Position: the Stochastic Parrot in the Coal Mine
Model Collapse is a Threat to Low-Resource Communities

Devon Jarvis Richard Klein Benjamin Rosman Steven James Stefano Sarao Mannelli

ICML 2026

Poster Session B; Hall A

July 9, 2026

5:00 PM – 6:45 PM

Our Position

Model Collapse:

- amplifies the existing harms of generative models for low-resource communities.

Our Position

Model Collapse:

- amplifies the existing harms of generative models for low-resource communities.
  - Shifts data distributions away from marginalised groups towards hegemonic viewpoints.

Our Position

Model Collapse:

- amplifies the existing harms of generative models for low-resource communities.
  - Shifts data distributions away from marginalised groups towards hegemonic viewpoints.
  - Amplified uneven energy costs and lowers best case performance for under-represented groups.

Our Position

Model Collapse:

- amplifies the existing harms of generative models for low-resource communities.
  - Shifts data distributions away from marginalised groups towards hegemonic viewpoints.
  - Amplified uneven energy costs and lowers best case performance for under-represented groups.

Addressing model collapse should therefore be a central concern for research on low-resource settings and AI fairness

Stochastic Parrots [1]

- An analogy used to explain how LLMs are
  - often ungrounded.

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜 . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*(pp. 610-623).

Stochastic Parrots [1]

- An analogy used to explain how LLMs are
  - often ungrounded.
  - repeat probable words based on language stats.

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜 . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*(pp. 610-623).

Stochastic Parrots [1]

- An analogy used to explain how LLMs are
  - often ungrounded.
  - repeat probable words based on language stats.

- Leads to a number of potential harms, including:
  - Value Lock and Static Data.

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜 . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*(pp. 610-623).

Stochastic Parrots [1]

- An analogy used to explain how LLMs are
  - often ungrounded.
  - repeat probable words based on language stats.

- Leads to a number of potential harms, including:
  - Value Lock and Static Data.
  - Reflects online data distributions and not reality.

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜 . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*(pp. 610-623).

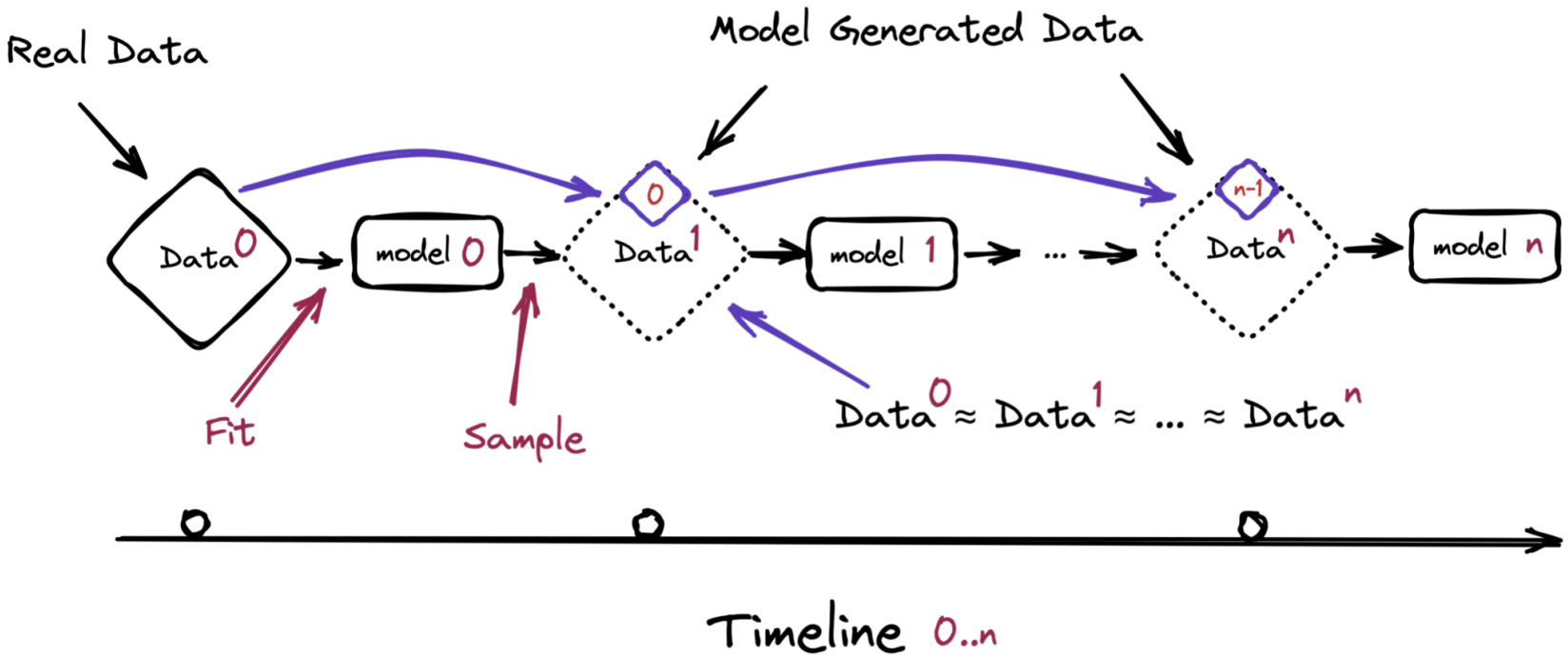
Stochastic Parrots [1]

