Amplifying Membership Signal Through Iterative Regeneration

Anonymous Authors[1]

Abstract

The tendency of large generative models to memorize training data has established sample verification as a critical necessity for privacy auditing and copyright enforcement. Current membership inference attacks (MIAs) often rely on "one-shot" generations, which yield weak signals and limited sensitivity across different modalities. Inspired by Model Autophagy Disorder (MAD), we introduce MADreMIA, a model-agnostic add-on framework that enhances white-, grey-, and black-box MIAs. Unlike conventional approaches that use a single query, MADreMIA utilizes chained generations – where each output informs the subsequent input – to amplify membership evidence. We demonstrate that training "re-members" exhibit significantly higher coherence and slower degradation during iterative regeneration than non-members generations. Our results across image, text, and audio modalities show that MADreMIA provides substantially richer signals for both membership and dataset inference across diverse model families, including IARs, diffusion, and large language models.

1. Introduction

The importance of Membership inference attacks (MIAs) (Shokri et al., 2017) and dataset inference (DIs) (Maini et al., 2021) has grown alongside the scale of generative models, which often ingest private or proprietary data without clear oversight. Practical auditing – ranging from protecting medical privacy (Zhang et al., 2022) to identifying licensed content (Dubiński et al., 2025) or detecting benchmark contamination (Maini et al., 2024; Singh et al., 2024; Zawalski et al., 2026) – requires looking beyond a model's general capabilities. The definitive test

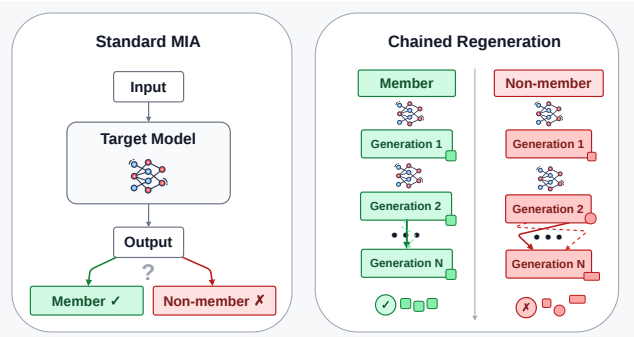


Figure 1. Comparison between conventional one-shot membership inference attack and our chained-generation approach. The former use a single query, which yields a weak signal that often fails to separate members from non-members. In the latter, each generation informs the next query, progressively amplifying membership evidence and improving separability: re-members ✓ are more coherent and degrade slower than re-non-members ✗.

becomes whether the model retains a structural "echo" of its training data, manifesting itself as a high-fidelity memorization signal that can be surfaced through targeted inference.

Most existing membership attacks extract evidence from a single query (Zhang et al., 2024; WU et al., 2024) or several loosely coupled samples (Choquette-Choo et al., 2021). We argue that such 'one-shot' interactions are insufficient to support high-confidence membership decisions, as they often fail to capture the subtle, deep-seated signals of memorization required for a robust audit.

Consider a suspect interrogation. A single question rarely distinguishes a plausible lie from the truth, but a sequential interrogation—where each query is conditioned on the previous answer—is far more revealing. In this setting, truthful narratives remain coherent under follow-up, whereas fabricated ones inevitably collapse. The diagnostic power lies not in an isolated response, but in the consistency of the evolving dialogue.

We view model auditing through this same lens. While a classical MIA effectively asks a single question – *Have you seen this example?* – and Dataset Inference (DI) asks a series of independent ones, we construct a chained interaction inspired by the recursive, *self-feeding* nature of Model Autophagy Disorder (MAD) (Alemohammad et al., 2023;

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by *The Impact of Memorization on Trustworthy Foundation Models* Workshop @ ICML. Do not distribute.

