The Many Faces of Multiplicity in ML

Prakhar Ganesh, Carol Long, Afaf Taik, Hsiang Hsu, Jamelle Watson-Daniels, Kathleen Creel, Flavio Calmon, Golnoosh Farnadi

Special thanks: Shomik Jain, Ashia Wilson

Tutorial Objectives

- Highlight the phenomenon of multiplicity in machine learning and call attention to its growing literature.
- Link multiplicity with other measures of instability: uncertainty and churn.
- Discuss the implications of multiplicity for fairness and explanability in algorithmic decision-making.
- Recognize exacerbated concerns and new forms of multiplicity in GenAI.
- Engaging the community on when and how to address multiplicity in various practical scenarios.
- Identifying open questions and motivating future research directions.

Several companies are planning to partially automate their entry-level hiring pipeline. They each receive over 100 applications every week, and recruiters don’t have time to review every application.

They are planning to train models and create automated hiring tools that will select the strongest applications every week for manual review.

Each of you represent a company!
You will all train your own separate models and make hiring decisions.

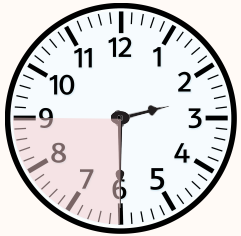
Let's train our own models for automated hiring

Source: dilbert.com; modified.

https://tinyurl.com/multiplicity-demo

Timeline for the Tutorial

Rashomon Set and Multiplicity



Multiplicity, Uncertainty, and Churn

Implications of Multiplicity for Fairness and Explainability

Multiplicity in the Age of GenAI

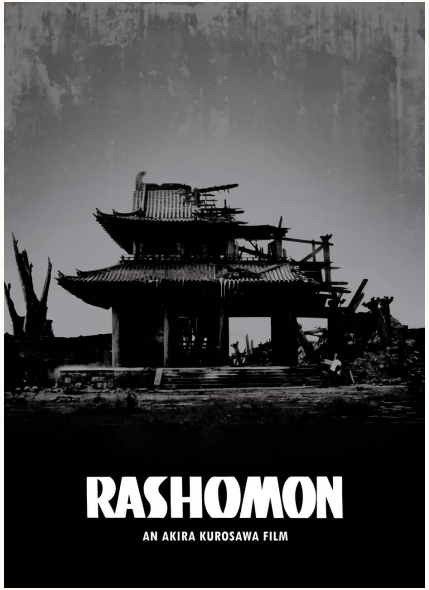
Practical Scenarios and Discussion

Future Directions

Rashomon Set and Multiplicity

Rashomon Effect

Based on *Rashomon (1950)*
by Akira Kurosawa



Source: Poster by Bo Kev (https://fineartamerica.com/featured/rashomon-bo-kev.html)

Rashomon Effect

Rashomon effect is *“a combination of a difference of perspective and equally plausible accounts, with the absence of evidence to elevate one above others…”*

Davis, B., Anderson, R. & Walls, J. (2015). *Rashomon Effects: Kurosawa, Rashomon and their legacies*. Routledge.

Rashomon Effect

Rashomon effect is "*a combination of a **difference of perspective and equally plausible accounts**, with the absence of evidence to elevate one above others…*"

Davis, B., Anderson, R. & Walls, J. (2015). *Rashomon Effects: Kurosawa, Rashomon and their legacies*. Routledge.

Rashomon Effect

Rashomon effect is “*a combination of a difference of perspective and equally plausible accounts, with the* ***absence of evidence to elevate one above others****…”*

Davis, B., Anderson, R. & Walls, J. (2015). *Rashomon Effects: Kurosawa, Rashomon and their legacies*. Routledge.

Rashomon Effect

Rashomon effect is *“a combination of a difference of perspective and equally plausible accounts, with the absence of evidence to elevate one above others…”*

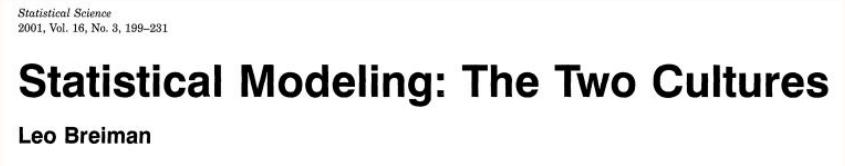
Davis, B., Anderson, R. & Walls, J. (2015). *Rashomon Effects: Kurosawa, Rashomon and their legacies*. Routledge.

Rashomon Effect in AI

*"What I call the Rashomon Effect is that there is often **a multitude of different descriptions** (equations f(x)) in a class of functions giving **about the same minimum error rate**."*

Breiman, L. (2001). Statistical modeling: The two cultures. Statistical science, 16(3), 199-231.

Rashomon Effect in AI



"What I call the Rashomon Effect is that there is often ***a multitude of different descriptions*** *(equations f(x)) in a class of functions giving* ***about the same minimum error rate****."*

Rashomon Set

Breiman, L. (2001). Statistical modeling: The two cultures. Statistical science, 16(3), 199-231.

Rashomon Effect in AI

The World

Annual Income

Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI

Loan Applications

The World

Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI

Loan Applications

The World

A Mathematical Model of the World

Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI

Loan Applications

The World

A Mathematical Model of the World

Chances of repaying depend on Annual Income and Credit Score

Annual Income

Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI

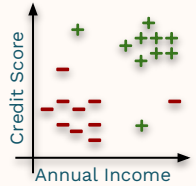
Loan Applications

The World

A Mathematical Model of the World

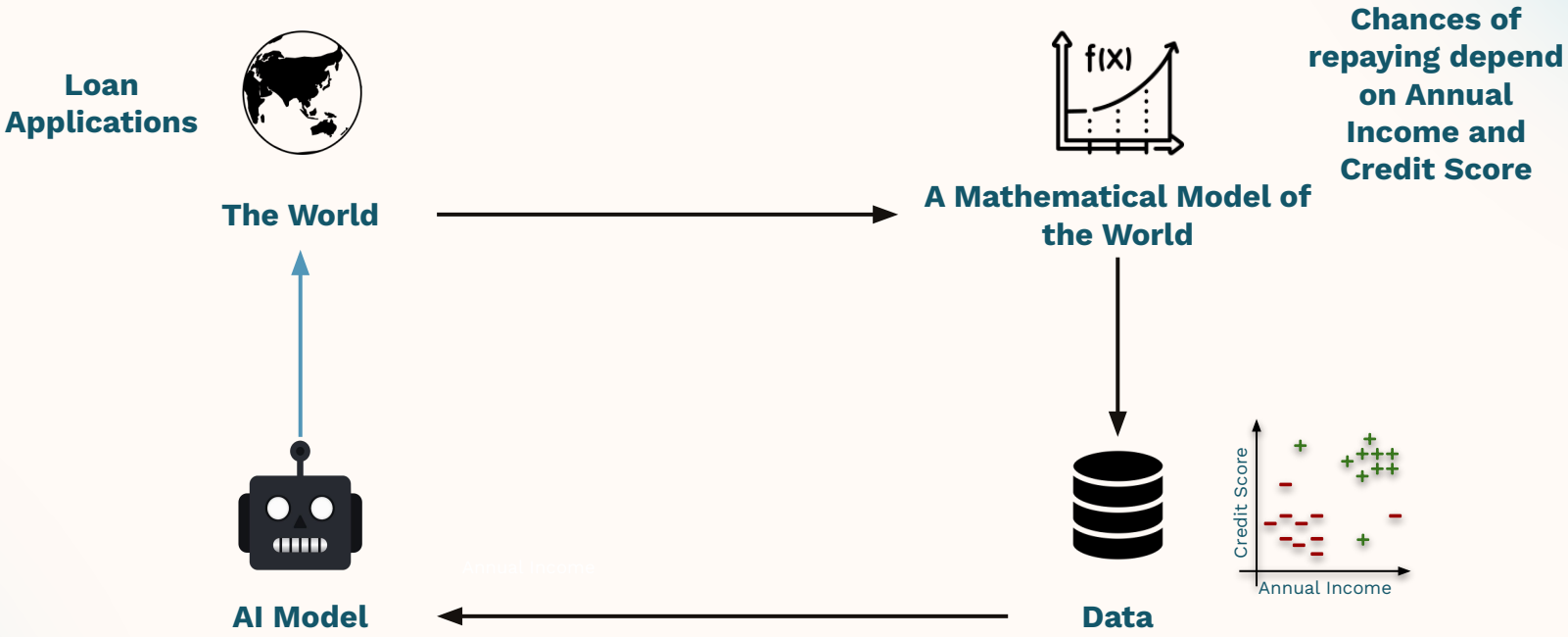
Chances of repaying depend on Annual Income and Credit Score

Data



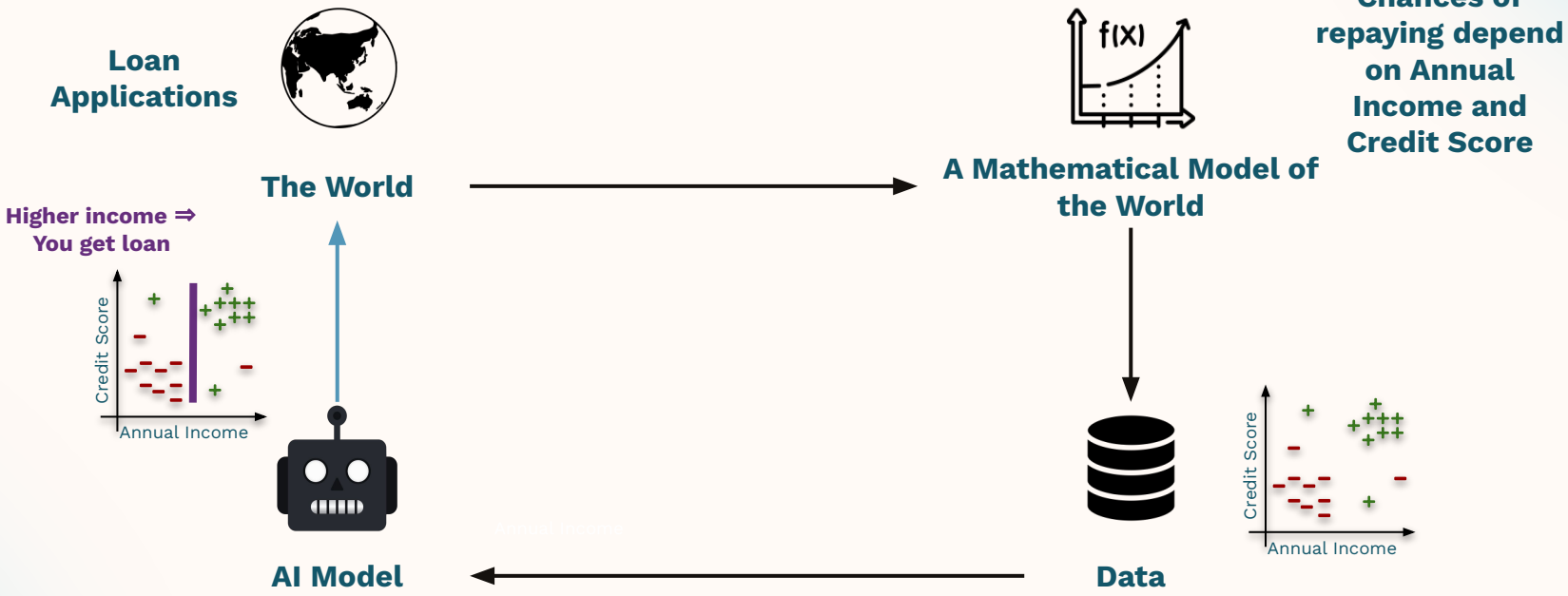
Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI



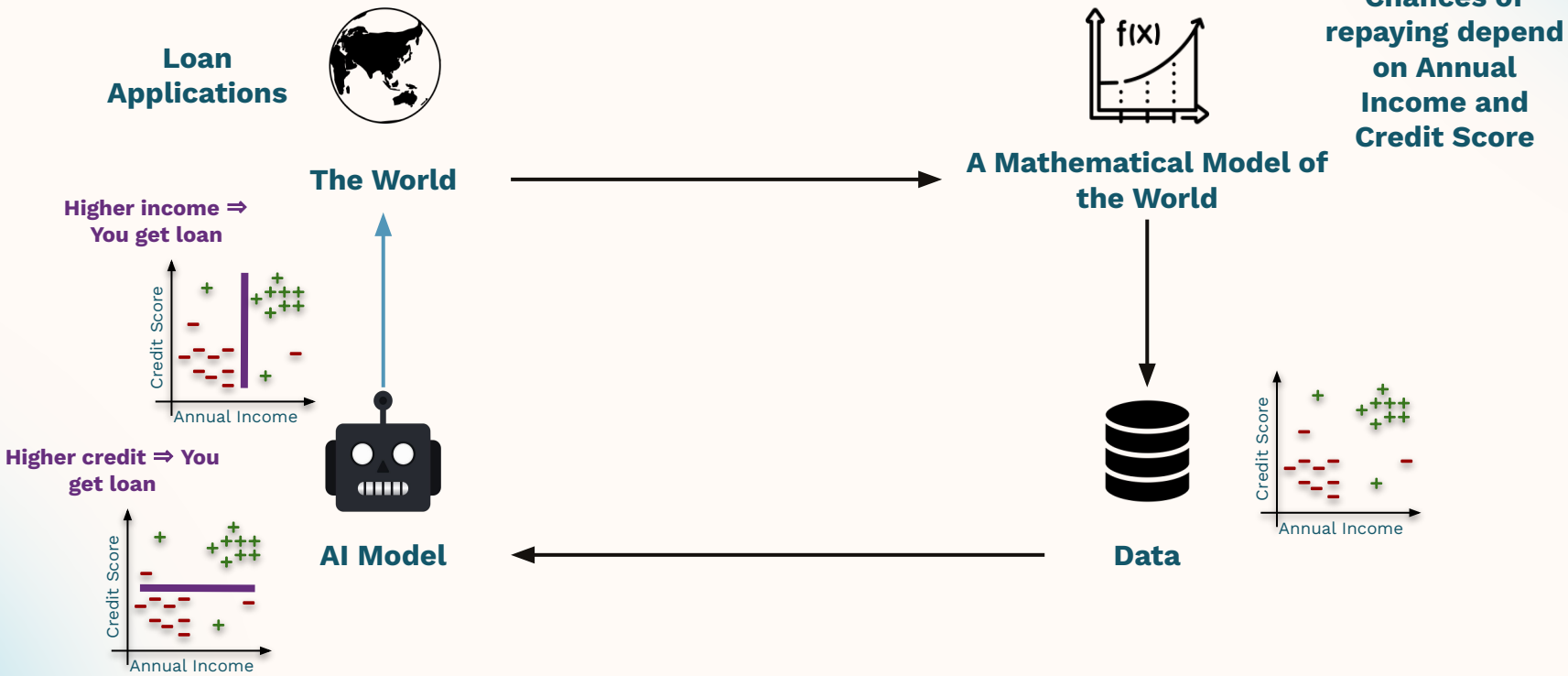
Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI



