation frameworks. As Benjamin argues [5], there is a need to *institutionalize* refusal so as to support people's capacity to collectively organize and challenge power. In this respect, popularizing evaluation frameworks that consider refusal could go a long way.

References

[1] Sara Ahmed. 2017. No. https://feministkilljoys.com/2017/06/30/no/

[2] Chelsea Barabas. 2022. Refusal in Data Ethics: Re-imagining the Code Beneath the Code of Computation in the Carceral State. https://doi.org/10.2139/ssrn.4094977

[3] Solon Barocas, Asia J. Biega, Benjamin Fish, Jędrzej Niklas, and Luke Stark. 2020. When not to design, build, or deploy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 695. https://doi.org/10.1145/3351095.3375691

[4] Eric P.S. Baumer, Morgan G. Ames, Jed R. Brubaker, Jenna Burrell, and Paul Dourish. 2014. Refusing, limiting, departing: why we should study technology non-use. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/2559206.2559224

[5] Ruha Benjamin. 2016. Informed Refusal: Toward a Justice-based Bioethics. *Science, Technology, & Human Values* 41, 6 (Nov. 2016), 967–990. https://doi.org/10.1177/0162243916656059

[6] Finn Brunton and Helen Nissenbaum. 2016. *Obfuscation: a user's guide for privacy and protest* (first mit press paperback edition ed.). The MIT Press, Cambridge, Massachusetts London.

[7] Alicia DeVrio, Motahhare Eslami, and Kenneth Holstein. 2024. Building, Shifting, & Employing Power: A Taxonomy of Responses From Below to Algorithmic Harm. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1093–1106. https://doi.org/10.1145/3630106.3658958

[8] Carl DiSalvo, Annabel Rothschild, Lara L. Schenck, Ben Rydal Shapiro, and Betsy DiSalvo. 2024. When Workers Want to Say No: A View into Critical Consciousness and Workplace Democracy in Data Work. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 156:1–156:24. https://doi.org/10.1145/3637433

[9] Seeta Gangadharan. 2021. 4. Digital Exclusion: A Politics of Refusal. In *Digital Technology and Democratic Theory*, Lucy Bernholz, Hélène Landemore, and Rob Reich (Eds.). University of Chicago Press, 113–140. https://doi.org/10.7208/9780226748603-005

[10] Patricia Garcia, Tonia Sutherland, Marika Cifor, Anita Say Chan, Lauren Klein, Catherine D'Ignazio, and Niloufar Salehi. 2020. No: Critical Refusal as Feminist Data Practice. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 199–202. https://doi.org/10.1145/3406865.3419014

[11] Patricia Garcia, Tonia Sutherland, Niloufar Salehi, Marika Cifor, and Anubha Singh. 2022. No! Re-imagining Data Practices Through the Lens of Critical Refusal. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 315:1–315:20. https://doi.org/10.1145/3557997

[12] André Gorz. 1968. *Strategy for labor: a radical proposal*. Beacon Press, Boston.

[13] Sandy Grande. 2018. Refusing the university. In *Toward what justice?*, Eve Tuck and K. Wayne Yang (Eds.). Routledge, New York, 47–65.

[14] Ben Green. 2019. "Good" isn't good enough. In *Proceedings of the AI for Social Good workshop at NeurIPS*, Vol. 17.

[15] Arne Hintz and Stefania Milan. 2009. At the margins of Internet governance: grassroots tech groups and communication policy. *International Journal of Media & Cultural Politics* 5, 1 (March 2009), 23–38. https://doi.org/10.1386/macp.5.1-2.23_1

[16] Nari Johnson, Sanika Moharana, Christina Harrington, Nazanin Andalibi, Hoda Heidari, and Motahhare Eslami. 2024. The Fall of an Algorithm: Characterizing the Dynamics Toward Abandonment. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Rio de Janeiro Brazil, 337–358. https://doi.org/10.1145/3630106.3658910

[17] J. Nathan Matias and Merry Mou. 2018. CivilServant: Community-Led Experiments in Platform Governance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173583

[18] Carole McGranahan. 2016. Theorizing Refusal: An Introduction. *Cultural Anthropology* 31, 3 (Aug. 2016), 319–325. https://doi.org/10.14506/ca31.3.01

[19] Procreate. 2024. Creativity is made, not generated. https://procreate.com/ai

[20] Audra Simpson. 2017. The ruse of consent and the anatomy of 'refusal': cases from indigenous North America and Australia. *Postcolonial Studies* 20, 1 (Jan. 2017), 18–33. https://doi.org/10.1080/13688790.2017.1334283

[21] C Matthew Snipp. 2016. What does data sovereignty imply: what does it look like? In *Indigenous Data Sovereignty* (1st ed.), Tahu Kukutai and John Taylor (Eds.). ANU Press. https://doi.org/10.22459/CAEPR38.11.2016.03

[22] Irene Solaiman, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Canyu Chen, Hal Daumé III, Jesse Dodge, Isabella Duan, Ellie Evans, Felix Friedrich, Avijit Ghosh, Usman Gohar, Sara Hooker, Yacine Jernite, Ria Kalluri, Alberto Lusoli, Alina Leidinger, Michelle Lin, Xiuzhu Lin, Sasha Luccioni, Jennifer Mickel, Margaret Mitchell, Jessica Newman, Anaelia Ovalle, Marie-Therese Png, Shubham Singh, Andrew Strait, Lukas Struppek, and Arjun Subramonian. 2024. Evaluating the Social Impact of Generative AI Systems in Systems and Society. arXiv:2306.05949 (June 2024). http://arxiv.org/abs/2306.05949 arXiv:2306.05949 [cs].