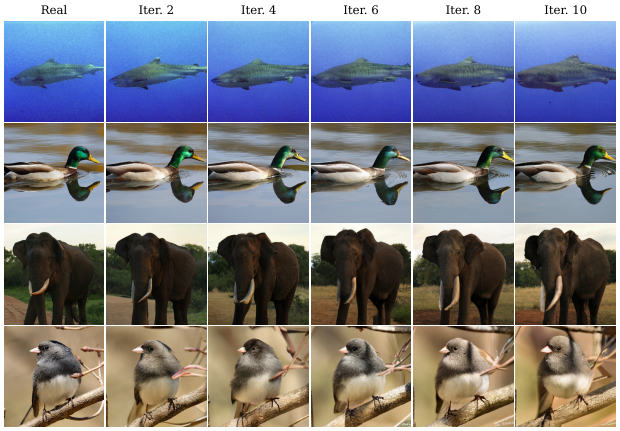


Figure 2. **VAR-d30 members: qualitative example.** Across iterative regeneration, image quality gradually declines, yet semantic coherence is largely preserved from one generation to the next.

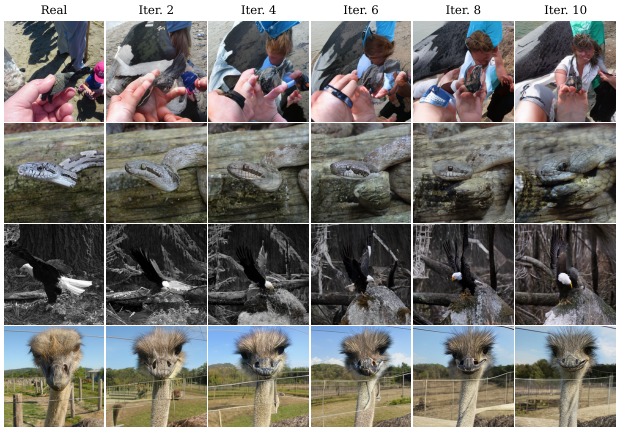


Figure 3. **VAR-d30 non-members: qualitative example.** Across iterative regeneration, image quality deteriorates faster and inter-generation coherence is less stable, with greater drift in content and structure.

Shumailov et al., 2024). Our approach, MADreMIA, is iteratively reconditioning the model on its own outputs – through image regeneration, audio "echo", or text expansion – to move beyond one-shot plausibility. We measure the model's ability to sustain a coherent chain of evidence, a signal that remains robust for training members but degrades rapidly for non-members.

Our framework is designed to be intentionally method-, model-, and modality-agnostic. By functioning as a modular add-on, it can be integrated with existing white-, grey-, or black-box membership inference attacks to significantly enhance their detection sensitivity. Although the technical implementation of the chaining mechanism varies by domain, the core principle remains universal: membership leaves an indelible footprint on the stability, fidelity, and self-consistency of iterative generations. This unified approach allows MADreMIA to transfer seamlessly across diverse architectures, including image autoregressive models (IARs), diffusion models (DMs), large language models (LLMs), and Voice Conversion (audio) models.

In summary, we challenge the status quo of one-shot auditing, arguing that it fundamentally underutilizes the information latent within generative models. We demonstrate that while a single output is often too noisy to be decisive, a chain of regenerated outputs acts as a powerful **signal amplifier**. By developing this concept into a model-agnostic add-on, we show that "re-members" maintain a structural persistence that non-members simply cannot sustain. Across diverse architectures and modalities, our results confirm that the key to a confident audit lies not in the first response, but in the consistency of the iterative dialogue.

2. Method

MADreMIA is a lightweight add-on to standard membership and dataset inference tasks (see Figure 4). In this section,

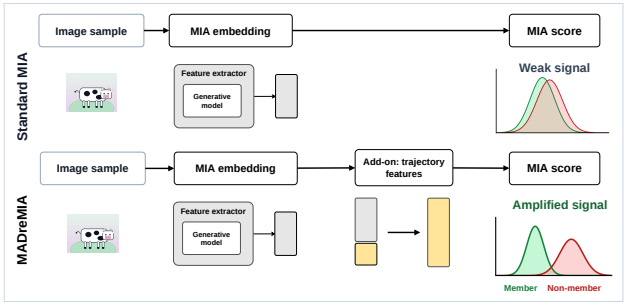


Figure 4. **MADreMIA: add-on chained regeneration features to strengthen the membership signal for generative models.** It is method type-agnostic: can be used for white-, grey- and black-box MIA methods.

we present it only as MIA add-on for simplicity. A one-shot MIA extracts features from a query sample and predicts a membership score, but this signal is often weak. We strengthen it by concatenating features from an iterative regeneration trajectory before calculating final MIA score.

