
Self-Verification Provably Prevents Model Collapse in Recursive Synthetic Training

Shi Fu¹ Yingjie Wang^{1*} Yuzhu Chen² Li Shen³ Dacheng Tao^{1*}

¹College of Computing and Data Science, Nanyang Technological University, Singapore,

²University of Science and Technology of China, Hefei, China,

³Shenzhen Campus of Sun Yat-sen University, Shenzhen, China

shi011@e.ntu.edu.sg, yingjiewang@upc.edu.cn,

cyzkrau@mail.ustc.edu.cn, mathshenli@gmail.com, dacheng.tao@gmail.com

Abstract

Large generative models are increasingly trained on synthetic data from earlier generations, raising concerns about *model collapse*, a progressive performance decline consistently observed in empirical studies. However, theoretical understanding of recursive training dynamics and their failure modes remains limited. In this work, we theoretically show that recursive training inherently leads to exponential error growth unless mitigated by sufficient real data. Addressing the growing scarcity of real data, we introduce a self-verification mechanism enabling models to filter their outputs based on internal confidence scores without external validation. Through rigorous analysis, we derive finite-sample error bounds demonstrating that self-verification alone can prevent collapse, even in fully synthetic training regimes. Our theoretical framework extends to large language models (LLMs), characterizing the conditions under which recursive training can maintain stability without performance degradation.

1 Introduction

High-quality training data is a critical foundation for the remarkable success of large generative models, such as LLMs [Fu et al., 2024a]. However, it is increasingly acknowledged that the pool of publicly available, real-world data is nearing exhaustion [Villalobos et al., 2022]. As a result, the training of modern LLMs increasingly relies on synthetic data produced by previous generations of the models [Briesch et al., 2023, Martínez et al., 2023, Xing et al., 2025]. Moreover, even unintentionally, models today are often trained on synthetic data, as many existing datasets are already polluted with synthetic content [Schuhmann et al., 2022]. This paradigm, known as recursive synthetic data training, has raised significant concerns regarding the risk of *model collapse*, a phenomenon characterized by a drastic and often irreversible deterioration in model performance across generations [Shumailov et al., 2024, Alemohammad et al., 2024a, Bertrand et al., 2024, Dohmatob et al., 2024b].

Numerous empirical studies have highlighted the potential risks associated with model collapse in recursive training loops, including substantial reductions in output diversity [Guo et al., 2023, Zhu et al., 2024], sharply increasing error rates [Dohmatob et al., 2024b,c], and the amplification of biases [Wyllie et al., 2024]. To mitigate these risks, researchers have proposed various approaches, such as incorporating sufficient real data [Alemohammad et al., 2024a], augmenting synthetic datasets [Gerstgrasser et al., 2024], and introducing external verification mechanisms for synthesized data [Feng et al., 2025, Firdoussi et al., 2025]. Nevertheless, despite these empirical advances, there remains a notable gap in rigorous theoretical understanding of both the observed phenomena and the underlying dynamics of recursive synthetic training loops.

*Corresponding authors

Initial theoretical explorations have started to address these gaps by analyzing population risk dynamics under specific assumptions [Bertrand et al., 2024, Dohmatob et al., 2024a, Seddik et al., 2024]. For instance, Bertrand et al. [2024] derived upper bounds on parameter deviations for likelihood-based models when real data is incorporated, establishing convergence under strict statistical and optimization assumptions. Further work by Fu et al. [2024b, 2025] relaxed these assumptions on error bounds through more fine-grained analyses. Fu et al. [2024b] provided bounds on the Total Variation distance for simplified diffusion models, linking distributional divergence to dataset size and the proportion of real data. Complementing this, Fu et al. [2025] introduced the concept of recursive stability to establish the first generalization error bounds. They demonstrated for transformer architectures that a constant proportion of real data is sufficient to prevent model collapse.

More recently, theoretical studies focusing on verification strategies have gained particular attention [Feng et al., 2025, Firdoussi et al., 2025]. These works investigated Gaussian mixture models with linear classifiers in non-recursive settings, modeling external verification of synthesized data as Bernoulli random variables. In particular, Feng et al. [2025] demonstrated, in the infinite-sample limit, that linear verifiers can significantly improve model robustness when training solely on synthetic data, while Firdoussi et al. [2025] extended these findings to scenarios involving a mixture of real and synthetic samples.

However, existing theoretical analyses suffer from critical limitations [Feng et al., 2025, Firdoussi et al., 2025]. Current frameworks lack rigorous external verification and rely on simplified probabilistic assumptions for data quality. These analyses remain narrowly focused on Gaussian mixture models, limiting their applicability to broader generative models, particularly transformer-based LLMs used in practice. Previous theoretical studies are also restricted to non-recursive training scenarios, whereas model collapse emerges specifically within recursive settings [Schaeffer et al., 2025]. Additionally, analyses typically rely on asymptotic regimes with infinite-sample assumptions, failing to capture realistic finite-sample conditions relevant to actual deployment. This paper addresses these gaps with the following contributions:

1. **Unified Framework for Self-Verification in Recursive Training:** We propose a general theoretical framework for recursive training with self-verification, where models filter their own outputs based on internally estimated confidence scores. This approach removes the need for external verification and enables model-internal quality control applicable to a broad class of generative architectures.
2. **Finite-Sample Error Bounds Across Training Regimes:** We provide rigorous finite-sample error bounds across three recursive training regimes. First, we show that naive recursive training leads to exponential error accumulation across generations. We then demonstrate that incorporating sufficient real data mitigates this degradation. Most notably, we prove that self-verification alone suffices to prevent model collapse, even in fully synthetic settings, by ensuring provable convergence of error.
3. **Theoretical Guarantees for Transformer-Based LLMs:** We extend our analysis to transformer-based LLMs, establishing convergence guarantees for self-verification through recursive coverage coefficient control and PAC-Bayesian analysis. Our results provide the first theoretical justification for stable, fully synthetic recursive training in high-capacity transformer architectures.

2 Related Work

High-quality training data is essential to the success of large generative models. However, the vast supply of human-annotated or naturally occurring data on the internet is nearing exhaustion [Villalobos et al., 2022]. As a result, LLMs are increasingly trained on data generated by earlier versions of themselves, forming recursive training loops, also known as self-consuming loops [Martínez et al., 2023, Bohacek and Farid, 2023, Shumailov et al., 2024, Tao et al., 2024, Schaeffer et al., 2025]. Recent high-profile research has highlighted the potential for dramatic degradation in model performance, a phenomenon known as *model collapse*, which represents a critical challenge for the future of AI development and deployment [Gibney, 2024, Alemohammad et al., 2024a, Dohmatob et al., 2024b].