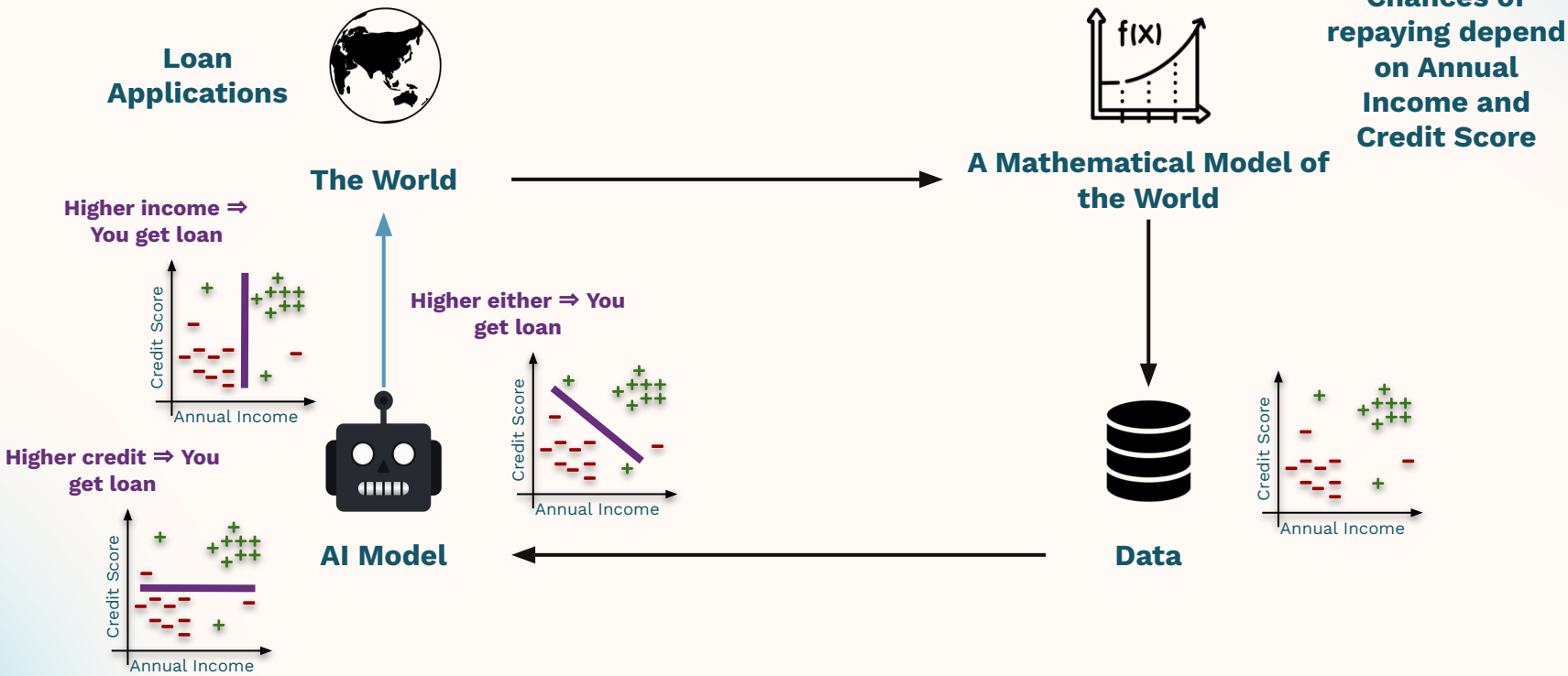
Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Effect in AI



Source: All icons from Flaticon (https://www.flaticon.com/)

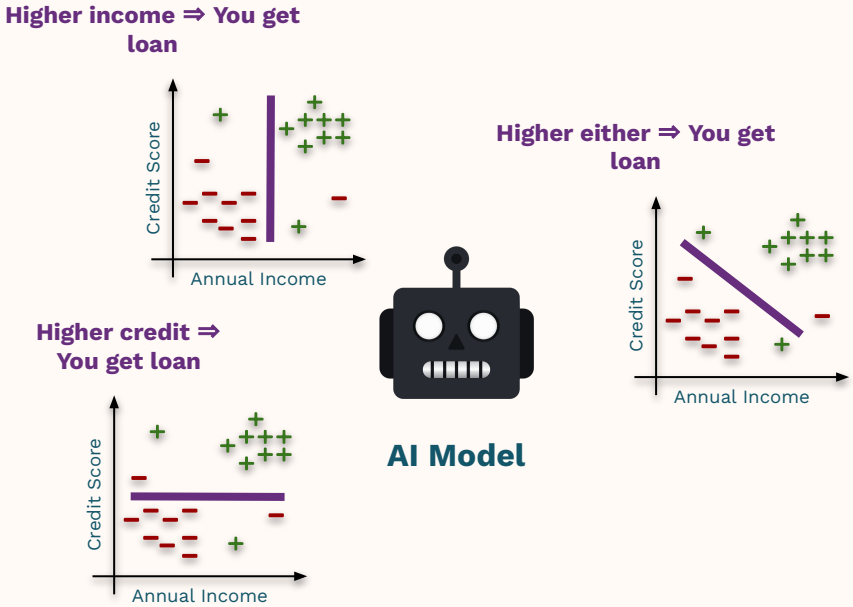
Rashomon Effect in AI



Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Set

Or a set of competing models, a set of good models, ε–Rashomon set, ε–Level set, etc.



Source: All icons from Flaticon (https://www.flaticon.com/)

Rashomon Set

Or a set of competing models, a set of good models, ε–Rashomon set, ε–Level set, etc.

Def (ε-Level Set): Given a baseline classifier h_0 and a hypothesis class $\mathcal{H}$, the ε-level set around h_0 is the set of all classifiers $h \in \mathcal{H}$ with an error rate of at most $L(h_0) + \varepsilon$ on the training data,

$$S_\varepsilon(h_0) := \{h \in \mathcal{H} : L(h) \leq L(h_0) + \varepsilon\}$$

Marx, C., Calmon, F., & Ustun, B. (2020, November). Predictive multiplicity in classification. In International Conference on Machine Learning (pp. 6765–6774). PMLR.

Rashomon Set

Or a set of competing models, a set of good models, ε–Rashomon set, ε–Level set, etc.

Def (ε-Level Set): Given ~~a baseline classifier h_0 and~~ a hypothesis class $\mathcal{H}$, the ε-level set ~~around h_0~~ is the set of all classifiers $h \in \mathcal{H}$ with an error rate of at most ~~$L(h_0) +$~~ ε on the training data,

$$S_\varepsilon := \{h \in \mathcal{H} : L(h) \leq \varepsilon\}$$

Hsu, H., & Calmon, F. (2022). Rashomon capacity: A metric for predictive multiplicity in classification. Advances in Neural Information Processing Systems, 35, 28988–29000.

Rashomon Set

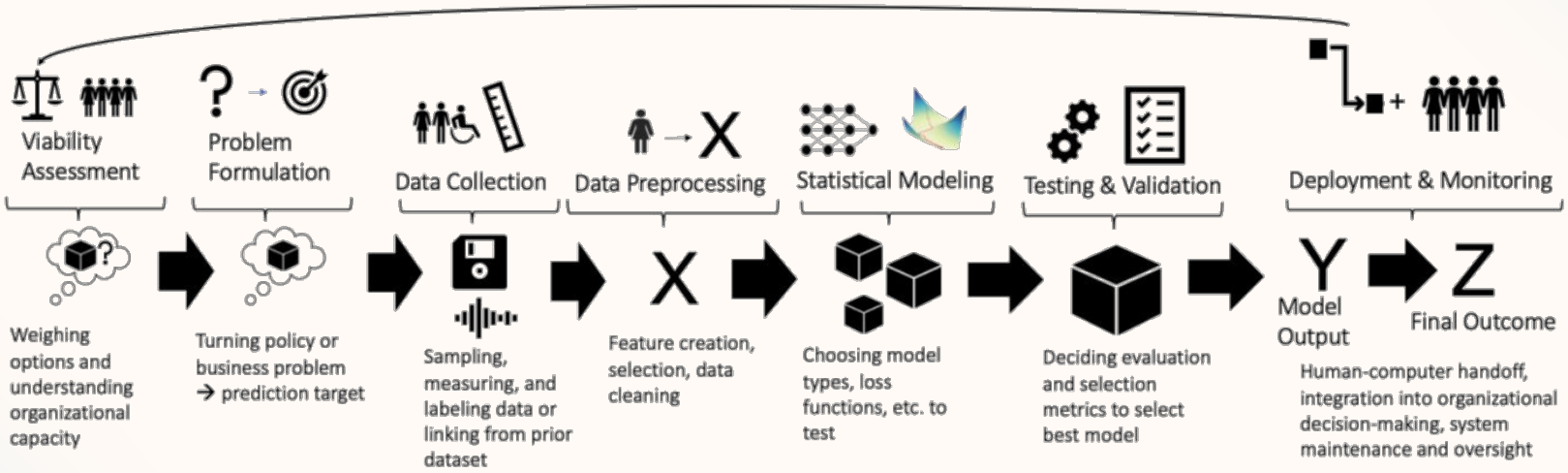
Or a set of competing models, a set of good models, ε–Rashomon set, ε–Level set, etc.

Def (ε-Level Set): ~~Given a baseline classifier h_0 and a hypothesis class $\mathcal{H}$,~~ the ε-level set ~~around h_0~~ is the set of all classifiers h ~~$\in \mathcal{H}$~~ with an error rate of at most ~~$L(h_0) +$~~ ε on the training data,

$$S_\varepsilon := \{h : L(h) \leq \varepsilon\}$$

Ganesh, P., Taik, A., & Farnadi, G. (2025, October). Systemizing multiplicity: The curious case of arbitrariness in machine learning. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 2, pp. 1032-1048).

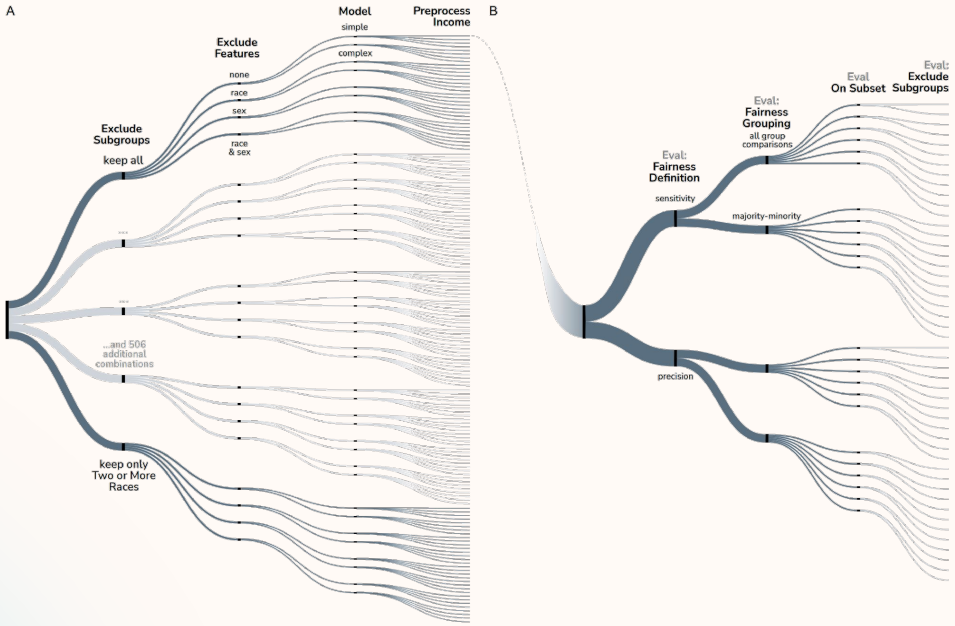
Developer Choices and the Rashomon Effect



Source: Black et al. (2024)

Black, E., Koepke, L., Kim, P., Barocas, S., & Hsu, M. (2024). The Legal Duty to Search for Less Discriminatory Algorithms. arXiv preprint arXiv:2406.06817.