- An analogy used to explain how LLMs are
  - often ungrounded.
  - repeat probable words based on language stats.

- Leads to a number of potential harms, including:
  - Value Lock and Static Data.
  - Reflects online data distributions and not reality.
  - Encodes explicit bias.

[1] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜 . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*(pp. 610-623).

Model Collapse [2,3,4]

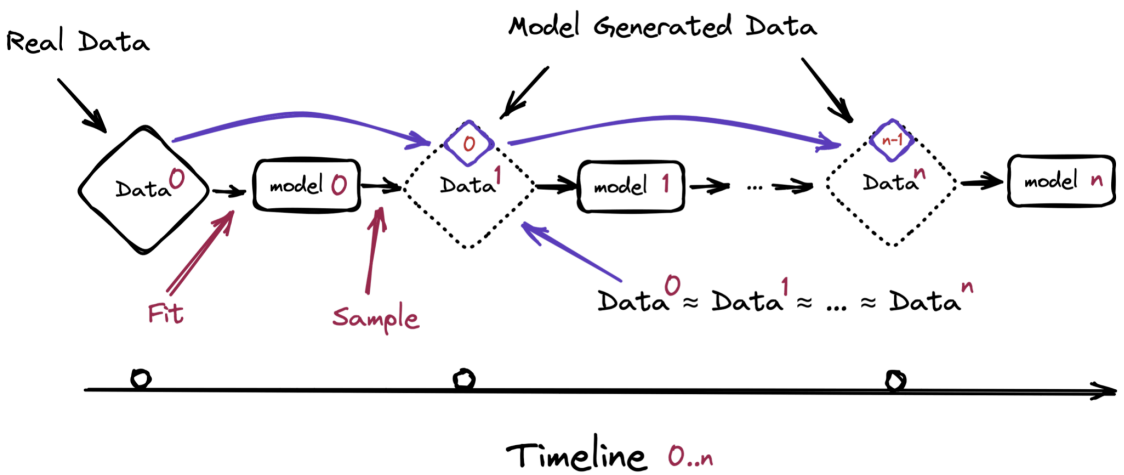


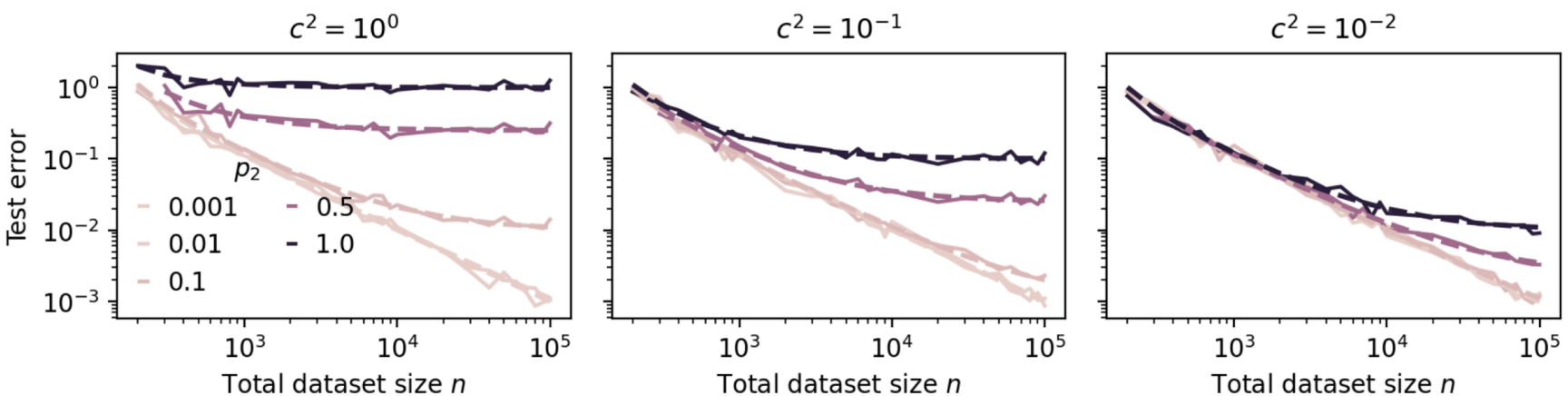
[2] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755-759.

[3] Dohmatob, Elvis, et al. "Strong model collapse." *International Conference on Learning Representations*. Vol. 2025. 2025.