Base MIA embedding. Given an input x, a baseline extractor ϕ_{mia} produces $z_{\text{mia}} = \phi_{\text{mia}}(x) \in \mathbb{R}^d$. This is the standard one-shot MIA representation.

Chained regeneration. We formalize our approach as a self-conditioned sequence:

$$x^{(t+1)} = \mathcal{R}(f, x^{(t)}), \quad x^{(0)} = x, \; t = 0, \ldots, T-1,$$

where f denotes the target generator and $\mathcal{R}$ represents a modality-specific regeneration function (e.g., image-to-image refinement, text continuation, or audio re-conditioning). Intuitively, if x is a training member, the model preserves structure and identity across generation trajectory more consistently; for non-members, the chain drifts faster and degrades in fidelity.

Feature fusion and scoring. Trajectory statistics are encoded as $z_{\text{traj}} = \psi(x^{(0)}, \ldots, x^{(T)}) \in \mathbb{R}^k$, and then concatenated with baseline evidence: $\tilde{z} = [z_{\text{mia}} \| \ z_{\text{traj}}]$. A scorer h outputs the final membership score $s(x) = h(\tilde{z})$. The scoring stage is unchanged; MADreMIA only enriches the input representation.

Key effect. MADreMIA converts weak one-shot evidence into a stronger temporal-consistency signal, improving member/non-member separability while remaining model- and modality-agnostic.

3. Experiments

We evaluate whether chained regeneration can be a signal amplifier for one-shot auditing across modalities, model families, and access regimes. Our preliminary analysis focuses on the following questions: **(Q1)** What distinguishes member/non-member chained generation trajectories? **(Q2)** Is it consistent between images, audio, and text? **(Q3)** Does MADreMIA increase member/non-member separability compared to one-shot MIA? **(Q4)** How does model size affect member/nonmember trajectory signals?

Experimental setup. To ensure a scientifically sound evaluation across our MIA tasks, we restrict our setup to models trained on public datasets with well-defined, identically and independently distributed (IID) training and test splits. We evaluate our method across three diverse modalities to demonstrate its broad applicability. For image generation, we analyze SOTA autoregressive models (VAR-d{20, 24, 30} (Tian et al., 2024), RAR-{L, XL, XXL} (Yu et al., 2024)) and diffusion models (DiT-RF-{XL, G} (Fei et al., 2024), UViT-T2I (Bao et al., 2023)), trained primarily on the ImageNet (Deng et al., 2009) or COCO (Veit et al., 2016) datasets for class-conditioned and text-to-image generation. We extend this evaluation to the audio domain using modern voice conversion frameworks (AutoVC (Qian et al., 2019), FreeVC (Li et al., 2023)), and to the language domain utilizing prominent LLMs (LLaMA-13B (Touvron et al., 2023), Mamba-1.4B (Gu & Dao, 2023), GPT-NeoX-20B (Black et al., 2022), OLMo-7B (Groeneveld et al., 2024)). Comprehensive details regarding all specific models and datasets used in experiments are provided in the Appendix B and C.

Metrics. To measure similarity between feature representations and their fidelity, we utilize the Fréchet Inception Distance (FID) (Heusel et al., 2017), and Fréchet Audio Distance (FAD) (Kilgour et al., 2018) for vision and audio models, respectively. For LLMs, we track Token Diversity, which we define as the Kullback-Leibler Divergence (KLD) between the empirical unigram token distributions of the initial evaluation iteration and the current iteration.