Developer Choices and the Rashomon Effect



Source: Simson et al. (2025)

Simson, J., Draxler, F., Mehr, S., & Kern, C. (2025, April). Preventing harmful data practices by using participatory input to navigate the machine learning multiverse. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1–30).

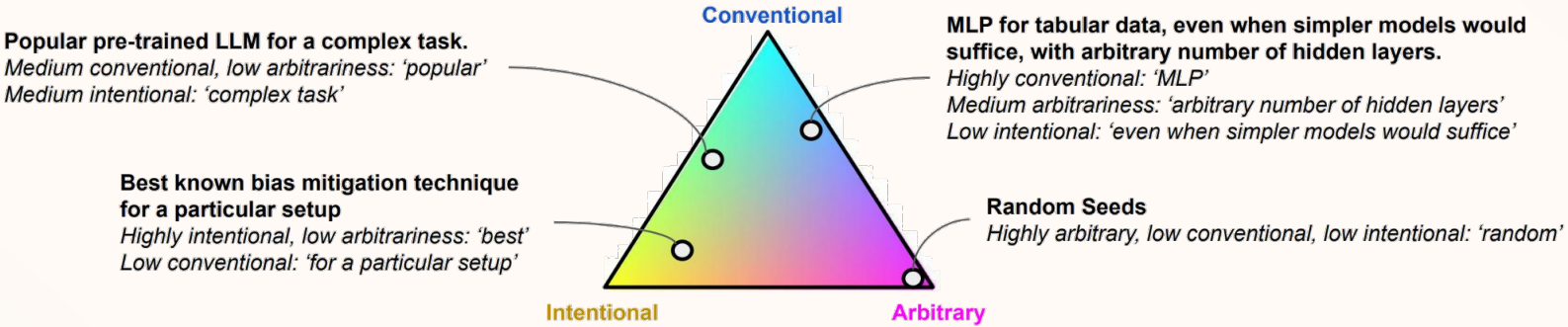
Developer Choices and the Rashomon Effect

"[...] access to the training data alone is typically insufficient for a rigorous audit. The primary reason is that the same training data can induce many different downstream models. [...] This model multiplicity or underspecification is an unavoidable feature of modern AI systems."

"An auditor cannot gain insight into the developers' process and reasoning from the trained model alone, which is not always satisfactory from a broader accountability perspective."

Cen, S. H., & Alur, R. (2024, October). From transparency to accountability and back: A discussion of access and evidence in ai auditing. In Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1–14).

Developer Choices and the Rashomon Effect



Source: Ganesh et al. (2025)

Ganesh, P., Taik, A., & Farnadi, G. (2025, October). Systemizing multiplicity: The curious case of arbitrariness in machine learning. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 2, pp. 1032–1048).

Predictive Multiplicity

*"we define predictive multiplicity as the ability of a prediction problem to admit **competing models** that assign **conflicting predictions**."*

Rashomon Set

Predictive Multiplicity

Marx, C., Calmon, F., & Ustun, B. (2020, November). Predictive multiplicity in classification. In International Conference on Machine Learning (pp. 6765-6774). PMLR.

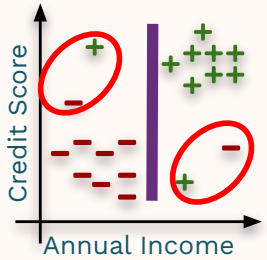
Predictive Multiplicity

Definition (Predictive Multiplicity): Given a baseline classifier h_0, a prediction problem exhibits predictive multiplicity over the ε-level set $S_\varepsilon(h_0)$ if there exists a model $h \in S_\varepsilon(h_0)$ such that $h(\boldsymbol{x}_i) \neq h_0(\boldsymbol{x}_i)$ for some $\boldsymbol{x}_i$ in the training set.

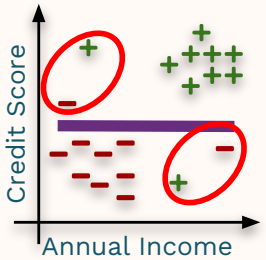
Marx, C., Calmon, F., & Ustun, B. (2020, November). Predictive multiplicity in classification. In International Conference on Machine Learning (pp. 6765-6774). PMLR.

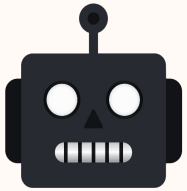
Predictive Multiplicity & Conflicting Outcomes

Higher income ⇒ You get loan



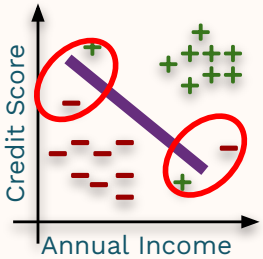
Higher credit ⇒ You get loan





AI Model

Higher either ⇒ You get loan



Source: All icons from Flaticon (https://www.flaticon.com/)

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Marx, C., Calmon, F., & Ustun, B. (2020, November). Predictive multiplicity in classification. In International Conference on Machine Learning (pp. 6765–6774). PMLR.

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Watson-Daniels, J., Parkes, D. C., & Ustun, B. (2023, June). Predictive multiplicity in probabilistic classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 9, pp. 10306-10314).

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Allocation Multiplicity: multiplicity in predictions => multiplicity in allocation outcomes

Jain, S., Wang, M., Creel, K., & Wilson, A. (2025). Allocation Multiplicity: Evaluating the Promises of the Rashomon Set. FAccT.

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Allocation Multiplicity: multiplicity in predictions => multiplicity in allocation outcomes

Procedural Multiplicity: multiplicity in model internals

Black, E., Raghavan, M., & Barocas, S. (2022, June). Model multiplicity: Opportunities, concerns, and solutions. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 850-863).

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Allocation Multiplicity: multiplicity in predictions => multiplicity in allocation outcomes

Procedural Multiplicity: multiplicity in model internals

Dataset Multiplicity: multiplicity in training data => multiplicity in predictions

Meyer, A. P., Albarghouthi, A., & D'Antoni, L. (2023, June). The dataset multiplicity problem: How unreliable data impacts predictions. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 193–204).

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Allocation Multiplicity: multiplicity in predictions => multiplicity in allocation outcomes

Procedural Multiplicity: multiplicity in model internals

Dataset Multiplicity: multiplicity in training data => multiplicity in predictions

Multi-Target Multiplicity: multiplicity in target variables => multiplicity in allocations

Watson-Daniels, J., Barocas, S., Hofman, J. M., & Chouldechova, A. (2023, June). Multi-target multiplicity: Flexibility and fairness in target specification under resource constraints. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 297-311).

The Many Forms of Multiplicity

Predictive Multiplicity for Binary Classification: multiplicity in prediction classes

Predictive Multiplicity for Probabilistic Classification: multiplicity in prediction probabilities

Allocation Multiplicity: multiplicity in predictions => multiplicity in allocation outcomes

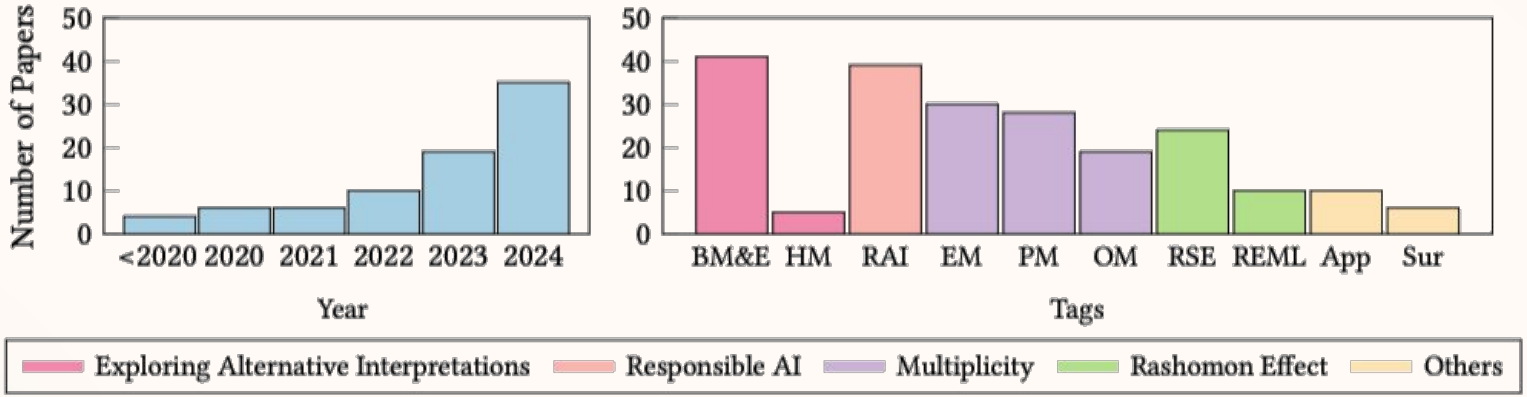
Procedural Multiplicity: multiplicity in model internals

Dataset Multiplicity: multiplicity in training data => multiplicity in predictions

Multi-Target Multiplicity: multiplicity in target variables => multiplicity in allocations

Explanation Multiplicity: multiplicity in model internals => multiplicity in explanations

Growing Literature About Multiplicity



BM&E: Better Models and Ensembles; **HE:** Hacking Metrics; **RAI:** Responsible AI; **PM:** Predictive Multiplicity; **EM:** Explanation Multiplicity; **OM:** Other Multiplicity; **RSE:** Rashomon Set Exploration; **REML:** Rashomon Effect in ML; **App:** Application; **Sur:** Survey.

Source: Ganesh et al. (2025)

Ganesh, P., Taik, A., & Farnadi, G. (2025, October). Systemizing multiplicity: The curious case of arbitrariness in machine learning. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 2, pp. 1032–1048).

