From Decisions to Multiplicity: Frameworks, Theories, and Applications

Prakhar Ganesh[1,2]

[1]McGill University,
[2]Mila - Quebec AI Institute
prakhar.ganesh@mila.quebec

Abstract

Model development in AI is shaped by developer decisions. While there is significant research on the opportunities and risks of multiplicity, little attention has been paid to how developer decisions impact multiplicity. My thesis focuses on (a) introducing broader frameworks to better analyze developer decisions in AI, (b) identifying theoretical connections to characterize the influence of these decisions on multiplicity, and (c) operationalizing insights across various applications, thus building responsible AI models with multiplicity.

Motivation

Artificial intelligence (AI) is experiencing a remarkable rise (Yuan 2023), and its increasing deployment has sparked concerns about patterns of biased treatment (fairness) (Mehrabi et al. 2021), brittleness to distribution shifts (robustness) (Subbaswamy, Adams, and Saria 2021), and information leakage risks (privacy) (Carlini et al. 2023), among others, studied under the umbrella of *responsible AI*. A crucial aspect of building responsible models is the concept of multiplicity (Black, Raghavan, and Barocas 2022).

What is multiplicity? Data serves as a proxy for the real world, but cannot fully capture its nuances. There is a fundamental loss of information in data-driven learning, leaving a gap where multiple interpretations (i.e., AI models) can arise, studied under the epistemological framework of *Rashomon effect*. The disagreement between interpretations under this Rashomon effect is known as *multiplicity*. While the study of multiplicity is not new (Breiman 2001), recent research has broadened the discourse on multiplicity to its real-world impact and potential in advancing responsible AI practices (Ganesh, Taik, and Farnadi 2025).

Multiplicity and Responsible AI. When multiple interpretations exist, some of them may exhibit certain desirable properties, such as better fairness, robustness, interpretability, etc. Thus, if managed well, multiplicity can help us select or aggregate models to minimize harm (Black, Raghavan, and Barocas 2022; Rudin et al. 2024). However, multiplicity also marks the inherent arbitrariness in data-driven learning that impacts individuals, necessitating a discussion on the expectations of automated decision makers in our society (Jain, Creel, and Wilson 2024; Creel and Hellman 2022; Kulynych et al. 2023).

Objective

Model development in AI is shaped by developer decisions. While there is significant research on the opportunities and risks of multiplicity, little attention has been paid to how developer decisions impact multiplicity. *My thesis focuses on (a) introducing broader frameworks to better situate and analyze developer decisions in AI (WP1), (b) identifying theoretical connections to characterize the influence of these decisions on multiplicity (WP2), and (c) operationalizing these insights across various applications, thus building responsible AI models with multiplicity (WP3, WP4).*

Work Packages

Outlined below are four work packages (WPs) essential to achieving the main objective of my thesis.

WP1 - Developer Decision Framework (Finished). Multiplicity is often studied through the lens of information loss, i.e., incomplete information can lead to multiple interpretations. While this perspective is valuable, it is limited in providing an operational framework for developers. Instead, attention should be directed toward the role of developer choices in model design, which lays a practical foundation for managing multiplicity. We proposed the *intention-convention-arbitrariness* (ICA) framework (Ganesh, Taik, and Farnadi 2025) to study developer choices. By structuring these choices, we promote transparency and connect developer decisions to subfields in multiplicity on creating steerable models (*intentional choices*), homogenization (*conventional choices*), and arbitrariness (*arbitrary choices*).

WP2 - Data Processing, Model Maintenance, and Multiplicity. Recent works have focused on understanding how developer choices during model training influence the resulting model, and thus, in turn, multiplicity. However, far less is known about the effects of decisions made during data processing before training and model maintenance, i.e., updates after deployment, on multiplicity. This gap is critical given the complexity of these relationships: directly predicting model behavior from data without training remains a dif-

ficult challenge, and characterizing how model updates behave under distribution shifts is equally non-trivial. We currently lack a strong theoretical foundation for understanding how these choices shape multiplicity. **Long-term expected outcome:** Research paper(s) that establish strong foundational connections between data preprocessing and model maintenance decisions and their impact on multiplicity. **Expected outcome by Doctoral Consortium:** A research paper that studies the impact of multiplicity on algorithmic recourse under distribution shift. **Current progress:** In a recent work (currently under review), we identified the impact of data preprocessing choices on the resulting multiplicity, without the need to explicitly train models, and even provided the tools for developers to steer multiplicity.