Empirical studies have documented a number of alarming phenomena associated with model collapse in recursive training loops, including a substantial reduction in output diversity [Guo et al., 2023, Shumailov et al., 2024, Briesch et al., 2023, Zhu et al., 2024], rapidly increasing error rates [Fu et al., 2024b], and bias amplification [Wyllie et al., 2024]. To mitigate these effects, prior work has proposed several strategies, such as incorporating real data into training [Alemohammad et al., 2024a,

Kanabar and Gastpar, 2025], enlarging synthetic datasets [Dohmatob et al., 2024a, Gerstgrasser et al., 2024, Kazdan et al., 2024], or guiding the data generation process using reward or control signals [Gillman et al., 2024, Alemohammad et al., 2024b, Feng et al., 2025, Firdoussi et al., 2025].

Despite the wealth of empirical investigations, theoretical understanding of recursive synthetic training loops remains relatively limited [Bertrand et al., 2024, Dohmatob et al., 2024c, Fu et al., 2024b, 2025, Seddik et al., 2024, Dey and Donoho, 2024]. A line of recent work has begun to explore this direction by analyzing population risk dynamics under specific modeling assumptions, such as linear contexts [Dohmatob et al., 2024a], Gaussian models [Shumailov et al., 2024, Alemohammad et al., 2024a, Suresh et al., 2024, Jain et al., 2024], and asymptotic regimes [Marchi et al., 2024]. Furthermore, Bertrand et al. [2024] derived upper bounds on parameter drift in likelihood-based models under strong statistical assumptions, while Fu et al. [2024b, 2025] relaxed these constraints and provided generalization bounds for simplified diffusion and attention-based architectures.

Most closely related to our work are Feng et al. [2025] and Firdoussi et al. [2025], which introduce verification strategies into training loops to prevent model collapse. These studies investigated Gaussian mixture models with linear classifiers in non-recursive settings under the infinite sample size assumption. They modeled external verification of synthesized data as Bernoulli random variables and examined the impact on model performance. Specifically, Feng et al. [2025] demonstrated that linear verifiers can provide feedback that improves the robustness of models trained on fully synthesized data, while Firdoussi et al. [2025] extended these findings to mixed data scenarios.

Our work further advances this line of research by introducing the self-verification mechanism, extending the theoretical framework to general architectures like transformers, analyzing error accumulation in recursive settings, and deriving practical finite sample bounds.

3 Preliminary

In this section, we formalize the recursive training framework, where each model is trained on data generated by its predecessor. We then outline the self-verification mechanism, which enables models to evaluate and filter their own outputs to ensure stable training.

Recursive Training Loops. Let $\{\mathcal{G}_t\}_{t=0}^T$ be a sequence of models trained recursively. The initial model $\mathcal{G}_0$ is trained on a real dataset $S_0 = \{(x_{0,j}, y_{0,j}^*)\}_{j=1}^n \subseteq \mathcal{X} \times \mathcal{Y}$, where $x_{0,j}$ are inputs sampled from the data distribution and $y_{0,j}^*$ are ground-truth labels. For each generation $t = 1, \dots, T$, model $\mathcal{G}_t$ is trained on a synthetic dataset $S_t = \{(x_{t,j}, \hat{y}_{t,j})\}_{j=1}^n$, where $x_{t,j} \sim \mathcal{X}$ and $\hat{y}_{t,j} \sim \mathcal{G}_{t-1}(x_{t,j})$ are labels predicted by the previous model. The process continues recursively until generation T .

Next, we introduce a structured framework that integrates self-verification into each training generation, consisting of two key components: (1) a confidence-guided filtering mechanism for self-verification, and (2) a recursive training procedure leveraging verified data. Together, these components enable robust self-evolution through selective data curation.

Self-Verification via Confidence-Guided Filtering. In the absence of external supervision, we introduce a self-verification mechanism that allows the model to evaluate the quality of its own outputs. The verification score is computed using a fixed evaluator $\mathcal{G}_0$, trained solely on the initial real dataset, and reflects the alignment of a generated output with the base model’s inductive biases. Formally, the score for output y given input x is defined as:

$$s_{\text{verify}}(y \mid x) := \log \mathcal{G}_0(y \mid x).$$

This design offers two main benefits: it anchors verification to a stable reference, mitigating distributional drift across generations, and acts as an implicit regularizer that discourages semantic deviation from the base distribution.

Given an input $x_{t,j}$, we sample N candidate outputs from the previous model $\mathcal{G}_{t-1}$, denoted $\{\hat{y}_{t,j,k}\}_{k=1}^N \sim \mathcal{G}_{t-1}(\cdot \mid x_{t,j})$. We then apply a verification-based filtering step that retains only candidates whose scores fall within a margin $\gamma > 0$ of the highest score in the sampled set:

$$\mathbf{y}_{t,j,\gamma}^+ = \left\{ \hat{y}_{t,j,k} \mid s_{\text{verify}}(\hat{y}_{t,j,k} \mid x_{t,j}) \geq \max_{1 \leq k' \leq N} s_{\text{verify}}(\hat{y}_{t,j,k'} \mid x_{t,j}) - \gamma \right\},$$

A verification-weighted distribution over $\mathbf{y}_{t,j,\gamma}^+$ is then defined via a softmax over the scores:

$$\mathcal{G}_{t,N}^*(\hat{y} \mid x_{t,j}) = \frac{\exp(s_{\text{verify}}(\hat{y} \mid x_{t,j}))}{\sum_{\hat{y}' \in \mathbf{y}_{t,j,\gamma}^+} \exp(s_{\text{verify}}(\hat{y}' \mid x_{t,j}))}, \text{ where } \hat{y} \in \mathbf{y}_{t,j,\gamma}^+.$$

Then, a verified output $\hat{y}_{t,j}$ is sampled from this distribution. This mechanism allows the model to act as both generator and verifier, enforcing quality control in recursive training.