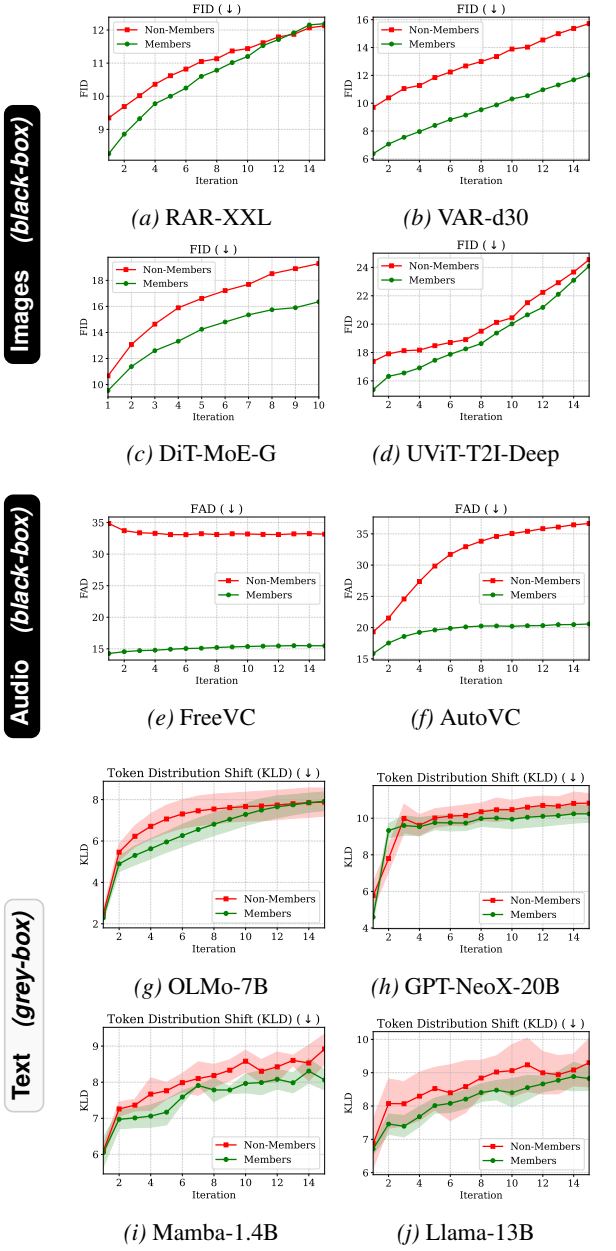


Figure 5. **Divergence trajectories across chained regeneration steps.** Rows represent image models (FID), audio models (FAD), and text models (KLD). Across modalities and access settings, member examples retain lower divergence and degrade more slowly than non-member examples, providing a robust signal for both membership and dataset inference.

Main qualitative finding: trajectory asymmetry. Across all modalities, members and non-members exhibit distinct regeneration dynamics. Members preserve structure longer and drift more slowly, while non-members degrade faster and diverge toward the model's generic prior. This pattern is visible both in per-step qualitative examples (Figures 2 and 3) and in aggregate divergence trajectories (Figure 5) comparing the quality of regenerations to base

Table 1. **MADreMIA Results.** (Quality: MSE_{sum} over trajectory iterations, Diversity:MSE_{std} within trajectory iterations). Δ denotes improvement over the respective Iter 0 baseline (standard MIA). Combined results utilize both quality and diversity features.

Method	TPR @ 1% FPR	AUC	ACC
VAR-d30			
Baseline (Iter 0)	0.041 ± 0.02	0.753 ± 0.01	0.617 ± 0.07
MADreMIA (Quality)	0.061 ± 0.07	0.747 ± 0.04	0.689 ± 0.04
Δ *vs Baseline*	+0.020	-0.006	+0.072
MADreMIA (Diversity)	0.070 ± 0.07	0.759 ± 0.03	0.700 ± 0.03
Δ *vs Baseline*	+0.029	+0.006	+0.083
MADreMIA (Combined)	**0.062** ± 0.07	**0.751** ± 0.04	**0.690** ± 0.03
Δ *vs Baseline*	**+0.021**	**-0.002**	**+0.073**
RAR-XXL			
Baseline (Iter 0)	0.048 ± 0.02	0.757 ± 0.02	0.566 ± 0.02
MADreMIA (Quality)	0.059 ± 0.06	0.759 ± 0.03	0.714 ± 0.03
Δ *vs Baseline*	+0.011	+0.002	+0.148
MADreMIA (Diversity)	0.060 ± 0.06	0.755 ± 0.03	0.709 ± 0.03
Δ *vs Baseline*	+0.012	-0.002	+0.143
MADreMIA (Combined)	**0.068** ± 0.07	**0.765** ± 0.03	**0.719** ± 0.03
Δ *vs Baseline*	**+0.020**	**+0.008**	**+0.153**

samples (FID for images, FAD for audio) and the drift of output token distribution in text model. The results presented support the core hypothesis that auto-regeneration trajectory contains multiple membership cues.