Multiplicity, Uncertainty, and Churn

Multiplicity

$$\mathbf{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$$

dataset

$$\mathbf{x}_i = [1, x_{i1}, \cdots, x_{id}] \in \mathbb{R}^{d+1}$$

features

$$y_i \in \{0, 1\}$$

outcome of interest

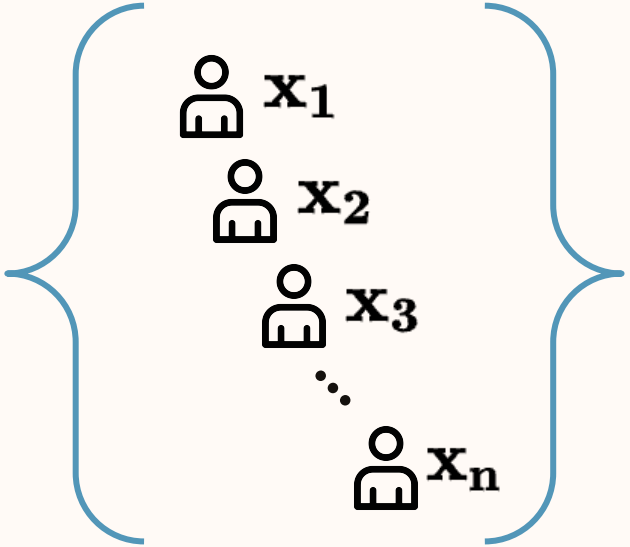
Multiplicity

$$\mathbf{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$$

$$\mathbf{x}_i = [1, x_{i1}, \cdots, x_{id}] \in \mathbb{R}^{d+1}$$

$$y_i \in \{0, 1\}$$

dataset



Model 1

$$p(y = +1|x)$$

$$Acc_{M1}$$

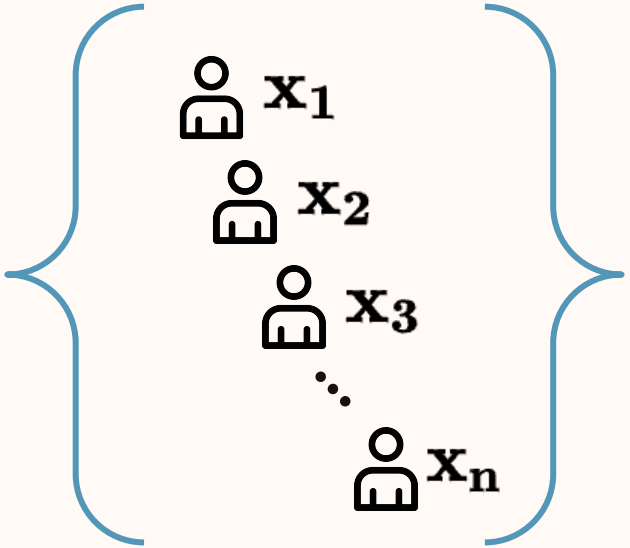
Multiplicity

$$\mathbf{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$$

$$\mathbf{x}_i = [1, x_{i1}, \cdots, x_{id}] \in \mathbb{R}^{d+1}$$

$$y_i \in \{0, 1\}$$

dataset



Model 1

$$p(y = +1|x)$$

$$Acc_{M1}$$

Model 2

$$p(y = +1|x)$$

$$Acc_{M1} - \epsilon$$

Model 3

$$p(y = +1|x)$$

$$Acc_{M1} - \epsilon$$

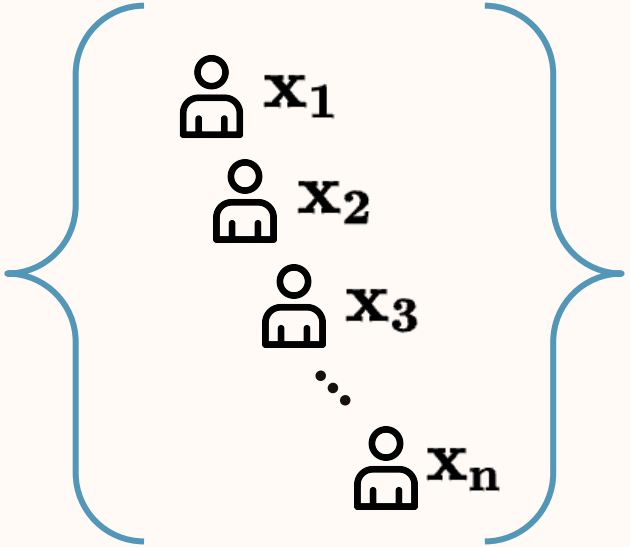
Multiplicity

$$\mathbf{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$$

$$\mathbf{x}_i = [1, x_{i1}, \cdots, x_{id}] \in \mathbb{R}^{d+1}$$

$$y_i \in \{0, 1\}$$

dataset



Rashomon Set

Model 1

$$p(y = +1|x)$$

$$Acc_{M1}$$

Model 2

$$p(y = +1|x)$$

$$Acc_{M1} - \epsilon$$

Model 3

$$p(y = +1|x)$$

$$Acc_{M1} - \epsilon$$

existence of **multiple good models** for a given problem

Rashomon Set

Model 1

$p(y = +1|x)$

Acc_{M1}

Model 2

$p(y = +1|x)$

$Acc_{M1} - \epsilon$

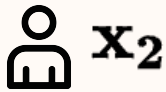
Model 3

$p(y = +1|x)$

$Acc_{M1} - \epsilon$

⋮

Predictive Multiplicity



$p_{M1}(\mathbf{x_2}) = 87\%$

$p_{M2}(\mathbf{x_2}) = 62\%$

$p_{M3}(\mathbf{x_2}) = 73\%$

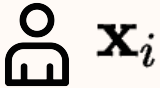
⋱

prediction variation over the set of good models

prediction variation over the set of good models is related to **general prediction uncertainty**

Predictive Multiplicity

How do individual predictions change over the set of good models?



Rashomon Set

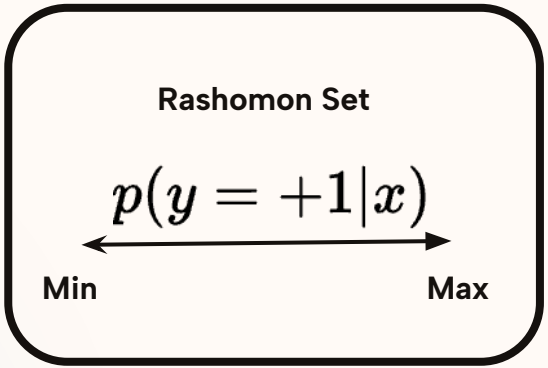
$$p(y = +1|x)$$

Min **Max**

Predictive Multiplicity

How do <u>individual predictions</u> change over the set of good models?

$\mathbf{x}_i$

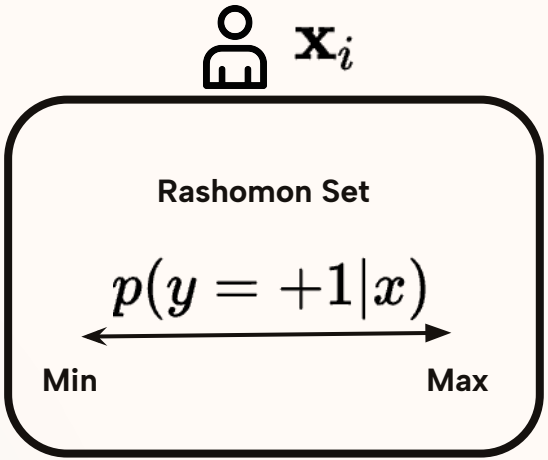


Defined based on setting e.g.

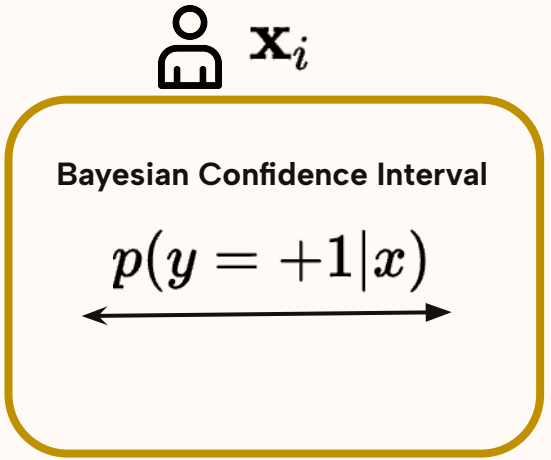
- Binary classification (yes/no flip)
- Probabilistic classification (range)
- Allocation (receive resource change)

Multiplicity and Uncertainty

How do individual predictions change over the set of good models?



How uncertain or consistent are model predictions?

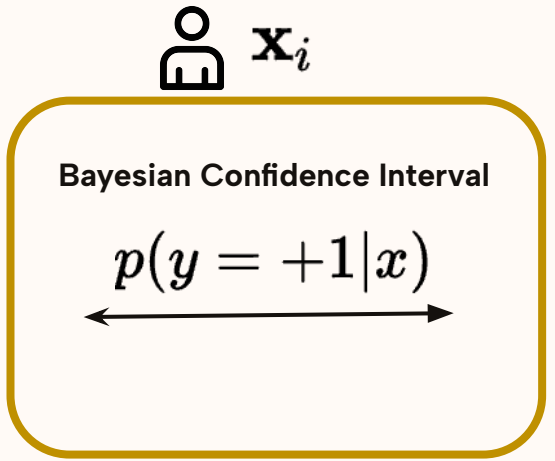


distinct phenomena though related

Multiplicity and Uncertainty

- Bayesian ML methods can be difficult to scale in applied settings
- Multiplicity is typically defined in a **non-bayesian regime** for practicality
- When compared, for the same problem setup, the Bayesian confidence interval is **not equivalent** to the Viable Prediction Range

How uncertain or consistent are model predictions?

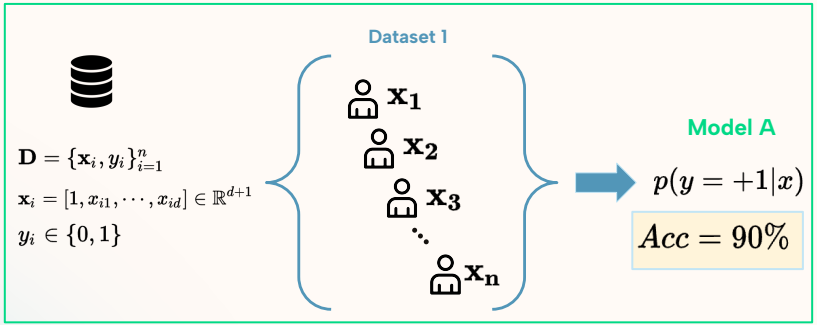


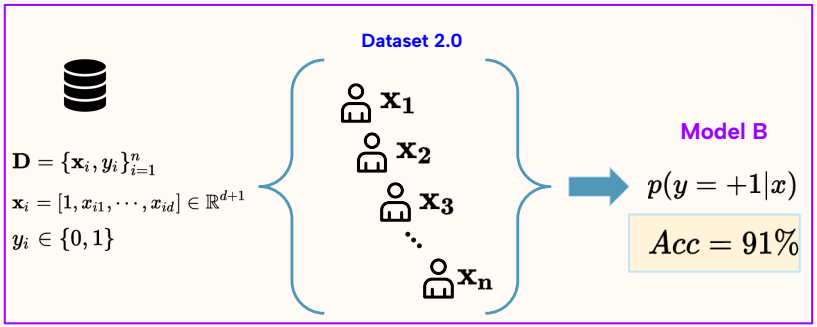
Watson-Daniels, J., Parkes, D. C., & Ustun, B. (2023, June). Predictive multiplicity in probabilistic classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 9, pp. 10306-10314).

How about other notions of model predictive instability?

Predictive Churn

Industry interest in **how model predictions change over time** with model updates **i.e. training dataset changes**

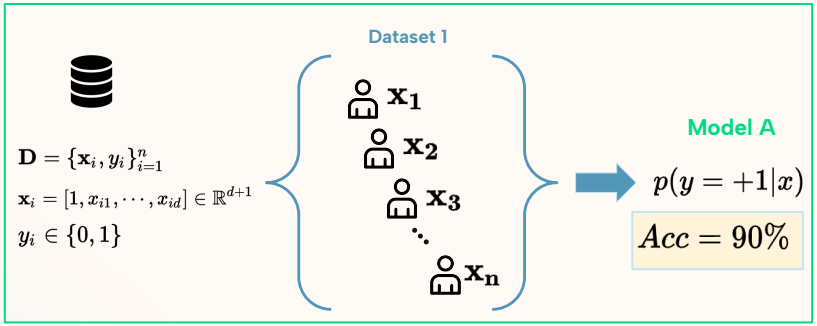


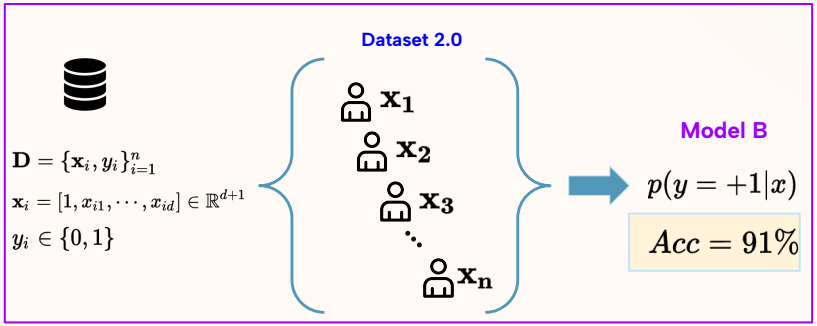


Watson-Daniels, J., du Pin Calmon, F., D'Amour, A., Long, C., Parkes, D. C., & Ustun, B. (2024, February). Predictive churn with the set of good models. In arXiv preprint arXiv:2402.07745.

Predictive Churn

Industry interest in **how model predictions change over time** with model updates **i.e. training dataset changes**





Best case (1%): Model B improves Model A

Worst case (19%): Model A is accurate on 9% B gets wrong, Model B correct on 10% A gets wrong

Watson-Daniels, J., du Pin Calmon, F., D'Amour, A., Long, C., Parkes, D. C., & Ustun, B. (2024, February). Predictive churn with the set of good models. In arXiv preprint arXiv:2402.07745.

Predictive Churn + Predictive Multiplicity

Industry interest in **how model predictions change over time** with model updates **i.e. training dataset changes**

What is the relationship between

(1) parts of the dataset that have **lots of prediction change over Rashomon set**

and

(2) parts of the dataset that have **lots of predictions change over time**?

Watson-Daniels, J., du Pin Calmon, F., D'Amour, A., Long, C., Parkes, D. C., & Ustun, B. (2024, February). Predictive churn with the set of good models. In arXiv preprint arXiv:2402.07745.

Predictive Churn + Predictive Multiplicity

Industry interest in **how model predictions change over time** with model updates **i.e. training dataset changes**

➔ Samples impacted by predictive multiplicity i.e. ambiguous examples are often the same samples impacted by predictive churn i.e. dataset changes

➔ Mitigating predictive multiplicity e.g. ensembling algorithm also decreases predictive churn

Predictive multiplicity analysis can help provide insight into **general predictive instability**

Watson-Daniels, J., du Pin Calmon, F., D'Amour, A., Long, C., Parkes, D. C., & Ustun, B. (2024, February). Predictive churn with the set of good models. In arXiv preprint arXiv:2402.07745.

As a concept, **predictive multiplicity** in a problem or in a given dataset is often correlated with **predictive instability**

Multiplicity and Uncertainty

"Uncertainty is better suited for modelling the information-theoretic relationships."

"Multiplicity is aligned with learning theory and hierarchical optimization, and is better suited for exploring alternative interpretations."

Ganesh, P., Taik, A., & Farnadi, G. (2025, October). Systemizing multiplicity: The curious case of arbitrariness in machine learning. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 2, pp. 1032-1048).