Recursive Training with Verified Data. The verified outputs from confidence-guided filtering are used to construct the synthetic training set for the next generation. For each input $x_{t,j}$, we sample a verified target $\hat{y}_{t,j} \sim \mathcal{G}_{t,N}^*(\cdot \mid x_{t,j})$, forming the dataset $S_t = \{(x_{t,j}, \hat{y}_{t,j})\}_{j=1}^n$. The next model $\mathcal{G}_t$ is trained by maximizing the log-likelihood over S_t :

$$\mathcal{G}_t = \arg \max_{\mathcal{G} \in \mathcal{H}} \sum_{j=1}^n \log \mathcal{G}(\hat{y}_{t,j} \mid x_{t,j}),$$

where $\mathcal{H}$ is the model hypothesis class. This process facilitates iterative self-improvement by leveraging verified high-quality outputs for training.

4 Main Results: Understanding and Preventing Model Collapse

In this section, we present our theoretical analysis of model collapse in recursive synthetic training to understand when and how this phenomenon can be prevented. We first demonstrate that, without intervention, error accumulates exponentially across generations, leading to inevitable model collapse. We then demonstrate that introducing real data mitigates this collapse. Finally, we prove that a self-verification mechanism alone can prevent collapse, even under fully synthetic training.

4.1 Assumptions and Definitions

We begin by stating the core assumptions and definitions that underpin our analysis.

Assumption 1 (Confidence-Calibrated Agreement). Let $\mathcal{G}_0$ denote the base model trained on a ground-truth real dataset $S_0 = \{(x_j, y_j^*)\}_{j=1}^n$. For any input $x \in \mathcal{X}$, define the verified high-confidence prediction as: $\mathbf{y}_\gamma^+ = \{y \mid s_{\text{verify}}(y \mid x) \geq \max_{y' \in \mathcal{Y}} s_{\text{verify}}(y' \mid x) - \gamma\}$, where s_{verify} is chosen as $\log \mathcal{G}_0$. Let $y_t^+(x) = \arg \max_{y \in \mathcal{Y}} \mathcal{G}_t(y \mid x)$ denote the prediction of $\mathcal{G}_t$. Then, we assume the existence of a constant $\tau \in (0, 1)$ such that:

$$\mathbb{P}_{x \sim \mathcal{X}} [y_t^+(x) = y^*(x) \mid \mathcal{G}_t(\mathbf{y}_\gamma^+(x) \mid x) \geq 1 - \tau] \geq 1 - \epsilon(\gamma),$$

where $\epsilon(\gamma)$ is a function that controls the residual error, τ is a fixed constant representing the confidence threshold, and $y^*(x)$ denotes the ground-truth label.

As the filtering becomes more selective ($\gamma \rightarrow 0$), we expect $\epsilon(\gamma) \rightarrow 0$ if the base model $\mathcal{G}_0$, trained on a real dataset, is well-calibrated. This expectation is reasonable, as extensive prior work has explored confidence calibration techniques that enable models to achieve well-calibrated predictions [Mehrshah et al., 2020, Wang et al., 2021, Zhu et al., 2022, Liu et al., 2025]. Furthermore, similar assumptions have been extensively employed in the literature on self-learning paradigms [Huang et al., 2025a,b].

Definition 1 (Minimum Confidence). Let $y_t^+(x) = \arg \max_{y \in \mathcal{Y}} \mathcal{G}_t(y \mid x)$ denote the model's top prediction at generation t . We define the minimum confidence level of these predictions as: $C_0 = \min_{x \sim \mathcal{X}, t \in [0, T]} \mathcal{G}_t(y_t^+(x) \mid x)$.

The quantity C_0 denotes the minimum confidence assigned by any model in the sequence to its top prediction over the input distribution. Ensuring $C_0 > 0$ guarantees that a non-trivial level of certainty is maintained throughout the recursive training process. This assumption is standard in prior works [Zhang et al., 2023, Huang et al., 2025a].

Definition 2 (Test Error). We define the test error of the final model $\mathcal{G}_T$ after T recursive generations as: $\text{Err}(\mathcal{G}_T) = \mathbb{P}_{x \sim \mathcal{X}} [y_T^+(x) \neq y^*(x)]$, where $y_T^+(x) = \arg \max_{y \in \mathcal{Y}} \mathcal{G}_T(y \mid x)$ is the model's prediction, and $y^*(x)$ is the ground-truth label.

This metric provides a clear measure of model performance and serves as the foundation for analyzing error propagation in recursive training.

4.2 Model Collapse in Naive Recursive Training

We begin by analyzing the failure mode of naive recursive training, where models are trained solely on synthetic data generated by their predecessors without any verification. In this setting, performance inevitably degrades over generations, leading to eventual collapse. Our first theorem shows that prediction errors propagate and amplify exponentially across recursive steps.

Theorem 1 (Error Propagation in Naive Recursive Training). *Let $\{\mathcal{G}_t\}_{t=0}^T$ denote a sequence of models trained via naive recursive training, where each $\mathcal{G}_t$ is trained solely on synthetic data generated by its predecessor $\mathcal{G}_{t-1}$, without any verification. Assume that for all $t \in [1, T]$, the Total Variation (TV) distance between successive generations satisfies $D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t-1}(\cdot | x)) \asymp D_{\text{TV}}$, where D_{TV} denotes the characteristic magnitude of distributional shift across generations. Define the coverage coefficient as $C_\gamma := \mathbb{E}_{x \sim \mathcal{X}} [1/\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)]$. Suppose the final model $\mathcal{G}_T$ satisfies Assumption 1. Then the test error after T recursive steps satisfies:*

$$\text{Err}(\mathcal{G}_T) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \cdot D_{\text{TV}} + \left(C_\gamma + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \right) \cdot \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T. \quad (1)$$

Remark 1. Exponential Error Amplification in Naive Recursive Training. This theorem underscores the core limitation of naive recursive training: the prediction error grows exponentially with the number of generations T when the condition $\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} > 1$ is met. Here, C_0 , defined in Definition 1, denotes the minimum confidence assigned to the model’s top prediction across all generations. In practice, C_0 tends to be small, making the condition readily satisfied. Notably, when $C_0 \leq 0.5$, the error term in Equation 1 exhibits exponential growth as T increases. This growth stems from the fact that each model not only inherits information from its predecessor but also compounds its biases and errors, leading to eventual model collapse.

This result aligns with prior theoretical works [Bertrand et al., 2024, Gillman et al., 2024], which shows that recursive training without control leads to exponential error amplification. However, these analyses directly assume upper bounds on optimization and statistical errors from finite sampling, and typically posit unrealistic iterative retraining where each model inherits its predecessor’s parameters and optimizer state [Schaeffer et al., 2025]. In contrast, we analyze training dynamics for specific generative models, including Transformer-based LLMs (Appendix I, Theorem 7), under a more practical setting where each model is trained from scratch on synthetic data.