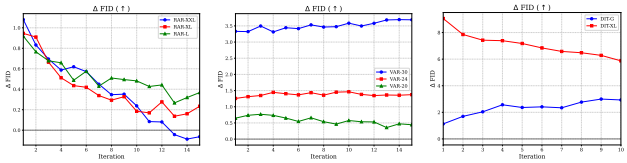
The key trajectory asymmetry findings are:

1. **Fidelity and degradation:** Re-members maintain high structural quality throughout the trajectory, whereas re-non-members exhibit rapid perceptual and semantic degradation.

2. **Persistence and divergence:** Re-members demonstrate significant structural persistence and coherence across iterations. Conversely, re-non-members diverge more quickly, drifting toward the model's general distribution and losing the specific characteristics of the original input.

The asymmetry is present across diverse models and modalities. We test broad architectural diversity: image autoregressive and diffusion models, audio voice conversion/generation models, and text generative models. Figure 5 summarizes trajectory behavior using modality-appropriate divergence metrics 3. This design directly tests whether our proposed signal amplification is model- and modality-agnostic.

MADreMIA amplifies baseline MIA. Table 1 evaluates the MADreMIA framework by comparing auxiliary-signal-augmented attacks against the one-shot baseline (Iter 0) across two distinct model families. The results demonstrate that augmenting the baseline with reconstruction-based quality (MSE_{sum}) and diversity (MSE_{std}) signals consistently improves attack effectiveness. Our MADreMIA (Combined) approach proves particularly potent for the larger RAR-XXL model, yielding a TPR of 0.068 at a 1% FPR and a substantial +15.3 p.p. absolute gain in classification accuracy. This robust performance across architectures confirms that iterative reconstruction features provide critical, scalable signals for identifying training set membership.



(a) RAR *(b)* VAR *(c)* DiT-MoE

Figure 6. **Ablation: Trajectory asymmetry scaling across model families.** Membership separation (Δ FID) persists across model scales, confirming that iterative trajectory chaining consistently amplifies membership signals compared to one-shot baselines. Δ FID is calculated as FID(nonmem) - FID(mem).

Trajectory asymmetry scaling across model families. As illustrated in Figure 6, the membership signal – quantified by $\Delta\text{FID} = \text{FID}_{\text{nonmem}} - \text{FID}_{\text{mem}}$ – persists across all model scales, suggesting that the observed asymmetry is a fundamental property rather than an artifact of specific parameter regimes. While the magnitude of this separation varies by architecture – exhibiting a strong positive correlation with model size for VAR and DiT-MoE, while remaining largely scale-invariant in RARs – the underlying trend is robust: iterative trajectory chaining consistently exposes a larger membership gap compared to standard one-shot generations.

4. Summary

As generative models increasingly exhibit high-fidelity memorization, verifying whether specific samples were seen during training becomes central to privacy auditing and copyright attribution. Inspired by the recursive self-feeding dynamics of Model Autophagy Disorder (MAD), we introduce MADreMIA as a model-agnostic add-on for white-, grey-, and black-box generative settings: instead of relying on a single generation, it builds chained generations in which each output conditions the next. Our results show a consistent asymmetry for this iterative protocol – training re-members retain coherence and degrade more slowly, while non-members drift and deteriorate faster across image, text, and audio generators, including IAR, diffusion, and large language model families. Finally, we show preliminary results that by fusing trajectory-derived evidence with baseline MIA features, MADreMIA amplifies membership signals and improves separability between members and non-members.

References

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., and Baraniuk, R. Self-consuming generative models go mad. In *The Twelfth International Conference on Learning Representations*, 2023.

Bao, F., Nie, S., Xue, K., Cao, Y., Li, C., Su, H., and Zhu, J. All are worth words: A vit backbone for diffusion models. In *CVPR*, 2023.

Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., et al. Gpt-neox-20b: An open-source autoregressive language model. In *Proceedings of BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models*, pp. 95–136, 2022.