[4] Shumailov, Ilia, et al. "The curse of recursion: Training on generated data makes models forget." *arXiv preprint arXiv:2305.17493* (2023).

Model Collapse [2,3,4]



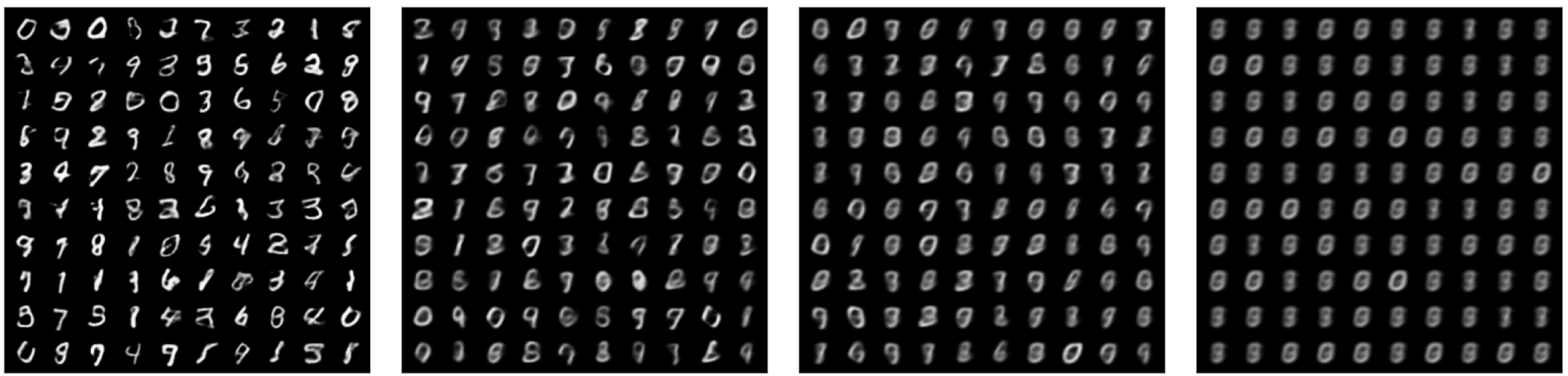


[2] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755-759.

[3] Dohmatob, Elvis, et al. "Strong model collapse." *International Conference on Learning Representations*. Vol. 2025. 2025.

[4] Shumailov, Ilia, et al. "The curse of recursion: Training on generated data makes models forget." *arXiv preprint arXiv:2305.17493* (2023).

Model Collapse [2,3,4]



(a) Original model (b) Generation 5 (c) Generation 10 (d) Generation 20

[2] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755-759.