Multiplicity and Uncertainty

"Separability informs predictive multiplicity"

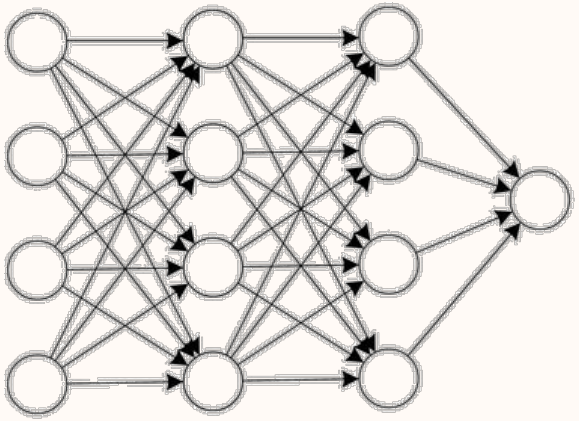
– Watson-Daniels et al. (2023)

"Noisier datasets lead to larger Rashomon ratios"

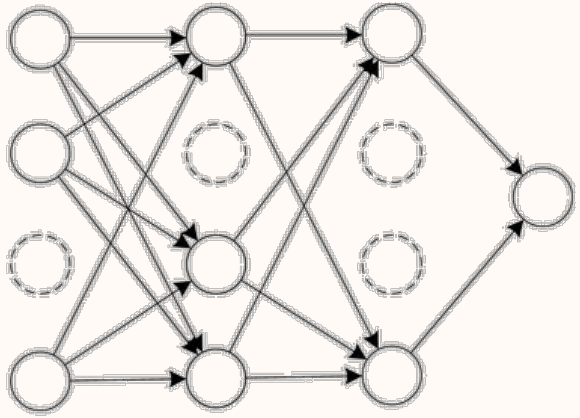
– Semenova et al. (2023)

Watson-Daniels, J., Parkes, D. C., & Ustun, B. (2023, June). Predictive multiplicity in probabilistic classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 9, pp. 10306-10314).
Semenova, L., Chen, H., Parr, R., & Rudin, C. (2023). A path to simpler models starts with noise. Advances in neural information processing systems, 36, 3362-3401.

Multiplicity and Uncertainty



(a) Standard Neural Network



(b) Network after Dropout

Inference time dropout, combined with filtering models in the Rashomon set, can help estimate multiplicity more efficiently!

Hsu, H., Li, G., Hu, S., & Chen, C. F. Dropout-Based Rashomon Set Exploration for Efficient Predictive Multiplicity Estimation. In The Twelfth International Conference on Learning Representations.

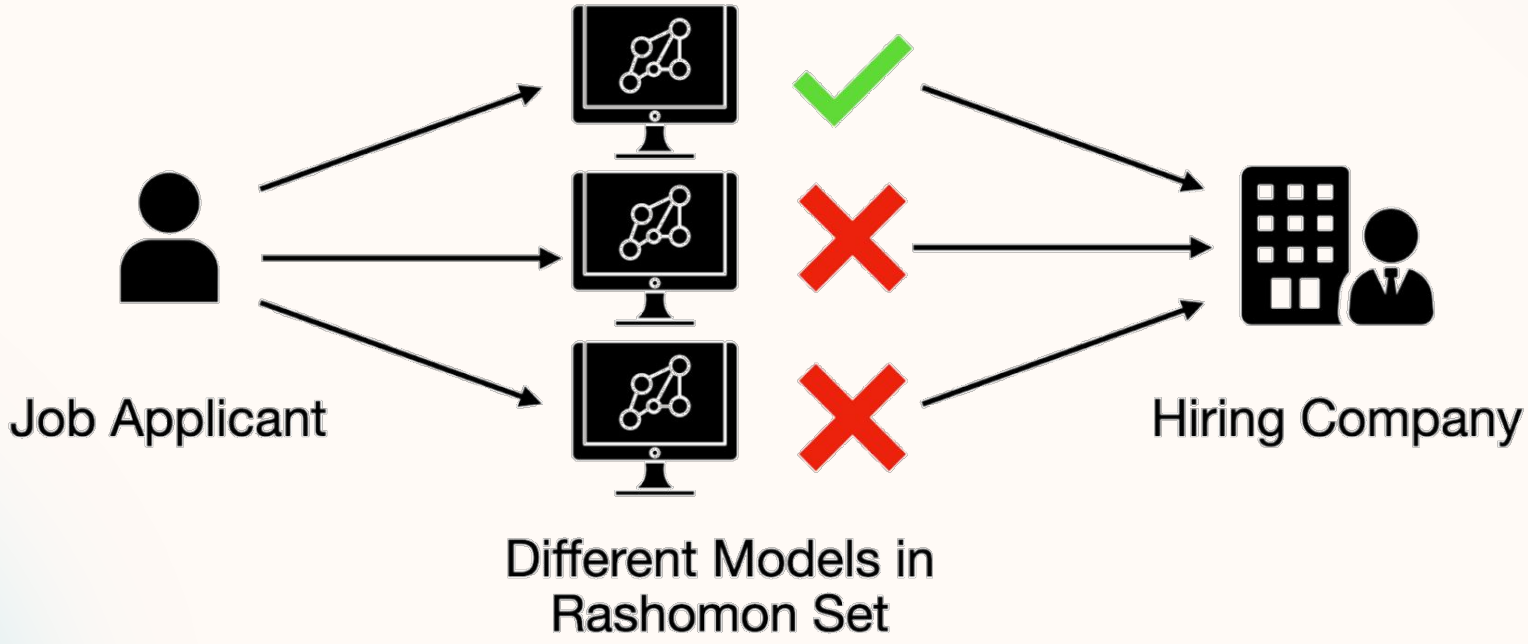
Open Areas of Research

- Develop additional computationally **efficient methods** for training the Rashomon Set → assist in uncertainty analysis
- Empirical – investigate whether methods to **reduce churn and uncertainty** can also address multiplicity concerns e.g. fairness
- Theoretical – **distinguish** between multiplicity and uncertainty more rigorously
- Algorithmic – connect multiplicity to established work in **algorithmic stability**
- And more!

Implications of Multiplicity for Fairness

Conflicting Outcomes

A job applicant may be accepted or rejected depending on the model chosen



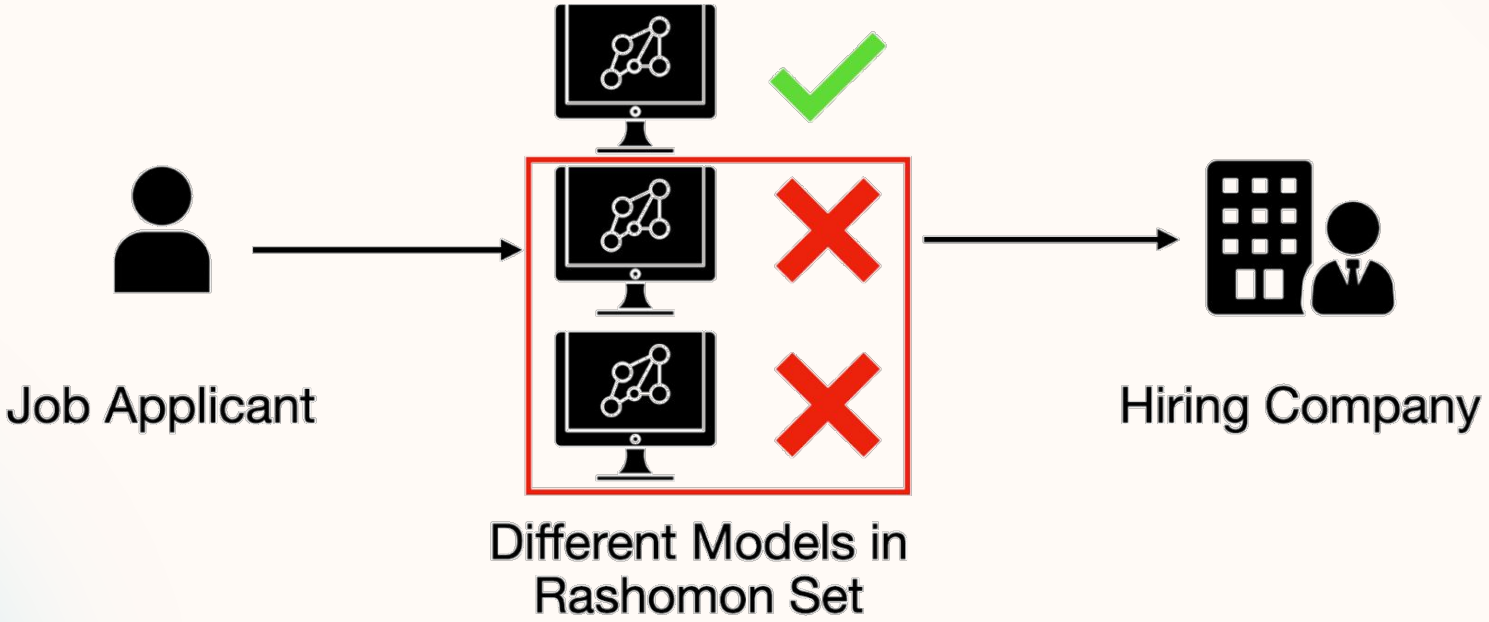
Implications of Multiplicity for Fairness

Different fairness concerns => Different multiplicity interventions

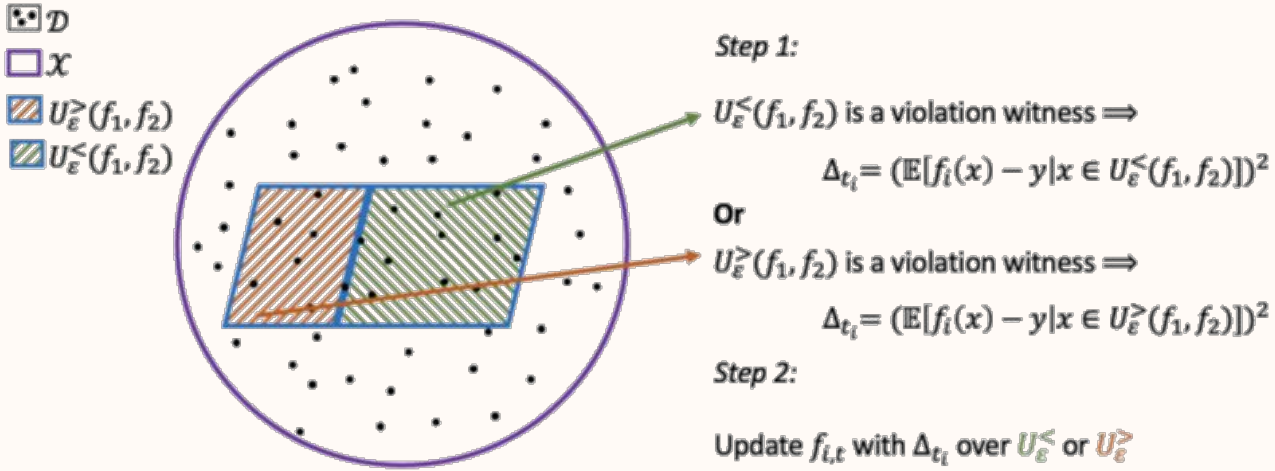
Fairness Concern	Multiplicity Intervention
Conflicting Outcomes	Combining Models
Outcome Homogenization	Randomizing Among Models
Group Fairness	Secondary Criteria to Choose Models

Ensemble of Rashomon Models

Is the ensemble of Rashomon models more fair than using a single model?



Reconciliation of Rashomon Models



Source: Behzad et al., 2025

Roth et al. “Reconciling Individual Probability Forecasts” (FAccT 2023)

Behzad et al. “Reconciling Predictive Multiplicity in Practice” (FAccT 2025)

Algorithmic Monoculture

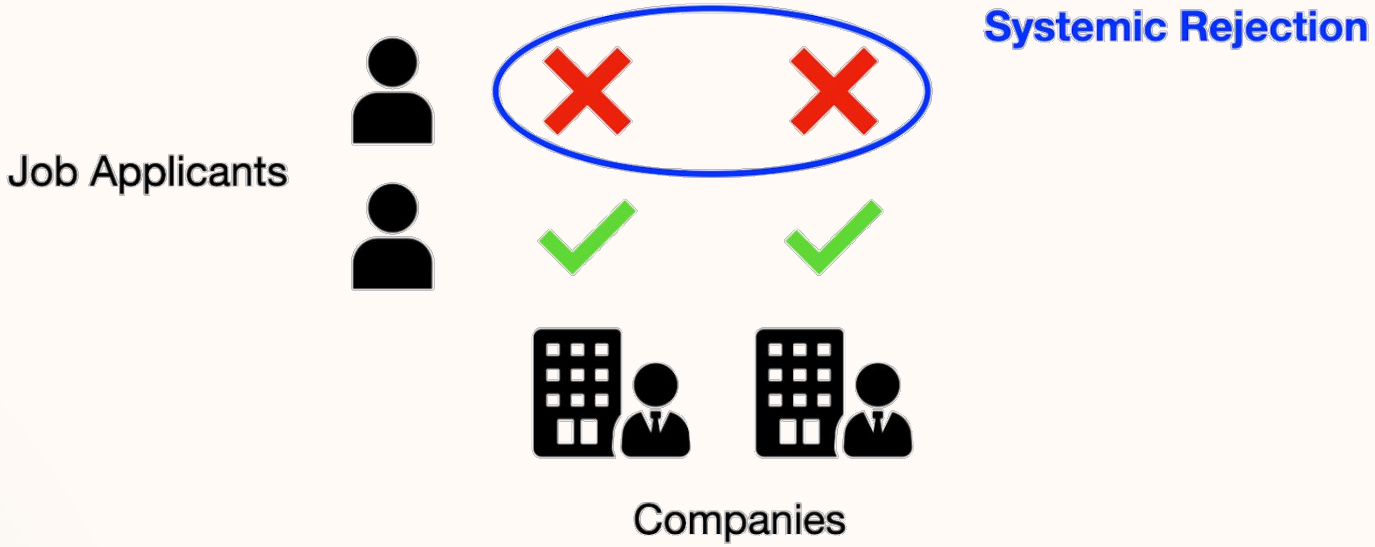
Decision-makers using the same or similar algorithm