Choquette-Choo, C. A., Tramer, F., Carlini, N., and Papernot, N. Label-only membership inference attacks. In *International conference on machine learning*, pp. 1964–1974. PMLR, 2021.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Dubiński, J., Kowalczuk, A., Boenisch, F., and Dziedzic, A. Cdi: Copyrighted data identification in diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 18674–18684, 2025.

Fei, Z., Fan, M., Yu, C., Li, D., and Huang, J. Scaling diffusion transformers to 16 billion parameters, 2024. URL https://arxiv.org/abs/2407.11633.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

Groeneveld, D., Beltagy, I., Walsh, E., Bhagia, A., Kinney, R., Tafjord, O., Jha, A., Ivison, H., Magnusson, I., Wang, Y., et al. Olmo: Accelerating the science of language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15789–15809, 2024.

Gu, A. and Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Kilgour, K., Zuluaga, M., Roblek, D., and Sharifi, M. Fr\'echet audio distance: A metric for evaluating music enhancement algorithms. *arXiv preprint arXiv:1812.08466*, 2018.

Li, J., Tu, W., and Xiao, L. Freevc: Towards high-quality text-free one-shot voice conversion. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Maini, P., Yaghini, M., and Papernot, N. Dataset inference: Ownership resolution in machine learning. *arXiv preprint arXiv:2104.10706*, 2021.

Maini, P., Jia, H., Papernot, N., and Dziedzic, A. LLM dataset inference: Did you train on my dataset? *CoRR*, abs/2406.06443, 2024. doi: 10.48550/arXiv.2406.06443. URL https://doi.org/10.48550/arXiv.2406.06443.

Qian, K., Zhang, Y., Chang, S., Yang, X., and Hasegawa-Johnson, M. Autovc: Zero-shot voice style transfer with only autoencoder loss. In *International Conference on Machine Learning*, pp. 5210–5219. PMLR, 2019.

Shi, W., Ajith, A., Xia, M., Huang, Y., Liu, D., Blevins, T., Chen, D., and Zettlemoyer, L. Detecting pretraining data from large language models. *arXiv preprint arXiv:2310.16789*, 2023.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pp. 3–18. IEEE, 2017.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., and Gal, Y. Ai models collapse when trained on recursively generated data. *Nature*, 631:755–759, 07 2024. doi: 10.1038/s41586-024-07566-y.

Singh, A. K., Kocyigit, M. Y., Poulton, A., Esiobu, D., Lomeli, M., Szilvasy, G., and Hupkes, D. Evaluation data contamination in llms: how do we measure it and (when) does it matter? *arXiv preprint arXiv:2411.03923*, 2024.

Soldaini, L., Kinney, R., Bhagia, A., Schwenk, D., Atkinson, D., Authur, R., Bogin, B., Chandu, K., Dumas, J., Elazar, Y., et al. Dolma: An open corpus of three trillion tokens for language model pretraining research. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15725–15788, 2024.

Tian, K., Jiang, Y., Yuan, Z., Peng, B., and Wang, L. Visual autoregressive modeling: Scalable image generation via next-scale prediction, 2024. URL https://arxiv.org/abs/2404.02905.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

Veit, A., Matera, T., Neumann, L., Matas, J., and Belongie, S. Coco-text: Dataset and benchmark for text detection and recognition in natural images. *arXiv preprint arXiv:1601.07140*, 2016.

WU, Y., Qiu, H., Guo, S., Li, J., and Zhang, T. You only query once: An efficient label-only membership inference attack. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=7WsivwyHrS.

Yamagishi, J., Veaux, C., and MacDonald, K. Cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit (version 0.92). *The Rainbow Passage which the speakers read out can be found in the International Dialects of English Archive:(http://web. ku. edu/˜ idea/readings/rainbow. htm)*., 2019.

Yu, Q., He, J., Deng, X., Shen, X., and Chen, L.-C. Randomized autoregressive visual generation, 2024. URL https://arxiv.org/abs/2411.00776.

Zawalski, M., Boubdir, M., Bałazy, K., Nushi, B., and Ribalta, P. Detecting data contamination in LLMs via in-context learning. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=YlpaaYxx4t.