WP3 - Model Selection and Benchmarking under Multiplicity (Finished). Multiplicity introduces a significant challenge in benchmarking models, as a one-dimensional comparison of algorithms without accounting for multiplicity can create a false sense of superiority (Ganesh et al. 2025). Improving our understanding of the role of multiplicity in model benchmarking is thus imperative to tackling multiplicity in real-world applications. In our works, we have studied the role of training stochasticity in fairness (Ganesh et al. 2023), highlighting the dominant impact of data order. We also showed a similar concern of variance across several responsible AI metrics beyond fairness (Ganesh 2024), proposing the use of *multiplicity sheets* when benchmarking models to tackle multiplicity. Most recently, we recognized the value of context when benchmarking bias mitigation algorithms (Ganesh et al. 2025), showing that different techniques might be the "best" depending on the context of model development.

WP4 - Multiplicity in the Age of GenAI. The recent adoption of generative AI (GenAI) has reshaped how developers design and deploy models (Yuan 2023). Rather than training task-specific models, we see an increasing reliance on powerful pre-trained models (Dhar 2024). In this paradigm, designing effective prompts—not models—is the new objective. Multiplicity in GenAI both exacerbates existing challenges, such as increased risk of homogenization when everyone relies on a few model providers, and introduces new ones, including prompt multiplicity, reward multiplicity, and preference multiplicity, among others. **Long-term expected outcomes:** Research and Position paper(s) highlighting the value of learning from the existing multiplicity literature in the age of GenAI. **Current progress:** We have shown that studying prompts through the lens of multiplicity can help us quantify the true risk of information leakage in LLMs (More, Ganesh, and Farnadi 2025), more precisely evaluate hallucination concerns (Ganesh, Shokri, and Farnadi 2025), and improve real-world applicability of LLM auditing frameworks (Chataigner et al. 2025).

References

Black, E.; Raghavan, M.; and Barocas, S. 2022. Model multiplicity: Opportunities, concerns, and solutions. In *FAccT*.

Breiman, L. 2001. Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3): 199–231.

Carlini, N.; Hayes, J.; Nasr, M.; Jagielski, M.; Sehwag, V.; Tramer, F.; Balle, B.; Ippolito, D.; and Wallace, E. 2023. Extracting training data from diffusion models. In *USENIX*.

Chataigner, C.; Ma, R.; Ganesh, P.; Taïk, A.; Creager, E.; and Farnadi, G. 2025. Say It Another Way: A Framework for User-Grounded Paraphrasing. *arXiv preprint arXiv:2505.03563*.

Creel, K.; and Hellman, D. 2022. The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. *Canadian Journal of Philosophy*, 52(1): 26–43.

Dhar, V. 2024. The paradigm shifts in artificial intelligence. *Communications of the ACM*, 67(11): 50–59.

Ganesh, P. 2024. An Empirical Investigation into Benchmarking Model Multiplicity for Trustworthy Machine Learning: A Case Study on Image Classification. In *2024 IEEE/CVF WACV*. IEEE.

Ganesh, P.; Chang, H.; Strobel, M.; and Shokri, R. 2023. On the impact of machine learning randomness on group fairness. In *FAccT*, 1789–1800.

Ganesh, P.; Gohar, U.; Cheng, L.; and Farnadi, G. 2025. Different Horses for Different Courses: Comparing Bias Mitigation Algorithms in ML. In *Proceedings of the AFME Workshop*, volume 279 of *Proceedings of Machine Learning Research*, 96–118. PMLR.

Ganesh, P.; Shokri, R.; and Farnadi, G. 2025. Rethinking Hallucinations: Correctness, Consistency, and Prompt Multiplicity. In *ICLR 2025 Workshop on Building Trust in Language Models and Applications*.

Ganesh, P.; Taik, A.; and Farnadi, G. 2025. The Curious Case of Arbitrariness in Machine Learning. *Proceedings of the 2025 AAAI/ACM AIES*.

Jain, S.; Creel, K.; and Wilson, A. C. 2024. Position: Scarce Resource Allocations That Rely On Machine Learning Should Be Randomized. In *Forty-first International Conference on Machine Learning*.

Kulynych, B.; Hsu, H.; Troncoso, C.; and Calmon, F. P. 2023. Arbitrary decisions are a hidden cost of differentially private training. In *FAccT*, 1609–1623.

Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6): 1–35.

More, Y.; Ganesh, P.; and Farnadi, G. 2025. Towards more realistic extraction attacks: An adversarial perspective. *TACL*.

Rudin, C.; Zhong, C.; Semenova, L.; Seltzer, M.; Parr, R.; Liu, J.; Katta, S.; Donnelly, J.; Chen, H.; and Boner, Z. 2024. Amazing things come from having many good models. *arXiv preprint arXiv:2407.04846*.

Subbaswamy, A.; Adams, R.; and Saria, S. 2021. Evaluating model robustness and stability to dataset shift. In *AISTATS*, 2611–2619. PMLR.

Yuan, Y. 2023. On the power of foundation models. In *ICML*, 40519–40530. PMLR.