[3] Dohmatob, Elvis, et al. "Strong model collapse." *International Conference on Learning Representations*. Vol. 2025. 2025.

[4] Shumailov, Ilia, et al. "The curse of recursion: Training on generated data makes models forget." *arXiv preprint arXiv:2305.17493* (2023).

Model Collapse [4]

Example of text outputs of an OPT-125m model affected by *Model Collapse*– models degrade over generations, where each new generation is trained on data produced by the previous generation.

Input: some started before 1360 — was typically accomplished by a master mason and a small team of itinerant masons, supplemented by local parish labourers, according to Poyntz Wright. But other authors reject this model, suggesting instead that leading architects designed the parish church towers based on early examples of Perpendicular

Outputs:

Gen 0: Revival architecture such as St. John's Cathedral in London. The earliest surviving example of Perpendicular Revival architecture is found in the 18th @-@ century Church of Our Lady of Guernsey, which dates from the late 19th century. There are two types of perpendicular churches : those

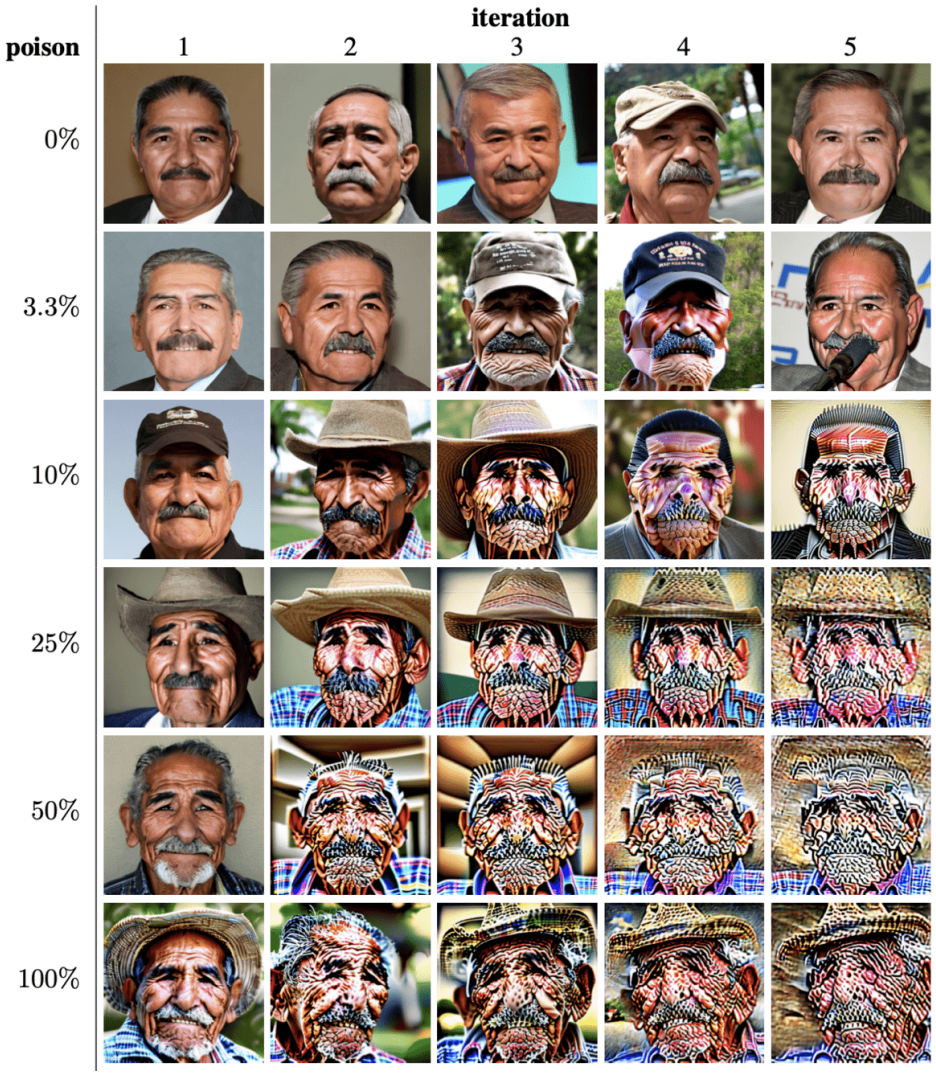
Gen 1: architecture such as St. Peter's Basilica in Rome or St. Peter's Basilica in Buenos Aires. There is no evidence that any of these buildings were built during the reign of Pope Innocent III, but it is possible that they may have been built during the reign of his successor, Pope Innocent