The phenomenon of model collapse observed here is also consistent with empirical observations in recursive training setups, where the absence of quality control mechanisms results in severe performance degradation over successive generations [Shumailov et al., 2024, Alemohammad et al., 2024a]. This theoretical result formalizes these observations, providing a rigorous explanation for the inevitable instability in naive recursive training.

4.3 Mitigating Collapse with Real Data

A common approach to prevent model collapse is to inject real data during training. While this strategy has been empirically validated in previous work [Alemohammad et al., 2024a], we provide a theoretical characterization of how real data injection affects error propagation.

Theorem 2 (Error Propagation with Real Data). *Let $\{\mathcal{G}_t\}_{t=0}^T$ be a sequence of models trained via recursive training that incorporates a proportion $\alpha \in [0, 1]$ of real data $\mathcal{D}_{\text{real}} = \{(x, y^*)\}$ at each generation. The remaining $1 - \alpha$ proportion of the training data comes from synthetic data generated by the previous model $\mathcal{G}_{t-1}$. Suppose that the TV distance satisfies $D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t-1}(\cdot | x)) \asymp D_{\text{TV}}$ for all t , and that the final model $\mathcal{G}_T$ satisfies Assumption 1. Assume $\alpha > 1 - \frac{1}{2(1/C_0 - 1)}$. Then as $T \rightarrow +\infty$, the final error after T recursive steps satisfies:*

$$\text{Err}(\mathcal{G}_T) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1 - \alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}.$$

Remark 2. Error Stabilization through Real Data Injection. This theorem demonstrates the stabilizing effect of incorporating real data into recursive training loops. Specifically, when the proportion of real data α satisfies the condition $\alpha > 1 - \frac{1}{2(1/C_0 - 1)}$, the exponential error growth observed in naive recursive training (Theorem 1) is mitigated. The resulting error bound becomes stable as $T \rightarrow +\infty$, with the dominant terms depending on $\epsilon(\gamma)$, the total variation distance D_{TV} , and the proportion of synthetic data $(1 - \alpha)$.

Notably, as discussed in Assumption 1, $\epsilon(\gamma)$ can be made arbitrarily small through confidence calibration techniques [Liu et al., 2025]. Meanwhile, the TV distance D_{TV} can diminish under certain theoretical conditions. Prior studies have shown that $D_{\text{TV}}(n)$ converges at a rate of $\mathcal{O}(n^{-1/4})$ in diffusion models [Fu et al., 2024b], with similar rates established in GANs [Liang, 2021]. These results suggest that as the number of training samples n increases, the distributional shift across generations diminishes, further stabilizing the overall error bound.

The key insight from this result is that introducing even a modest fraction of real data, $\alpha > 1 - \frac{1}{2(1/C_0-1)}$, fundamentally changes the error dynamics and ensures that the exponential amplification factor from Theorem 1 is suppressed. From a practical perspective, this theorem underscores the importance of incorporating real data in recursive training loops to ensure the long-term stability of model performance. Even a small fraction of real data acts as a corrective mechanism, counteracting the biases, distributional shifts, and errors introduced by synthetic data. This result also aligns with prior empirical and theoretical findings [Briesch et al., 2023, Bertrand et al., 2024, Fu et al., 2024b, 2025, Seddik et al., 2024] that emphasize the critical role of real data in mitigating error propagation.

Remark 3. Limitations of Real Data Dependency and Motivation for Self-Verification. While this result confirms the effectiveness of incorporating real data in preventing model collapse, it also highlights a significant limitation: the approach relies on access to an ongoing stream of high-quality real data. In many practical scenarios, such data may be scarce, expensive to obtain, or entirely unavailable [Villalobos et al., 2022, Alemohammad et al., 2024a]. Furthermore, in publicly available datasets, it may be challenging to accurately distinguish between real and synthetic data [Sadasivan et al., 2023], which can undermine the stability of the training process.

This limitation motivates the need for alternative strategies like self-verification. By internally validating and correcting the quality of synthetic data, a self-verification mechanism can prevent error propagation across generations, offering a more resource-efficient and generalizable solution to model collapse without depending on external data.

4.4 Mitigating Collapse through Self-Verification

Our next main contribution is to demonstrate that stable recursive training is possible without real data or external supervision, by leveraging a self-verification mechanism. This builds on the insight that a model’s verification ability often exceeds its generation capability [Weng et al., 2023, Huang et al., 2025a], allowing it to effectively filter its own outputs. We then present our main theoretical result, showing that self-verification can prevent model collapse:

Theorem 3 (Error Bound with Self-Verification). *Let $\{\mathcal{G}_t\}_{t=0}^T$ be a sequence of models obtained via recursive training with the self-verification mechanism described in Section 3. Assume that for all $t \in [1, T]$, the TV distance satisfies $D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t-1}(\cdot | x)) \asymp D_{\text{TV}}$. Suppose the final model $\mathcal{G}_T$ also satisfies Assumption 1. Then, after T recursive generations, the final model $\mathcal{G}_T$ satisfies:*

$$\text{Err}(\mathcal{G}_T) \lesssim \epsilon(\gamma) + \frac{1}{\tau} D_{\text{TV}} + \left(\frac{1}{NC_0} \log \frac{1}{C_0} \right)^T \frac{C_\gamma}{N} \log \frac{1}{\tau} + \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \right) \log \frac{1}{\tau}.$$

Furthermore, if $N > \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)$, the exponential term vanishes as $T \rightarrow \infty$, yielding:

$$\text{Err}(\mathcal{G}_T) \lesssim \epsilon(\gamma) + \frac{1}{\tau} D_{\text{TV}} + \frac{1}{N} \left(1 + \frac{1}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \right) \log \frac{1}{\tau}. \quad (2)$$

Remark 4. The Role of Self-Verification in Error Control. Theorem 3 demonstrates how self-verification prevents the exponential error growth shown in Theorem 1 during recursive training loops. Specifically, the method enforces quality control using a verification score $s_{\text{verify}}(y | x) = \log \mathcal{G}_0(y | x)$ which quantifies how well a generated output aligns with the distribution of the initial model. This reference model, trained on real data, remains fixed across generations and serves as a stable anchor for self-verification. By selecting outputs within a margin γ of the maximum score, the mechanism creates a filtered subset of high-quality candidates for training. This anchoring to the initial model $\mathcal{G}_0$ prevents distributional drift from compounding, ensuring each recursive generation remains tethered to a consistent inductive prior. As generations progress, outputs deviating from the original data manifold are excluded, effectively suppressing error amplification even without external supervision.