[23] Eve Tuck and K. Wayne Yang. 2014. R-Words: Refusing Research. In *Humanizing Research: Decolonizing Qualitative Inquiry with Youth and Communities*, Django Paris and Maisha T. Winn (Eds.). SAGE Publications, Inc., 1 Oliver's Yard, 55 City Road London EC1Y 1SP, 223–247. https://doi.org/10.4135/9781544329611

[24] Eve Tuck and K. Wayne Yang. 2014. Unbecoming Claims: Pedagogies of Refusal in Qualitative Research. *Qualitative Inquiry* 20, 6 (July 2014), 811–818. https://doi.org/10.1177/1077800414530265

[25] Sally Wyatt. 2005. Non-users also matter: The construction of users and non-users of the Internet. In *How Users Matter: The Co-Construction of Users and Technology*, Nelly Oudshoorn and Trevor Pinch (Eds.). MIT Press, Cambridge, MA.

[26] Jonathan Zong and J. Nathan Matias. 2024. Data Refusal from Below: A Framework for Understanding, Evaluating, and Envisioning Refusal as Design. *ACM J. Responsib. Comput.* 1, 1 (March 2024), 10:1–10:23. https://doi.org/10.1145/3630107

A Appendix

A.1 Background

A.1.1 Refusal

Refusal as a concept has long been theorized in Indigenous and feminist scholarship [11]. We credit our conception of refusal to prominent theorists in Indigenous Studies such as Audra Simpson [20], Sandy Grande [13], Eve Tuck [23], and K. Wayne Yang [24]. In their work, refusal is not only an academic theory, but also a methodological and political stance embodied in anticolonial practices that respond against settler colonialism [20]. Recognizing that refusal emerged from the context of anticolonial struggle is critical to understanding its theorization.

Simply put, refusal "is to say no" but "it is not just that" [18]. Refusal is also a generative practice that envisions liberatory alternatives to current forms of domination by intentionally setting boundaries in rejection of the status quo. To refuse is to articulate and embrace that a limit has been reached, and to develop and negotiate alternative practices in light of that limit. For instance, Tuck and Yang advocate for refusal as an analytic practice in qualitative research to limit the reach of the academy's commonly voyeuristic, extractive, and predatory lines of inquiry into the pain and humiliation of

Othered communities [24]. They argue that "there are some forms of knowledge that the academy doesn't deserve", and that "not selling" and "not telling" at times can be compelling acts of refusal that reinforce the sovereignty of Othered individuals, who would be otherwise objectified in the settler colonial knowledge production process, where there is no reciprocity between researchers and those researched [24, 23]. Tuck and Yang note that academic research is not always an appropriate or helpful intervention in the genuine interests of Othered communities, but refusal can generate the space for other decolonial forms of knowledge and representation to surface.

Benjamin adds that "the [generative] potential of refusal [is] not only to negate colonial forms of knowledge production but also to create new, more equitable relationships between researchers [and] subjects" [5].

A.1.2 Technology Refusal

The refusal of technology has been increasingly examined by scholars and technologists at communities like FAccT [3, 7, 16], CSCW [8, 10, 11], and CHI [4]. The question of when and how to refuse participation in harmful technology systems continues to grow more and more pertinent from both the perspectives of developers and users [3].

An example of institutionalizing technology refusal is the decision made by Procreate, a popular iPad app for creating digital art, to deviate from industry norms and *refuse* to incorporate GenAI into their software due to their explicit belief that GenAI is predicated on data theft, and that it undermines and devalues human creativity [19].

A.2 Limitations

This tiny paper is quite short and is mostly a theoretical engagement with issues in evaluation frameworks for GenAI. As such, it does not point out finer-grained provocations that might arise from case studies or empirical validation of certain evaluation frameworks' responses to or incorporation of refusal.

A.3 Broader Impacts

This paper proposes the incorporation of refusal into evaluation frameworks. This proposal aims to provide further impetus for technologists to trust marginalized critiques of GenAI technologies—in effect, it encourages technologists to trust and react to reports of harmful broader impacts, and to consider the broader impacts on affected communities of evaluation frameworks which do not fully engage with marginalized critiques.

At the same time, incorporating refusal always has the potential to introduce barriers to the positive and important use of technology. We have attempted to stress the *incorporation* of refusal into frameworks, rather than its prioritization over all other considerations, to combat this.