Kleinberg and Raghavan. “Algorithmic Monoculture and Social Welfare” (PNAS 2021)

Bommasani et al. “Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?” (NeurIPS 2022)

Outcome Homogenization

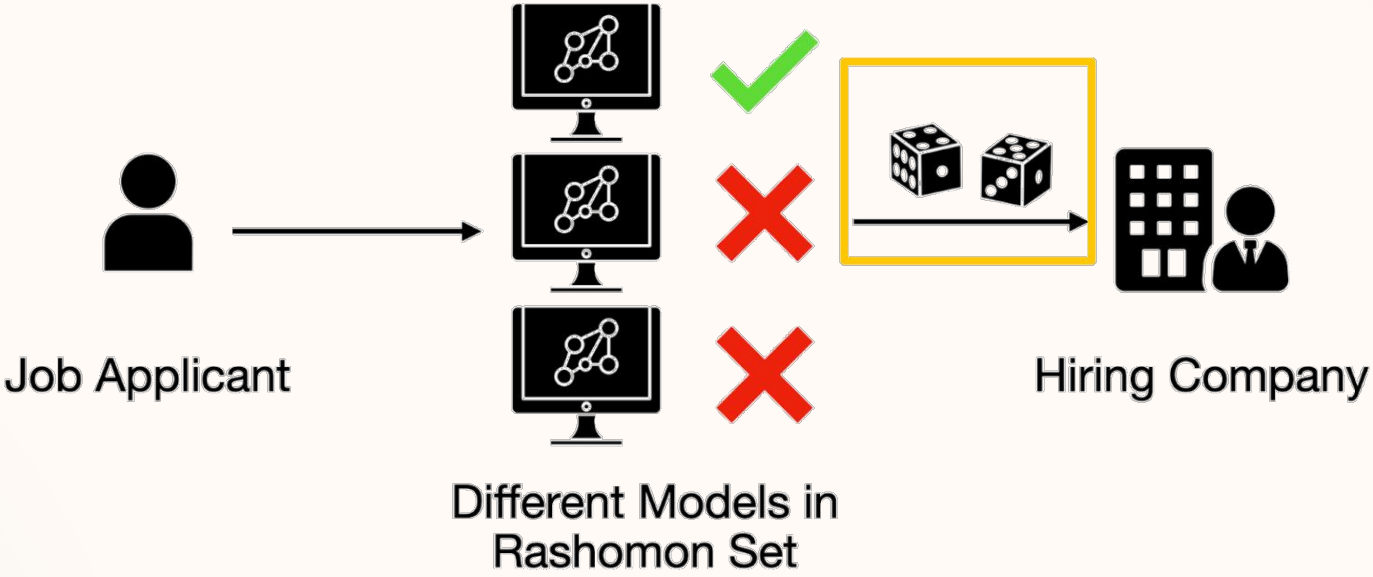
When individuals receive negative outcomes from all decision-makers



Bommasani et al. “Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?” (NeurIPS 2022)

Randomly Choosing Among Rashomon Models

Reducing Outcome Homogenization Without Compromising Accuracy



Jain et al. “Scarce Resource Allocations That Rely On Machine Learning Should Be Randomized” (ICML 2024)

How should decision-makers address multiplicity?

Domain-Specific Considerations

- **Factual vs Normative Decisions:** Whether there exists a single ground truth for the target variable
- **Single-Shot vs Multi-Shot Settings**: Whether individuals may receive the resource or opportunity from another decision-maker at a later time
- **Low vs High Aleatoric Uncertainty**: Whether there is variability in the target variable for the same inputs
- **Public vs Private Decisions**: Whether decision-makers are public entities or private entities

Gur-Arieh, S., & Lee, C. (2025, June). Consistently Arbitrary or Arbitrarily Consistent: Navigating the Tensions Between Homogenization and Multiplicity in Algorithmic Decision-Making. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency.

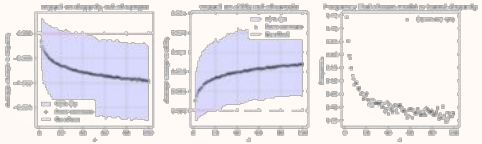
Choosing a Less Discriminatory Algorithm

Black et al. "The Legal Duty to Search For Less Discriminatory Algorithms" (FAccT 2024)

Choosing a Less Discriminatory Algorithm

Statistical Angle

Challenge: Firms commit to their policy using limited information. Finding an LDA in-sample in retrospect is much easier than finding generalizable, fair models.



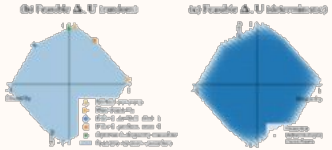
However, it is possible to reduce disparities out-of-sample using ML.

Math Angle

Challenge: It is sometimes impossible to reduce disparity at given accuracy.

Example:Feasible polygon for the realized dataset

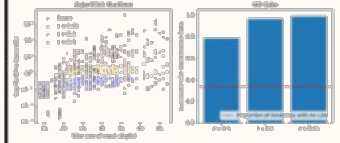
	+	-
1	15	20
2	5	10



However, this limit is only binding at near-perfect accuracy, unrealistic classifiers.

Comp. Angle

Challenge: Finding whether an LDA exists, given a fixed, fully known distribution, is NP-hard in general.



However, it is (1-ε)-approximable. (a full polynomial-time approximation exists).

Welfare Angle

Challenge: An LDA could leave consumers strictly worse off.

Example:
The distribution specified right, with + outcome probabilities ϱ and support σ. Utility of firm is TPR - 0.5*FPR. U of consumer is TPR-2*FPR. Then h' is an LDA leaving consumers worse off.

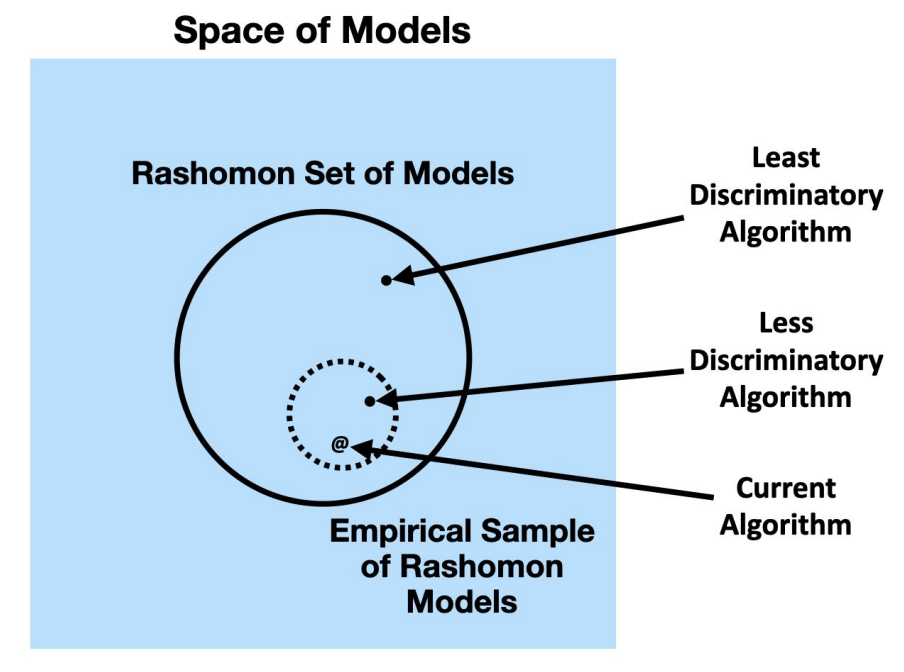
$\mathcal{X}$	$\mathcal{G}$	ϱ	σ	$h^{\mathcal{D}}$	h'
1	1	0.0	0.0	0	0
	2	0.7	0.0	0	0
2	1	0.8	0.5	1	0
	2	0.1	0.5	1	0
3	1	0.1	0.7	1	1
	2	0.1	0.7	1	1
4	1	0.1	1.0	1	1
	2	0.1	1.0	1	1

However, accounting for consumer welfare is easy.

Source: Laufer et al. (2025)

Laufer, B., Raghavan, M., & Barocas, S. (2025, March). What Constitutes a Less Discriminatory Algorithm?. In Proceedings of the 2025 Symposium on Computer Science and Law (pp. 136-151).

Choosing a Less Discriminatory Algorithm



Source: Jain et al. (2025)

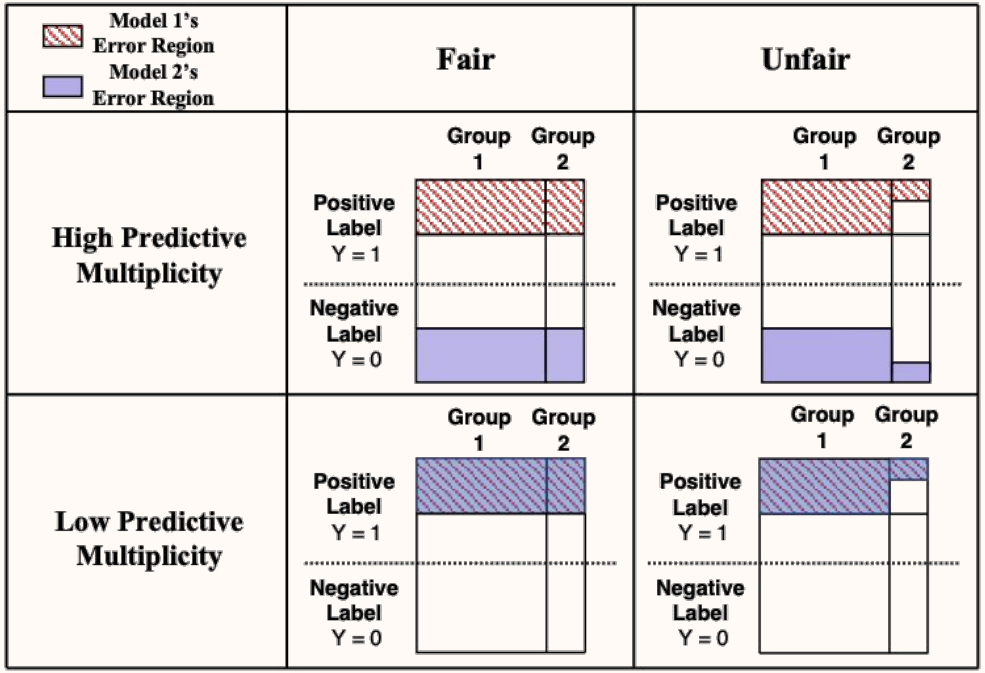
Jain, S., Wang, M., Creel, K., & Wilson, A. (2025, June). Allocation Multiplicity: Evaluating the Promises of the Rashomon Set. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 2040-2055).

Implications of Multiplicity for Fairness

Different fairness concerns => Different multiplicity interventions

Fairness Concern	Multiplicity Intervention
Conflicting Outcomes	Combining Models
Outcome Homogenization	Randomizing Among Models
Group Fairness	Secondary Criteria to Choose Models

Implications of Fairness for Multiplicity

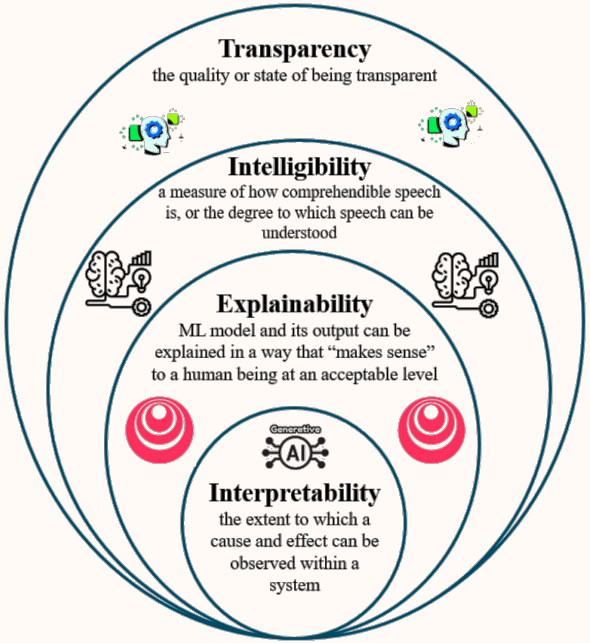


Source: Long et al. (2023)

Long, C., Hsu, H., Alghamdi, W., & Calmon, F. (2023). Individual arbitrariness and group fairness. Advances in Neural Information Processing Systems, 36, 68602-68624.

Implications of Multiplicity for Explainability

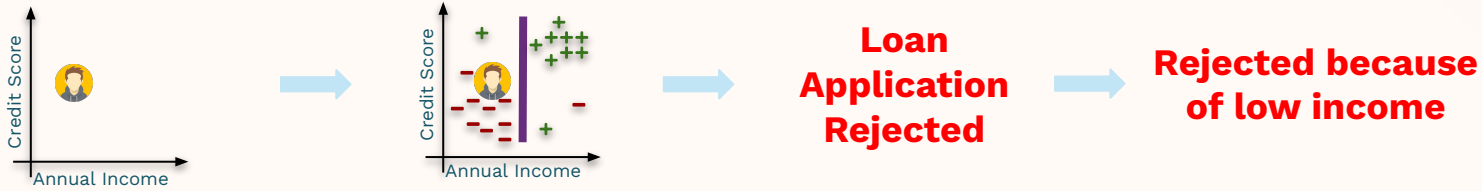
What is an Explanation?