Specifically, Theorem 3 establishes that the exponential decay term vanishes asymptotically when $N > \log(1/C_0)/C_0$, where N , as described in Section 3, represents the number of data samples

used by the self-verification mechanism to select a subset of high-quality samples for training. This condition highlights the importance of both the size of the data pool and the quality of the selected samples in stabilizing the recursive training process. When this condition is met, the error bound converges to an asymptotic form, ensuring that the error remains bounded even as $T \rightarrow \infty$.

This result is particularly significant because it demonstrates that, through proper quality control mechanisms performed entirely within the model itself, recursive training can achieve sustained improvement without relying on real data or external verification. By leveraging self-verification to prioritize reliable outputs and suppress low-quality generations, the training process becomes robust to error amplification, enabling scalable and stable model development in resource-constrained or fully synthetic environments.

Remark 5. Comparison with previous works. The study of model collapse in recursive training loops was initiated by theoretical insights from Shumailov et al. [2024], Alemohammad et al. [2024a], which analyzed simplified Gaussian models under fully synthetic data settings. Most closely related to our work is the emerging line of research that introduces verification strategies into training loops to prevent model collapse [Feng et al., 2025, Firdoussi et al., 2025]. Specifically, Feng et al. [2025] and Firdoussi et al. [2025] investigated Gaussian mixture models with linear classifiers in non-recursive settings. Their analysis examined how external verification of synthesized data influences model performance as training dataset size approaches infinity. Feng et al. [2025] primarily demonstrated that appropriate feedback mechanisms can significantly improve the robustness of models trained on fully synthetic data, while Firdoussi et al. [2025] extended these findings to mixed data scenarios.

Our work makes significant theoretical advancements over Feng et al. [2025] and Firdoussi et al. [2025] in several key aspects:

1. **Self-Verification vs. External Verification:** While prior work often relies on external verifiers to guide data quality [Feng et al., 2025, Firdoussi et al., 2025], we propose a self-verification framework in which models assess their own outputs. This intrinsic approach removes the reliance on external quality signals, enables autonomous self-improvement. For advanced or superhuman models, internal verification capabilities may surpass what weak external mechanisms can offer [Wang et al., 2025].
2. **A General and Practical Theoretical Framework:** Our work considers a more general and practical theoretical framework compared to Feng et al. [2025] and Firdoussi et al. [2025], which focused on Gaussian mixture models and linear classifiers. Importantly, we do not impose any assumptions about data distribution, making our findings applicable to a wide range of real-world scenarios. Moreover, our framework addresses more realistic language model settings, and in Section 5, we extend our theoretical results to Transformer-based LLMs, demonstrating their applicability to modern and widely-used architectures.
3. **Recursive Training vs. Non-Recursive Settings:** While Feng et al. [2025] and Firdoussi et al. [2025] investigated non-recursive training settings, our work tackles the more complex and challenging recursive training setup, which is the primary scenario in which model collapse occurs [Shumailov et al., 2024, Alemohammad et al., 2024a, Schaeffer et al., 2025]. Recursive training involves compounding errors and distributional shifts across multiple generations, making it significantly harder to analyze theoretically. By addressing this setting, our work provides deeper insights into the dynamics of recursive training loops and offers solutions that are more meaningful and impactful for preventing model collapse in practical scenarios.
4. **Finite-Sample Guarantees vs. Asymptotic Analyses:** Prior work [Feng et al., 2025, Firdoussi et al., 2025] primarily established asymptotic results under infinite-sample assumptions, limiting their applicability to real-world training regimes. In contrast, we derive finite-sample generalization bounds that remain valid under practical recursive training scenarios.

In summary, we advance the field by introducing self-verification, developing a general framework applicable to transformer-based LLMs, analyzing recursive training dynamics, and deriving finite-sample guarantees. These contributions mark a significant step toward understanding and mitigating model collapse in generative models.

Remark 6. Proof Sketch of Theorem 3. Theorem 3 provides a recursive generalization bound for self-verifying models under synthetic training, and shows how error amplification can be effectively controlled through confidence-guided filtering. The key innovation lies in leveraging *recursive coverage coefficient control* to quantify the reliability of verified data and to suppress error propagation across recursive generations.

Step 1: Verification-Induced Quality Control and Error Decomposition. We leverage the verification score $s_{\text{verify}}(y | x) = \log \mathcal{G}_0(y | x)$ to filter model outputs and restrict supervision to high-quality generations. Let $\mathbf{y}_\gamma^+(x)$ denote the verified candidate set, and define the confident support region:

$$\mathcal{D}_1^+ := \{x \in \mathcal{X} \mid \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) \geq \log(2/\tau)/N\}.$$

Using standard decomposition, the failure probability of the first-generation model can be split as:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_1^+] + \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+].$$

The first term is controlled via a Monte Carlo coverage argument and a TV distance bound between $\mathcal{G}_1$ and the empirical verification distribution $\mathcal{G}_{1,N}^*$, while the second is bounded using the expected inverse coverage. This yields:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \frac{4}{\tau} \cdot D_{\text{TV}}(\mathcal{G}_1, \mathcal{G}_{1,N}^*) + \frac{C_\gamma}{N} \log \frac{2}{\tau},$$

with coverage coefficient $C_\gamma := \mathbb{E}_x [1/\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)]$. This completes the base case.

Step 2: Recursive Coverage Propagation and Error Control. To extend the analysis across generations, we study how coverage evolves recursively. Specifically, the coverage coefficient at generation t satisfies the recurrence:

$$C_{t+1} \lesssim \beta + \alpha C_t, \quad \text{with} \quad \alpha = \frac{1}{NC_0} \log \left(\frac{1}{C_0} \right), \quad \beta = \frac{1}{N} + \frac{D_{\text{TV}}}{NC_0} \log \left(\frac{1}{D_{\text{TV}}} \right),$$

where C_0 is the minimum confidence. Solving this recurrence gives:

$$C_t \lesssim \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \cdot \frac{C_\gamma}{N} + \frac{1 + \frac{A}{C_0} \log \left(\frac{1}{A} \right)}{N - \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)}.$$

Substituting this into the error bound completes the proof.