Zen, H., Dang, V., Clark, R., Zhang, Y., Weiss, R. J., Jia, Y., Chen, Z., and Wu, Y. Libritts: A corpus derived from librispeech for text-to-speech. *arXiv preprint arXiv:1904.02882*, 2019.

Zhang, J., Sun, J., Yeats, E. C., Ouyang, Y., Kuo, M., Zhang, J., Yang, H., and Li, H. H. Min-k%++: Improved baseline for detecting pre-training data from large language models. *CoRR*, abs/2404.02936, 2024. doi: 10.48550/arXiv.2404.02936. URL https://doi.org/10.48550/arXiv.2404.02936.

Zhang, Z., Yan, C., and Malin, B. A. Membership inference attacks against synthetic health data. *Journal of biomedical informatics*, 125:103977, 2022.

A. Impact Statement

This work advances methods for auditing generative models by improving membership and dataset inference through chained regeneration. The primary positive impact is stronger accountability: MADreMIA can help detect memorization of sensitive, proprietary, or benchmark data, supporting privacy audits, copyright verification, and unlearning validation across model families and modalities.

While enhanced inference capabilities can assist in model auditing and transparency, they also require responsible application to avoid potential misuse. We frame MADreMIA as a tool for research evaluation, compliance monitoring, and internal red-teaming. It is important to note that our method provides statistical evidence rather than a definitive proof of data inclusion; therefore, results should be interpreted alongside additional forensic and procedural evidence within a broader data governance framework.

B. Model Details

In our experiments, we consider two vision model families: image autoregressive models (IARs) and diffusion models. The IAR category includes VAR (Tian et al., 2024) and RAR (Yu et al., 2024) variants, while the diffusion category includes DiT-MoE (Fei et al., 2024) and UViT-T2I (Bao et al., 2023). Furthermore, as others modalities, we evaluate large language models (LLMs) and voice conversion (VC) models. The LLMs include Mamba (Gu & Dao, 2023), OLMo (Groeneveld et al., 2024), GPT-NeoX (Black et al., 2022), and Llama (Touvron et al., 2023), while the VC models consist of AutoVC (Qian et al., 2019) and FreeVC (Li et al., 2023). Across all settings, we focus on representative, high-performing model variants.

Table 2. Vision model details.

	IAR Models						**Diffusion Models**		
	VAR-d30	VAR-d24	VAR-d20	RAR-XXL	RAR-XL	RAR-L	DiT-MoE-G	DiT-MoE-XL	UViT-T2I-Deep
Model parameters	2.1B	1.0B	600M	1.5B	955M	462M	16.5B	4.1B	141M
Training epochs	350	300	250	400	400	400	—	—	—
FID	1.92	2.33	2.95	1.48	1.50	1.70	1.72	2.10	5.48

Table 3. Language model details.

	Mamba	**OLMo**	**GPT-NeoX**	**Llama**
Model parameters	1.4B	7B	20B	13B
Training tokens	300B	2.46T	400B	1T

Table 4. Audio model details.

	AutoVC	**FreeVC**
Model parameters	28M	39M
Training data (hours)	44	40
SMOS (seen-to-seen)	3.5	4.1

C. Dataset Details

For vision and audio models that have publicly known and available train/test splits we use these datasets. For most LLMs we use established MIA benchmarks (e.g. WikiMIA), but for OLMo and GPT-Neox we use their corresponding training sets and the Global News as non-member set as suggested in (Zawalski et al., 2026).

Table 5. Datasets used to construct member and non-member sets for each model family in our experiments, spanning vision, language, and speech domains.

Model	**Members**	**Non-members**
VAR	ImageNet (Deng et al., 2009)	ImageNet
RAR	ImageNet	ImageNet
MAR	ImageNet	ImageNet
DiT-MoE	ImageNet	ImageNet
UViT-T2I	COCO (Veit et al., 2016)	COCO
Mamba	WikiMIA (Shi et al., 2023)	WikiMIA
GPT-Neox	The Pile (Gao et al., 2020)	Global News
OLMo	Dolma (Soldaini et al., 2024)	Global News
Llama	WikiMIA	WikiMIA
AutoVC	VCTK (Yamagishi et al., 2019)	LibriTTS (Zen et al., 2019)
FreeVC	VCTK	LibriTTS