Source: Shafik et al. (2024)

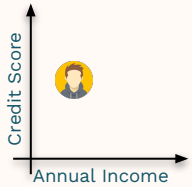
Shafik, W., Hidayatullah, A.F., Kalinaki, K., Gul, H., Zakari, R.Y. and Tufail, A., 2024. A Systematic Literature Review on Transparency and Interpretability of AI models in Healthcare: Taxonomies, Tools, Techniques, Datasets, OpenResearch Challenges, and Future Trends.

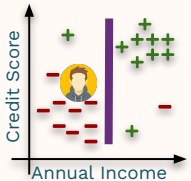
What is an Explanation?



Source: All icons from Flaticon (https://www.flaticon.com/)

Explanation Multiplicity



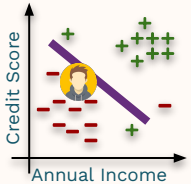


Decision

Loan Application Rejected

Explanation

Rejected because of low income



Decision

Loan Application Rejected

Explanation

Rejected because of low income and credit score

Source: All icons from Flaticon (https://www.flaticon.com/)

Implications of Multiplicity for Explainability

~~Fairness~~ Explainability Concern	Multiplicity Intervention
Conflicting ~~Outcomes~~ Explanations	Combining Models
~~Outcome Homogenization~~	~~Randomizing Among Models~~
~~Group Fairness~~ Interpretability	Secondary Criteria to Choose Models

Explanation Multiplicity

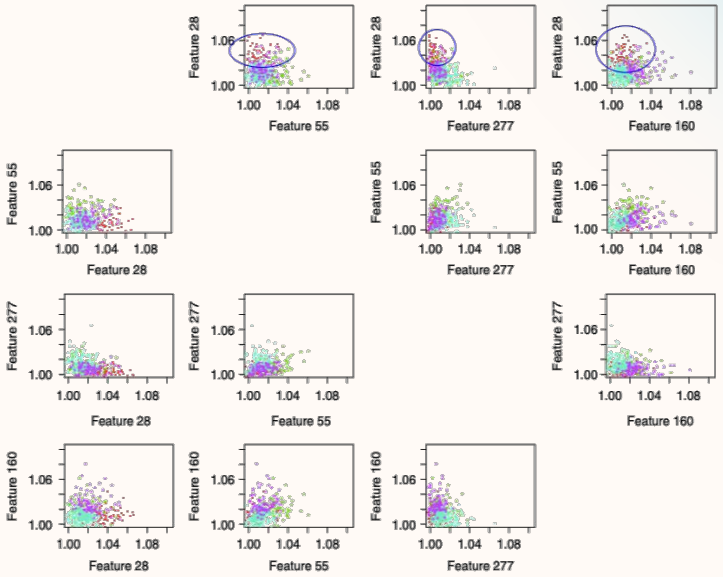
Attribution-based Explanations

Recourse/Counterfactuals

Explanation Multiplicity

Attribution-based Explanations

Recourse/Counterfactuals



Variable Importance Clouds

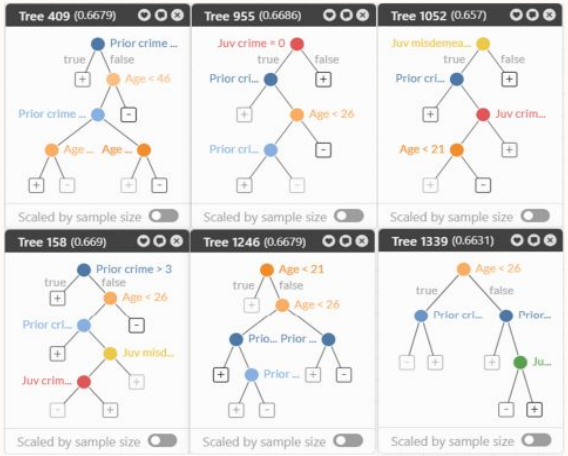
Source: Dong et al., 2020

Dong, J., & Rudin, C. (2020). Exploring the cloud of variable importance for the set of all good models. Nature Machine Intelligence, 2(12), 810-824.

Explanation Multiplicity

Attribution-based Explanations

Recourse/Counterfactuals



Visualizing the Rashomon Set

Source: Rudin et al., 2024

Rudin, C., Zhong, C., Semenova, L., Seltzer, M., Parr, R., Liu, J., ... & Boner, Z. (2024, July). Position: amazing things come from having many good models. In Proceedings of the 41st International Conference on Machine Learning (pp. 42783-42795).

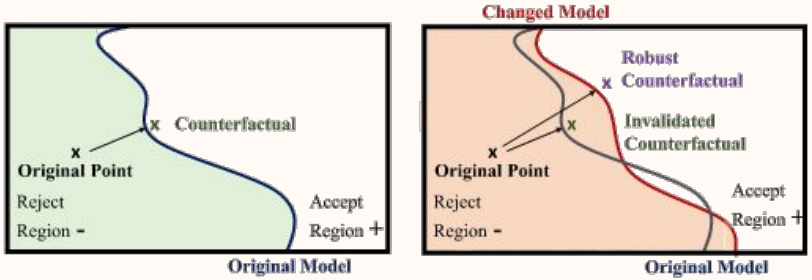
Xin, R., Zhong, C., Chen, Z., Takagi, T., Seltzer, M., & Rudin, C. (2022). Exploring the whole rashomon set of sparse decision trees. Advances in neural information processing systems, 35, 14071-14084.

Wang, Z. J., Zhong, C., Xin, R., Takagi, T., Chen, Z., Chau, D. H., ... & Seltzer, M. (2022, October). Timbertrek: Exploring and curating sparse decision trees with interactive visualization. In 2022 IEEE Visualization and Visual Analytics (VIS) (pp. 60-64). IEEE.

Explanation Multiplicity

Attribution-based Explanations

Recourse/Counterfactuals



Source: Hamman et al., 2023

Hamman, F., Noorani, E., Mishra, S., Magazzeni, D., & Dutta, S. (2023, July). Robust counterfactual explanations for neural networks with probabilistic guarantees. In International Conference on Machine Learning (pp. 12351-12367). PMLR.

Explanation Multiplicity

Attribution-based Explanations

Recourse/Counterfactuals

ROBUSTNESS AGAINST MC.
Assume an input x and a model M. Let x' be a CE for x. Robustness against MC requires that whenever the model M **changes** to M', and this change is **sufficiently small**, then $M(x') = M'(x')$.

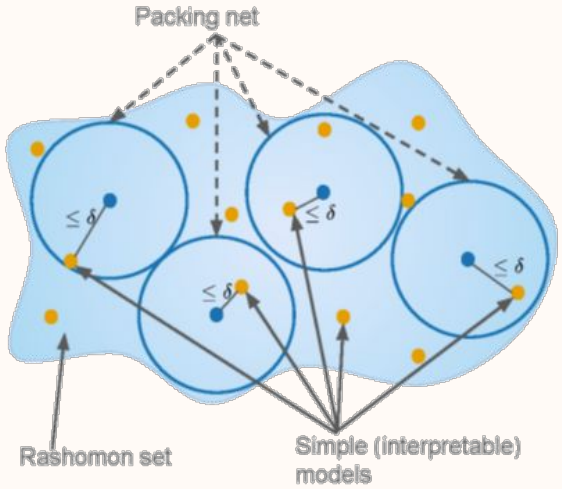
Source: Jiang et al., 2024

Jiang, J., Leofante, F., Rago, A., & Toni, F. (2024, August). Robust counterfactual explanations in machine learning: a survey. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (pp. 8086-8094).

Simpler Models and Multiplicity

Larger Rashomon ratios lead to better chances of finding simpler models

$$\text{Rashomon Ratio} := \frac{\text{Volume of Rashomon Set}}{\text{Volume of Hypothesis Space}}$$

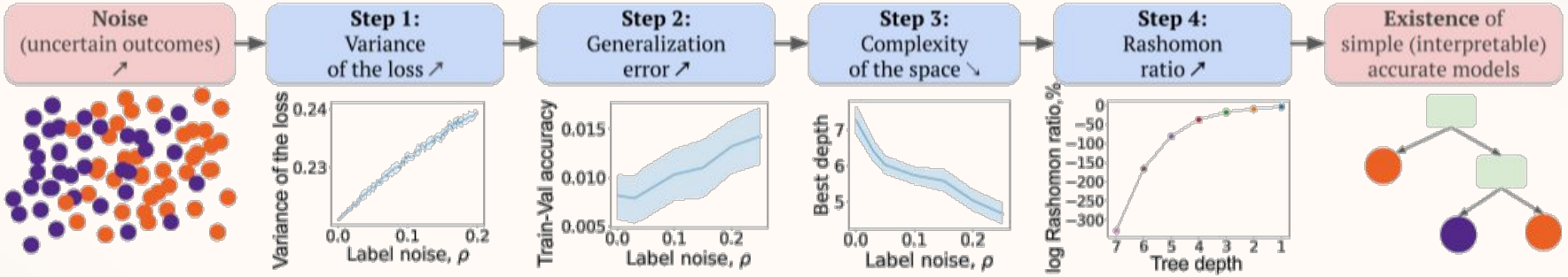


Source: Rudin et al., 2024

Rudin, C., Zhong, C., Semenova, L., Seltzer, M., Parr, R., Liu, J., ... & Boner, Z. (2024, July). Position: amazing things come from having many good models. In Proceedings of the 41st International Conference on Machine Learning (pp. 42783-42795).
Semenova, L., Rudin, C., & Parr, R. (2022, June). On the existence of simpler machine learning models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1827-1858).

Simpler Models and Multiplicity

Noisier settings lead to larger Rashomon ratios



Source: Rudin et al., 2024

Rudin, C., Zhong, C., Semenova, L., Seltzer, M., Parr, R., Liu, J., ... & Boner, Z. (2024, July). Position: amazing things come from having many good models. In Proceedings of the 41st International Conference on Machine Learning (pp. 42783-42795).

Semenova, L., Chen, H., Parr, R., & Rudin, C. (2023). A path to simpler models starts with noise. Advances in neural information processing systems, 36, 3362-3401.

Multiplicity in the Age of Generative AI

Homogenization vs Multiplicity

Multiplicity (loosely defined):
Varied model outputs across similar models or similar prompts

Open Area of Research:

How to rigorously define multiplicity for generative AI

Homogenization vs Multiplicity

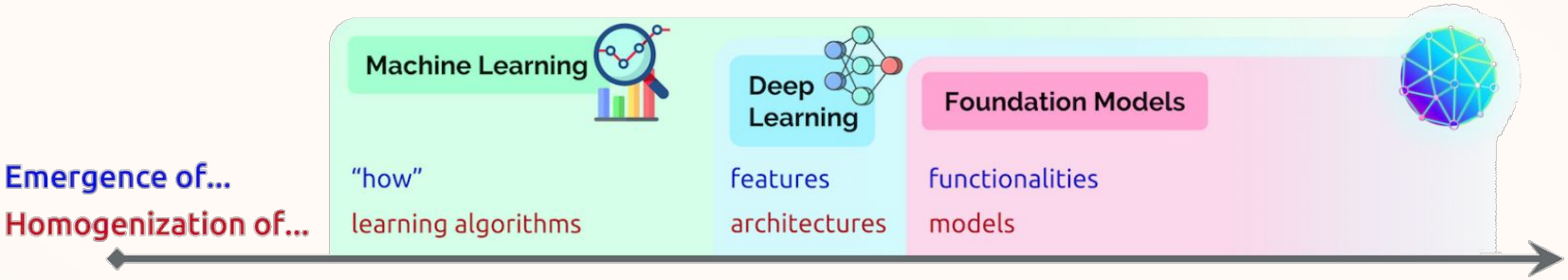
Homogenization:
Similar model outputs across different models or different prompts

LLMs often produce homogeneous outputs, but the impact depends on the task

- **Creative tasks:**
  - Example: If a user asks for a joke and the model always gives a “knock-knock” joke, output **homogenization limits creativity and user engagement**
- **Objective tasks:**
  - Example: For math problems, consistent correct answers are desirable.
  - However, **variation in explanations or problem-solving strategies can enhance understanding**.