Key Insight: Recursive Coverage Coefficient Control. Unlike classical generalization bounds, which depend solely on data size and model capacity, recursive synthetic training introduces a new challenge: maintaining high-quality supervision under model-generated data. This requires bounding the evolution of coverage across generations: $C_{\gamma,t} := \mathbb{E}_x [1/\mathcal{G}_t(\mathbf{y}_\gamma^+(x) | x)]$, which quantifies the expected inverse confidence over the verified candidate region. Intuitively, a smaller $C_{\gamma,t}$ implies that the model assigns higher probability mass to its own verified predictions, indicating better self-consistency and internal alignment.

5 Application to Transformer-based LLMs

This section instantiates our theoretical framework on Transformer-based LLMs, which form the foundation of modern commercial systems such as GPT-4 [Achiam et al., 2023], Claude [Anthropic, 2024], LLaMA [Touvron et al., 2023], and Gemini [Team et al., 2023]. Our formalization considers token sequences $S_t = (x_1, \dots, x_t)$ within a structured space $\mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$, where $\mathcal{X}$ represents the token vocabulary and ℓ denotes maximum sequence length. The family of distributions induced by the model is denoted as $\{\mathcal{G}_\theta^{\text{LLM}} \mid \theta \in \Theta_{\text{LLM}}\}$, where θ represents the model parameters. Then, the initial model $\mathcal{G}_{\theta_0}^{\text{LLM}}$ is trained on a corpus of real token sequences $\{(x_1^{(j)}, \dots, x_\ell^{(j)})\}_{j=1}^m$ using maximum likelihood estimation:

$$\theta_0 = \arg \max_{\theta \in \Theta_{\text{LLM}}} \frac{1}{n} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \mathcal{G}_\theta^{\text{LLM}}(x_{t+1}^{(j)} | S_t^{(j)}),$$

where $S_t^{(j)} = (x_1^{(j)}, \dots, x_t^{(j)}) \in \mathcal{S}$ represents each prefix sequence, and $n = m\ell$ is the total sample count. This initial model then participates in recursive training as described in Section 3, using the verification score function defined as $s_{\text{verify}}(x | S) := \log \mathcal{G}_{\theta_0}^{\text{LLM}}(x | S)$. Then, We adopt the standard Transformer architecture [Vaswani et al., 2017, Brown, 2020, Zhang et al., 2025, Hu et al., 2024], consisting of D layers with multi-head self-attention (MHA) and feed-forward network (FFN) modules. For input matrix $X^{(0)} \in \mathbb{R}^{L \times d}$, computation in layer $t \in [D]$ follows:

$$Y^{(t)} = \Pi_{\text{norm}} \left[\text{MHA}(X^{(t-1)}, W^{(t)}) + \gamma_1^{(t)} X^{(t-1)} \right], \quad X^{(t)} = \Pi_{\text{norm}} \left[\text{FFN}(Y^{(t)}, A^{(t)}) + \gamma_2^{(t)} Y^{(t)} \right],$$

where $\gamma_1^{(t)}, \gamma_2^{(t)} \in [-1, 1]$ are residual weights, and Π_{norm} normalizes rows to the unit ℓ_2 -ball. The FFN module is defined as $\text{FFN}(Y^{(t)}, A^{(t)}) = \text{ReLU}(Y^{(t)} A_1^{(t)}) A_2^{(t)}$, with $A_1^{(t)} \in \mathbb{R}^{d \times d_F}$ and $A_2^{(t)} \in \mathbb{R}^{d_F \times d}$. Token prediction logits are produced via $Y^{(D+1)} = \text{softmax}(I_L^\top X^{(D)} A^{(D+1)} / (L\beta))$, where $I_L \in \mathbb{R}^L$ is the all-ones vector, $A^{(D+1)} \in \mathbb{R}^{d \times d_y}$ is the output projection, $\beta \in (0, 1]$ controls temperature, and d_y is the output dimension.

For theoretical tractability, we assume uniformly bounded model parameters, as detailed in Appendix H. This standard assumption in Transformer analysis aligns with prior work [Li et al., 2023, Hu et al., 2024, Zhang et al., 2025, Fu et al., 2025]. In Appendix I, Theorem 7 demonstrates how naive recursive LLM training leads to exponential error growth, while Theorem 8 in Appendix J shows how incorporating real data prevents this collapse. We now present our key contribution: error bounds for LLMs with self-verification that maintain training stability without external real data.

Theorem 4 (Error Bound for Transformer-based LLMs with Self-Verification). *Let $\mathcal{G}_{\theta_t}^{LLM}(x \mid S)$ denote the conditional generation distribution of a Transformer-based LLM at the t -th recursive generation, where $S \in \mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$ is a token prefix sequence and $x \in \mathcal{X}$ is the next token. Then, after T recursive updates, the final prediction error is bounded with probability at least $1 - \delta$:*

$$\text{Err}(\mathcal{G}_{\theta_T}^{LLM}) \lesssim \epsilon(\gamma) + \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \frac{1}{\tau} \right) \frac{\log(n\tilde{B})\tilde{D}}{\sqrt{n}} \log \frac{1}{\delta} + \left(\frac{1}{N} + \left(\frac{1}{NC_0} \log \frac{1}{C_0} \right)^T \frac{C_\gamma}{N} \right) \log \frac{1}{\tau}.$$

Furthermore, if $N > \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)$, the exponential term vanishes as $T \rightarrow \infty$, yielding:

$$\text{Err}(\mathcal{G}_{\theta_T}^{LLM}) \lesssim \epsilon(\gamma) + \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \frac{1}{\tau} \right) \frac{\log(n\tilde{B})\tilde{D}}{\sqrt{n}} \log \frac{1}{\delta} + \frac{1}{N} \log \frac{1}{\tau},$$

where $n = m\ell$ is the total number of training samples, and $\tilde{D}, \tilde{B}$ reflect the model capacity and prior radius respectively. A complete version of the theorem is presented as Theorem 5 in Appendix H.