Gen 7: architecture in England. In an interview with The New York Times, Wright said : " I don 't think there is anything wrong with me being able to do what I want to do. It just doesn 't work for me. " He added : " I don 't know if you can call it funny,

Gen 9: architecture. In addition to being home to some of the world's largest populations of black @-@ tailed jackrabbits, white @-@ tailed jackrabbits, blue @-@ tailed jackrabbits, red @-@ tailed jackrabbits, yellow @-

[4] Shumailov, Ilia, et al. "The curse of recursion: Training on generated data makes models forget." *arXiv preprint arXiv:2305.17493* (2023)

Model Collapse [5]



[5] Bohacek, Maty, and Hany Farid. "Nepotistically trained generative image models collapse." *ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models*, 2023

Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
Synthetic-data feedback loops (Whitney & Norman, 2024)	Small real-data distributions are more easily overwhelmed by generated content.	Dominant patterns are reinforced across generations, while rare linguistic and cultural forms are pushed further into the tail.	Representation loss
Tail erasure under model collapse (Wyllie et al., 2024)	Low-resource languages, dialects, and marginalized viewpoints often already occupy low-probability regions of the data distribution.	Loss of diversity, reduced expressivity, and weaker model performance on communities whose data is already scarce.	Cultural erosion
Changed scaling behaviour with synthetic data (Shumailov et al., 2024)	If more data yields less improvement, communities with limited access to real data and compute have fewer viable routes to improvement.	Higher data and compute requirements, widening the gap between high- and low-resource settings.	Unequal access
High cost of real data acquisition (Kandpal & Raffel, 2025)	Collecting, validating, and maintaining culturally representative data is expensive and often unpaid or under-resourced.	Low-resource communities may be pushed toward synthetic supplementation, increasing the risk of poisoning their own data.	Data inequity
Automated curation and synthetic-data detection (Boutadjine et al., 2025)	Synthetic detection and filtering systems may be unreliable, opaque, or biased against uncommon language use.	Authentic but rare or reclaimed language may be removed, while synthetic or hegemonic content remains.	Curation bias
Value-lock and hegemonic viewpoints (Bender et al., 2021)	LLMs tend to reproduce frequent historical patterns rather than socially evolving or locally grounded meanings.	Outdated or dominant viewpoints persist, while new terms, norms, and movements struggle to enter the model distribution.	Bias reintroduction
Multilingual transfer and language drift (Lu et al., 2020)	Low-resource languages often depend on shared multilingual representations learned from high-resource languages.	Fine-tuning or repeated generation may erode low-resource capabilities faster than high-resource ones.	Language drift
Tokenizer and morphology mismatch (Rajab, 2022)	Agglutinative languages, code-switching, and non-standard orthographies may create longer-tailed token distributions.	Model collapse may manifest differently across languages, making universal mitigation strategies unreliable.	Measurement gap

Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
Synthetic-data feedback loops (Whitney & Norman, 2024)	Small real-data distributions are more easily overwhelmed by generated content.	Dominant patterns are reinforced across generations, while rare linguistic and cultural forms are pushed further into the tail.	Representation loss
Tail erasure under model collapse (Wyllie et al., 2024)	Low-resource languages, dialects, and marginalized viewpoints often already occupy low-probability regions of the data distribution.	Loss of diversity, reduced expressivity, and weaker model performance on communities whose data is already scarce.	Cultural erosion

Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
Value-lock and hegemonic viewpoints (Bender et al., 2021)	LLMs tend to reproduce frequent historical patterns rather than socially evolving or locally grounded meanings.	Outdated or dominant viewpoints persist, while new terms, norms, and movements struggle to enter the model distribution.	Bias reintroduction

Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
Changed scaling behaviour with synthetic data (Shumailov et al., 2024)	If more data yields less improvement, communities with limited access to real data and compute have fewer viable routes to improvement.	Higher data and compute requirements, widening the gap between high- and low-resource settings.	Unequal access

Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

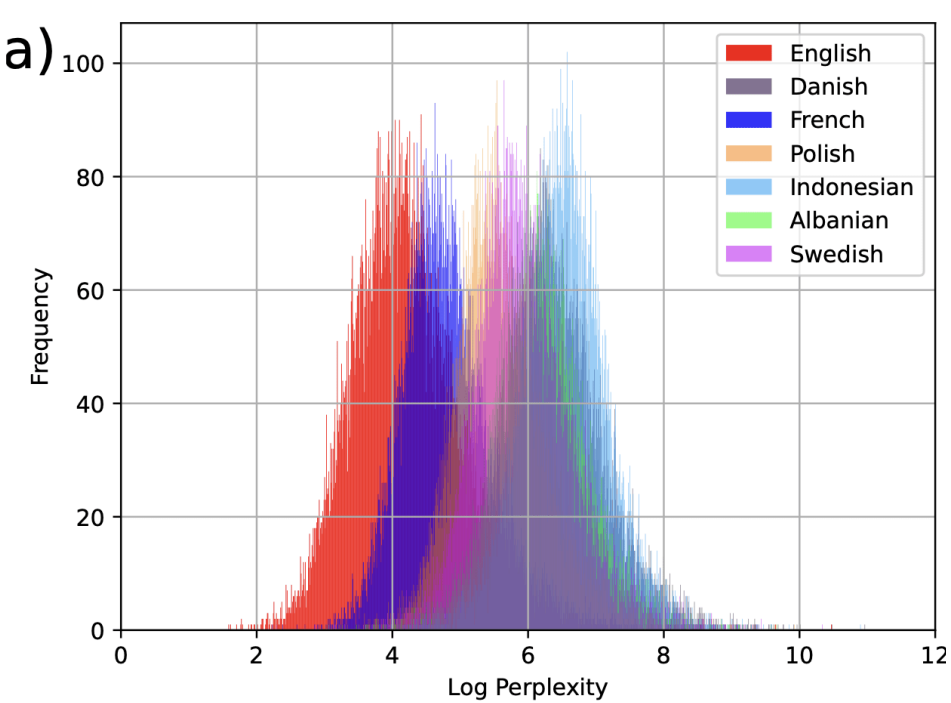
Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
High cost of real data acquisition (Kandpal & Raffel, 2025)	Collecting, validating, and maintaining culturally representative data is expensive and often unpaid or under-resourced.	Low-resource communities may be pushed toward synthetic supplementation, increasing the risk of poisoning their own data.	Data inequity
Automated curation and synthetic-data detection (Boutadjine et al., 2025)	Synthetic detection and filtering systems may be unreliable, opaque, or biased against uncommon language use.	Authentic but rare or reclaimed language may be removed, while synthetic or hegemonic content remains.	Curation bias

Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

Cause / mechanism	Why it matters for low-resource communities	Likely effect	Concerns
Multilingual transfer and language drift (Lu et al., 2020)	Low-resource languages often depend on shared multilingual representations learned from high-resource languages.	Fine-tuning or repeated generation may erode low-resource capabilities faster than high-resource ones.	Language drift
Tokenizer and morphology mismatch (Rajab, 2022)	Agglutinative languages, code-switching, and non-standard orthographies may create longer-tailed token distributions.	Model collapse may manifest differently across languages, making universal mitigation strategies unreliable.	Measurement gap