(1) Jain, S., Wang, M., Creel, K., & Wilson, A. (2025). Allocation Multiplicity: Evaluating the Promises of the Rashomon Set. FAccT 2025
(2) Jain, S., Lanchantin, J., Nickel, M., Ullrich, K., Wilson, A., & Watson-Daniels, J. (2025). LLM Output Homogenization is Task Dependent. arXiv:2509.21267

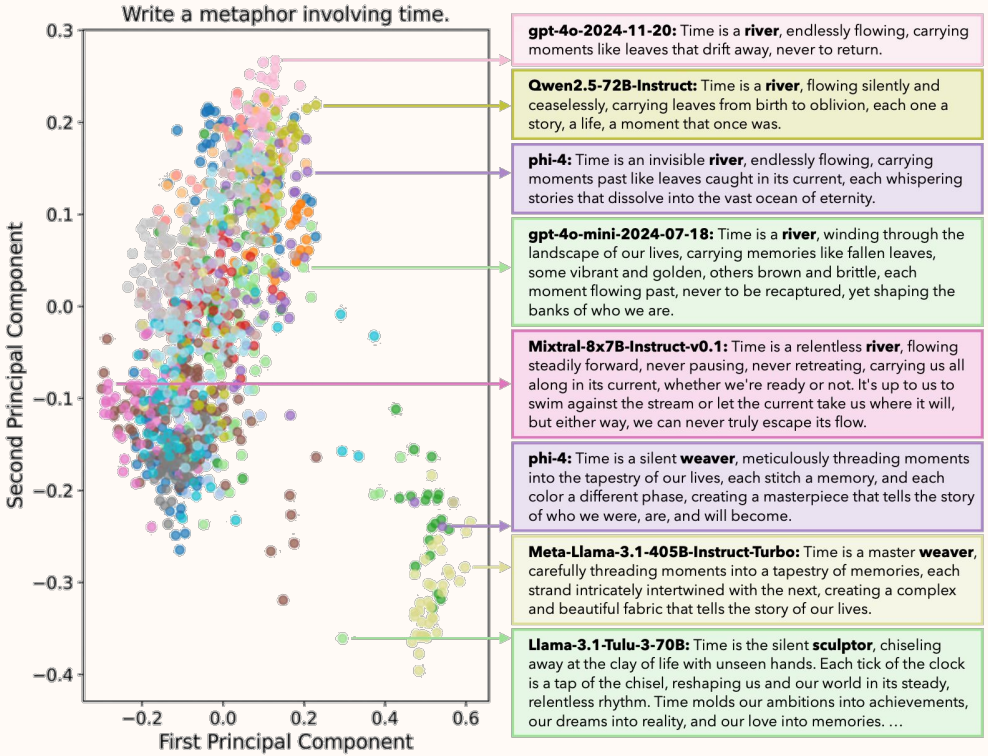
Homogenization and Monoculture



Source: Bommasani et al., 2021

Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Homogenization and Monoculture



Source: Jiang et al., 2025

Jiang, L., Chai, Y., Li, M., Liu, M., Fok, R., Dziri, N., … & Choi, Y. Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond). In The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track.

Homogenization vs Multiplicity

Homogenization:
Similar model outputs across different models or different prompts

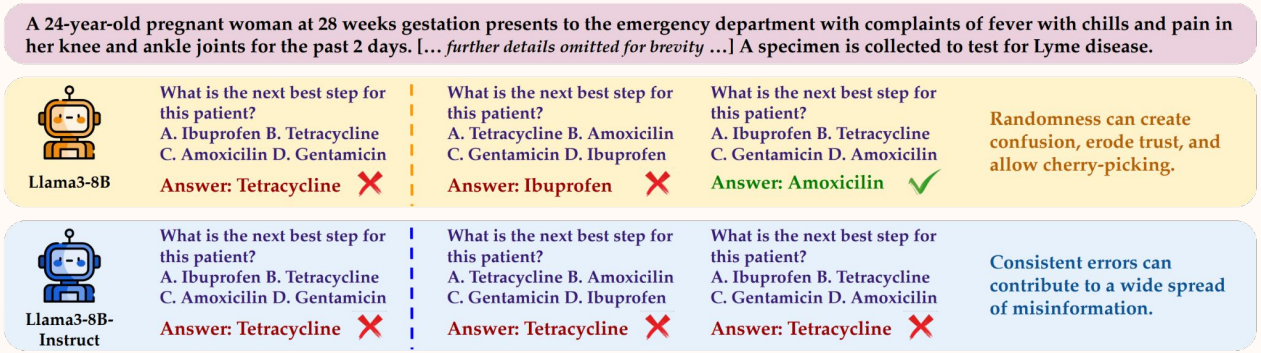
Multiplicity:
Varied model outputs across similar models or similar prompts

“Each narrative finds some empirical support, but neither tightly fits the observations. ”

Gorecki, M., & Hardt, M. (2025). Monoculture or Multiplicity: Which Is It? NeurIPS 2025 poster. https://openreview.net/pdf?id=DO5LtJc80w

Prompt Multiplicity

"Traditionally, machine learning workflows centered around the training of task-specific models. In contrast, we are now seeing [...] a growing focus on designing effective prompts to elicit desirable behavior. In this new paradigm, prompts have become the focus of system design."



Source: Ganesh et al., 2025

Ganesh, P. (2025, October). From Model Multiplicity to Prompt Multiplicity: Emerging Arbitrariness Concerns in the Age of Generative AI. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 3, pp. 2872-2874).
Ganesh, P., Shokri, R., & Farnadi, G. (2025). Rethinking hallucinations: Correctness, consistency, and prompt multiplicity. In ICLR 2025 Workshop on Building Trust in Language Models and Applications.

What is Multiplicity in GenAI Models?

Multiplicity through choices between discrete classes ⇒ Works well for traditional machine learning models.

But, this is not how we actually use GenAI models!

How to measure multiplicity moving from classification to generation?

Open Areas of Research

- **Compare** conceptualization of multiplicity in discriminative models (classification) to generative models
- Theoretical grounding for **multiplicity as a concept** in generative AI
- Algorithms to intervene and **balance homogenization vs multiplicity** considerations
- And more!

Discussion

Scenario 1: Hiring

A large company is planning to partially automate their entry-level hiring pipeline. They receive over 100 applications each week, and recruiters don't have time to review every application. They are planning to train a model that will select the top 25 applicants every week for manual review.

Scenario 1: Hiring

When developing their model, the company notices there is predictive multiplicity, so they search for 100 models in the Rashomon Set. All models satisfy the 80% rule in demographic parity.

Imagine you are an applicant for the company. How would you prefer they address multiplicity?

1. **Use an ensemble model that averages predictions**
2. **Randomize among individuals that would be in the top 25 applicants under any Rashomon model**
3. **Use the model with the best demographic parity on race & gender**

Slido: #2634641

Scenario 2: Hiring Product

A company is designing a hiring product as a service that will be used by many other businesses to partially automate their entry-level hiring pipeline. The company is planning to train a model that can select the top applicants from the pool of applicants. For each business that works with this company, they will take their pool of applicants and provide them with the top applicants based on their algorithm.

Scenario 2: Hiring Product

When developing their model, the company notices there is predictive multiplicity, so they search for 100 models in the Rashomon Set. All models satisfy the 80% rule in demographic parity.

Imagine you are an applicant for many businesses that all use this company. How would you prefer they address multiplicity?

1. **Use an ensemble model that averages predictions**
2. **Randomize among individuals that would be in the top 25 applicants under any Rashomon model**
3. **Use the model with the best demographic parity on race & gender**

Slido: #2634646

Scenario 3: Tax Audit

A government tax agency is planning to use an AI system to detect potential tax fraud. Everyone flagged by the AI will be audited, and audits will occur solely based on AI flags. However, taxpayers will only face penalties if a human investigator confirms the fraud. The tax agency checks that the model works well by seeing how often it agrees with human experts' choices of who to audit. The AI and the human experts agree 85% of the time.

Adapted from Meyer et al. "Perceptions of the Fairness Impacts of Multiplicity in Machine Learning" (CHI 2025)

Scenario 3: Tax Audit

When developing their model, the government notices there is predictive multiplicity, so they search for 100 models in the Rashomon Set. Imagine you are a taxpayer. How would you prefer they address multiplicity?

1. **Use an ensemble model that averages predictions, but only flag individuals with high risk**
2. **Randomize so the chance of being flagged depends on average risk score**
3. **Use the model with the most similar false positive rates across income levels**

Slido: #2634642

Scenario 4: Loan Application

A bank is planning to use an AI system to automate the process of small loan applications. All small loan applications will require filling a structured form, and the loan application decision will be made in an automated manner by an AI system. If the application is rejected, the bank also wants the system to provide recourse to the applicant.

Scenario 4: Loan Application

When developing their model, the bank notices explanation multiplicity. They search for 100 models in the Rashomon set and find that even when an application is rejected by them all, the suggested recourse differs between them. Imagine you are an applicant who was rejected. How would you prefer the bank addresses multiplicity?

1. **Use the model with the easiest recourse for the candidate**
2. **Use the model with the most 'robust' recourse. Same as finding recourse that flips the decisions for most models**
3. **Use the model with the most 'diverse' recourse options**

Slido: #2634643

Scenario 5: Recognizing Fault Points

A company that manufactures duplex steel is planning to use an AI system to study the causes of heating silver faults in their manufacturing pipeline. These faults can occur at various points in the process, and the company wants to use AI systems to recognize which parts of the pipeline need to be improved or replaced.

Adapted from Gianlucca, et al. "Predicting heating sliver in duplex stainless steels manufacturing through rashomon sets." IJCNN. IEEE, 2021.

Scenario 5: Recognizing Fault Points

When developing their model, the company notices multiplicity: different models recognize different parts of the pipeline to be faulty even though they have similar overall accuracy. Imagine you are responsible for maintenance. How would you address multiplicity?

1. **Use majority-voting to find the most likely failure point**
2. **Cluster models based on their recognized failure points, and then test each possibility in a random order**
3. **Only consider the failure points recognized by the model that had the best recall, i.e., recognized the most faults**

Slido: #2634644

Scenario 6: Diagnosis Support

A hospital is planning to use an AI system to help doctors with preliminary diagnosis. Medical records of patients are fed to the AI system, and a diagnosis along with an explanation is shown to the doctor. The doctor will study the explanation provided by the AI system to judge its reliability, and to guide their own diagnosis. The final diagnosis will be provided by the doctor.

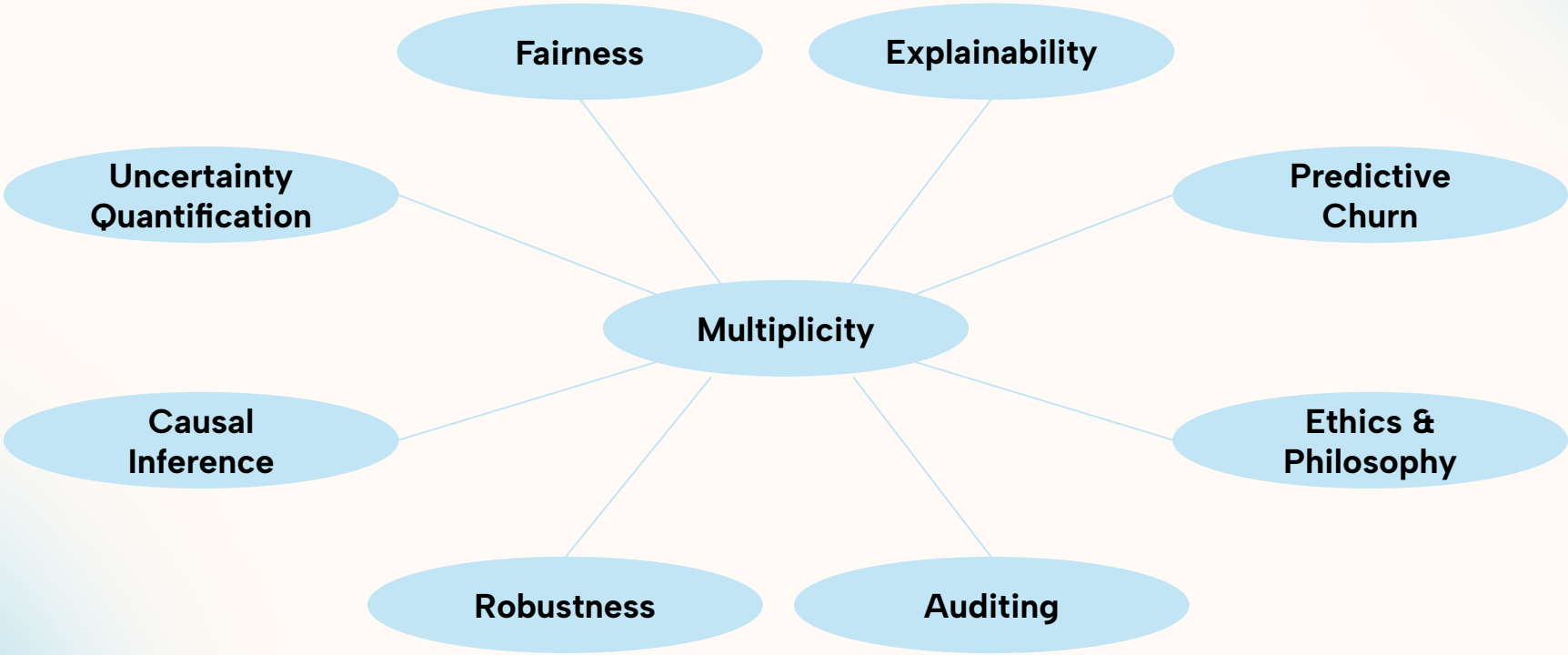
Scenario 6: Diagnosis Support

When developing their model, the hospital notices there is explanation multiplicity even when two models agree on their diagnosis. Imagine you are a doctor. How would you prefer the AI system developers address multiplicity?

1. **Aggregate explanations to find the most common explanation**
2. **Provide all explanations under a sample of Rashomon models**
3. **Only provide a diagnosis if majority of explanations are aligned, otherwise do not provide any diagnosis or explanations**

Slido: #2634645

The Many Faces of Multiplicity in ML



The Many Faces of Multiplicity in ML

There is a growing debate about the implications of multiplicity for algorithmic decision-making.

With this tutorial, we hope to have raised awareness of different perspectives on multiplicity and their connection to broader discussions in the community.


Prakhar Ganesh


Carol Long


Afaf Taik


Hsiang Hsu


Jamelle Watson-Daniels


Kathleen Creel


Flavio Calmon


Golnoosh Farnadi

Feedback Form
Connect with Us
Find Slides and Demo
Keep track of Future Events
Multiplicity Reading List