Remark 7. Stabilizing Recursive Training of LLMs via Verification Mechanisms. This theorem establishes that self-verification mechanisms, when incorporated into Transformer-based LLMs, effectively eliminate the exponential error accumulation typically observed in recursive synthetic training loops. Through rigorous analysis of the error bound, we identify three principal components governing model performance: (i) the confidence-calibrated agreement error $\epsilon(\gamma)$, which can be minimized when the verification slack parameter γ is sufficiently small, leveraging established confidence calibration methods [Liu et al., 2025]; (ii) the statistical generalization gap, characterized by $\log(n\tilde{B})\tilde{D}/\sqrt{n}$, where $\tilde{B}$ and $\tilde{D}$ represent the prior radius and effective dimensionality of the model capacity—parameters that directly scale with the architectural complexity of Transformer networks; and (iii) a recursive decay term determined by verification sampling intensity.

Notably, when the verification sample size exceeds the critical threshold $N > \log(1/C_0)/C_0$, the decay term asymptotically vanishes, enabling stable convergence of the training process regardless of recursion depth. This result not only provides theoretical guarantees for the stability of recursive training language models but also offers practical guidance for calibrating verification intensity based on architectural complexity and confidence characteristics of specific Transformer implementations.

6 Conclusion

This paper establishes a theoretical foundation for understanding and mitigating model collapse in recursive synthetic training. We demonstrate that naive recursive training leads to exponential error growth, whereas incorporating a sufficient fraction of real data can stabilize performance. To address scenarios where real data is scarce or unavailable, we propose a self-verification mechanism that enables models to filter their own outputs based on internal confidence scores. Our analysis provides finite-sample error bounds showing that self-verification alone is sufficient to prevent collapse. We further extend these guarantees to Transformer-based LLMs, offering the first formal justification for stable, fully synthetic training in high-capacity generative models. Overall, our framework paves the way for principled and scalable self-evolution in future generations of LLMs without reliance on external supervision.

Acknowledgement

This project is supported by the National Research Foundation, Singapore, under its NRF Professorship Award No. NRF-P2024-001.

References

- J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- S. Alemohammad, J. Casco-Rodriguez, L. Luzi, A. I. Humayun, H. Babaei, D. LeJeune, A. Siahkoohi, and R. Baraniuk. Self-consuming generative models go mad. In *The Twelfth International Conference on Learning Representations*, 2024a.
- S. Alemohammad, A. I. Humayun, S. Agarwal, J. Collomosse, and R. Baraniuk. Self-improving diffusion models with synthetic data. *arXiv preprint arXiv:2408.16333*, 2024b.
- A. Anthropic. The claude 3 model family: Opus, sonnet, haiku. *Claude-3 Model Card*, 1:1, 2024.
- Q. Bertrand, J. Bose, A. Duplessis, M. Jiralerspong, and G. Gidel. On the stability of iterative retraining of generative models on their own data. In *The Twelfth International Conference on Learning Representations*, 2024.
- M. Bohacek and H. Farid. Nepotistically trained generative-ai models collapse. *arXiv e-prints*, pages arXiv-2311, 2023.
- M. Briesch, D. Sobania, and F. Rothlauf. Large language models suffer from their own output: An analysis of the self-consuming training loop. *arXiv preprint arXiv:2311.16822*, 2023.
- T. B. Brown. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.
- A. Dey and D. Donoho. Universality of the $\pi^2/6$ pathway in avoiding model collapse. *arXiv preprint arXiv:2410.22812*, 2024.
- E. Dohmatob, Y. Feng, and J. Kempe. Model collapse demystified: The case of regression. *arXiv preprint arXiv:2402.07712*, 2024a.
- E. Dohmatob, Y. Feng, A. Subramonian, and J. Kempe. Strong model collapse. *arXiv preprint arXiv:2410.04840*, 2024b.
- E. Dohmatob, Y. Feng, P. Yang, F. Charton, and J. Kempe. A tale of tails: Model collapse as a change of scaling laws. In *Forty-first International Conference on Machine Learning*, 2024c.
- Y. Feng, E. Dohmatob, P. Yang, F. Charton, and J. Kempe. Beyond model collapse: Scaling up with synthesized data requires verification. In *The Thirteenth International Conference on Learning Representations*, 2025.
- A. E. Firdoussi, M. E. A. Seddik, S. Hayou, R. ALAMI, A. Alzubaidi, and H. Hacid. Maximizing the potential of synthetic data: Insights from random matrix theory. In *The Thirteenth International Conference on Learning Representations*, 2025.
- S. Fu, Y. Chen, Y. Wang, and D. Tao. A theoretical survey on foundation models. *arXiv preprint arXiv:2410.11444*, 2024a.
- S. Fu, S. Zhang, Y. Wang, X. Tian, and D. Tao. Towards theoretical understandings of self-consuming generative models. In *Forty-first International Conference on Machine Learning*, 2024b.
- S. Fu, Y. Wang, Y. Chen, X. Tian, and D. Tao. A theoretical perspective: How to prevent model collapse in self-consuming training loops. In *The Thirteenth International Conference on Learning Representations*, 2025.
- M. Gerstgrasser, R. Schaeffer, A. Dey, R. Rafailov, T. Korbak, H. Sleight, R. Agrawal, J. Hughes, D. B. Pai, A. Gromov, D. Roberts, D. Yang, D. L. Donoho, and S. Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*, 2024.
- E. Gibney. Ai models fed ai-generated data quickly spew nonsense. *Nature*, 632(8023):18–19, 2024.
- N. Gillman, M. Freeman, D. Aggarwal, H. Chia-Hong, C. Luo, Y. Tian, and C. Sun. Self-correcting self-consuming loops for generative model training. In *Forty-first International Conference on Machine Learning*, 2024.
- Y. Guo, G. Shang, M. Vazirgiannis, and C. Clavel. The curious decline of linguistic diversity: Training language models on synthetic text. *arXiv preprint arXiv:2311.09807*, 2023.
- X. Hu, F. Zhang, S. Chen, and Z. Yang. Unveiling the statistical foundations of chain-of-thought prompting methods. *arXiv preprint arXiv:2408.14511*, 2024.