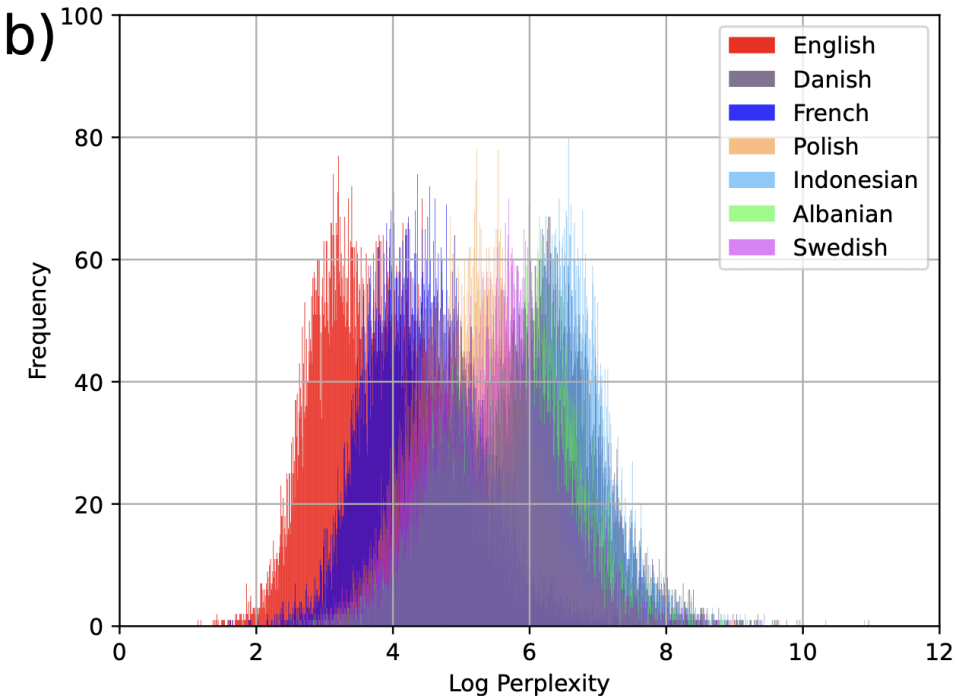
Table 1. Mechanisms by which model collapse may disproportionately affect low-resource and marginalized communities.

Transfer Learning



Original Data

15000 Short Stories



Generation 1 Data

10000 Stories + 5000 Synthetic

Alternative Views

- "Model Collapse" conflates a lot of phenomena in the literature and has not clear definition.
  - Model collapse may be less pessimistic and the assumptions of collapse do not match real world settings. [6]

[6] Schaeffer, R., Kazdan, J., Arulandu, A. C., & Koyejo, S. (2025). Position: Model collapse does not mean what you think. *arXiv preprint arXiv:2503.03150*

Alternative Views

- "Model Collapse" conflates a lot of phenomena in the literature and has not clear definition.
  - Model collapse may be less pessimistic and the assumptions of collapse do not match real world settings. [6]
- Model Collapse may be useful:
  - Human language evolves over generations in a similar manner and becomes more structured. [7]

[6] Schaeffer, R., Kazdan, J., Arulandu, A. C., & Koyejo, S. (2025). Position: Model collapse does not mean what you think. *arXiv preprint arXiv:2503.03150*.

[7] Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686.

Alternative Views

- "Model Collapse" conflates a lot of phenomena in the literature and has not clear definition.
  - Model collapse may be less pessimistic and the assumptions of collapse do not match real world settings. [6]
- Model Collapse may be useful:
  - Human language evolves over generations in a similar manner and becomes more structured. [7]
- The Stochastic Parrots analogy may be outdated:
  - Models are becoming more grounded and context aware.

[6] Schaeffer, R., Kazdan, J., Arulandu, A. C., & Koyejo, S. (2025). Position: Model collapse does not mean what you think. *arXiv preprint arXiv:2503.03150*.

[7] Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681-10686.

Conclusion and Call to Act

- Regardless of which definition we use or how much data is needed to avoid collapse, some culture on Earth is in a regime of worst case Model Collapse.

Conclusion and Call to Act

- Regardless of which definition we use or how much data is needed to avoid collapse, some culture on Earth is in a regime of worst case Model Collapse.

- Due to the potential to undermine efforts to democratise AI and the internet, more deliberate work is needed by low-resource and AI fairness communities.

Position: the Stochastic Parrot in the Coal Mine
Model Collapse is a Threat to Low-Resource Communities

Devon Jarvis Richard Klein Benjamin Rosman Steven James Stefano Sarao Mannelli

ICML 2026

Poster Session B; Hall A

July 9, 2026

5:00 PM – 6:45 PM