- A. Huang, A. Block, D. J. Foster, D. Rohatgi, C. Zhang, M. Simchowitz, J. T. Ash, and A. Krishnamurthy. Self-improvement in language models: The sharpening mechanism. In *The Thirteenth International Conference on Learning Representations*, 2025a.
- C. Huang, Z. Fan, L. Wang, F. Yang, P. Zhao, Z. Lin, Q. Lin, D. Zhang, S. Rajmohan, and Q. Zhang. Self-evolved reward learning for llms. In *The Thirteenth International Conference on Learning Representations*, 2025b.
- J. Huang, X. Chen, S. Mishra, H. S. Zheng, A. W. Yu, X. Song, and D. Zhou. Large language models cannot self-correct reasoning yet. *arXiv preprint arXiv:2310.01798*, 2023.
- A. Jain, A. Montanari, and E. Sasoglu. Scaling laws for learning with real and surrogate data. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- H. Jiang, Y. Ma, C. Ding, K. Luan, and X. Di. Towards intrinsic self-correction enhancement in monte carlo tree search boosted reasoning via iterative preference learning. *arXiv preprint arXiv:2412.17397*, 2024.
- R. Kamoi, Y. Zhang, N. Zhang, J. Han, and R. Zhang. When can llms actually correct their own mistakes? a critical survey of self-correction of llms. *Transactions of the Association for Computational Linguistics*, 12:1417–1440, 2024.
- M. Kanabar and M. Gastpar. Minimax discrete distribution estimation with self-consumption. *arXiv preprint arXiv:2501.19273*, 2025.
- J. Kazdan, R. Schaeffer, A. Dey, M. Gerstgrasser, R. Rafailov, D. L. Donoho, and S. Koyejo. Collapse or thrive? perils and promises of synthetic data in a self-generating world. *arXiv preprint arXiv:2410.16713*, 2024.
- A. Kumar, V. Zhuang, R. Agarwal, Y. Su, J. D. Co-Reyes, A. Singh, K. Baumli, S. Iqbal, C. Bishop, R. Roelofs, et al. Training language models to self-correct via reinforcement learning. *arXiv preprint arXiv:2409.12917*, 2024.
- Y. Li, M. E. Ildiz, D. Papailiopoulos, and S. Oymak. Transformers as algorithms: Generalization and stability in in-context learning. In *International Conference on Machine Learning*, pages 19565–19594. PMLR, 2023.
- T. Liang. How well generative adversarial networks learn distributions. *Journal of Machine Learning Research*, 22(228):1–41, 2021.
- H. Lightman, V. Kosaraju, Y. Burda, H. Edwards, B. Baker, T. Lee, J. Leike, J. Schulman, I. Sutskever, and K. Cobbe. Let’s verify step by step. In *The Twelfth International Conference on Learning Representations*, 2023.
- X. Liu, T. Chen, L. Da, C. Chen, Z. Lin, and H. Wei. Uncertainty quantification and confidence calibration in large language models: A survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 6107–6117, 2025.
- B. Luo, S. Wang, Y. Li, and H. Wei. Your pre-trained llm is secretly an unsupervised confidence calibrator. *arXiv preprint arXiv:2505.16690*, 2025.
- R. Ma, P. Wang, C. Liu, X. Liu, J. Chen, B. Zhang, X. Zhou, N. Du, and J. Li. s^2 r: Teaching llms to self-verify and self-correct via reinforcement learning. *arXiv preprint arXiv:2502.12853*, 2025.
- M. Marchi, S. Soatto, P. Chaudhari, and P. Tabuada. Heat death of generative models in closed-loop learning. *arXiv preprint arXiv:2404.02325*, 2024.
- G. Martínez, L. Watson, P. Reviriego, J. A. Hernández, M. Juárez, and R. Sarkar. Towards understanding the interplay of generative artificial intelligence and the internet. In *International Workshop on Epistemic Uncertainty in Artificial Intelligence*, pages 59–73. Springer, 2023.
- A. Mehrtash, W. M. Wells, C. M. Tempny, P. Abolmaesumi, and T. Kapur. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE transactions on medical imaging*, 39(12):3868–3878, 2020.
- V. Sadasivan, A. Kumar, S. Balasubramanian, W. Wang, and S. Feizi. Can ai-generated text be reliably detected? *arXiv preprint arXiv:2303.11156*, 2023.
- W. Saunders, C. Yeh, J. Wu, S. Bills, L. Ouyang, J. Ward, and J. Leike. Self-critiquing models for assisting human evaluators. *arXiv preprint arXiv:2206.05802*, 2022.
- R. Schaeffer, J. Kazdan, A. C. Arulandu, and S. Koyejo. Position: Model collapse does not mean what you think. *arXiv preprint arXiv:2503.03150*, 2025.
- C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35:25278–25294, 2022.

- M. E. A. Seddik, S.-W. Chen, S. Hayou, P. Youssef, and M. Debbah. How bad is training on synthetic data? a statistical analysis of language model collapse. *arXiv preprint arXiv:2404.05090*, 2024.
- I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal. Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.
- K. Stechly, K. Valmeekam, and S. Kambhampati. On the self-verification limitations of large language models on reasoning and planning tasks. *arXiv preprint arXiv:2402.08115*, 2024.
- A. T. Suresh, A. Thangaraj, and A. N. K. Khandavally. Rate of model collapse in recursive training. *arXiv preprint arXiv:2412.17646*, 2024.
- Z. Tao, T.-E. Lin, X. Chen, H. Li, Y. Wu, Y. Li, Z. Jin, F. Huang, D. Tao, and J. Zhou. A survey on self-evolution of large language models. *arXiv preprint arXiv:2404.14387*, 2024.
- G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- G. Tyen, H. Mansoor, V. Cărbune, P. Chen, and T. Mak. Llms cannot find reasoning errors, but can correct them given the error location. *arXiv preprint arXiv:2311.08516*, 2023.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- P. Villalobos, J. Sevilla, L. Heim, T. Besiroglu, M. Hobbhahn, and A. Ho. Will we run out of data? an analysis of the limits of scaling datasets in machine learning. *arXiv preprint arXiv:2211.04325*, 2022.
- R. Wang, J. Guo, C. Gao, G. Fan, C. Y. Chong, and X. Xia. Can llms replace human evaluators? an empirical study of llm-as-a-judge in software engineering. *arXiv preprint arXiv:2502.06193*, 2025.
- X. Wang, H. Liu, C. Shi, and C. Yang. Be confident! towards trustworthy graph neural networks via confidence calibration. *Advances in Neural Information Processing Systems*, 34:23768–23779, 2021.
- Y. Weng, M. Zhu, F. Xia, B. Li, S. He, S. Liu, B. Sun, K. Liu, and J. Zhao. Large language models are better reasoners with self-verification. In *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.
- S. Wyllie, I. Shumailov, and N. Papernot. Fairness feedback loops: training on synthetic data amplifies bias. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2113–2147, 2024.
- X. Xing, F. Shi, J. Huang, Y. Wu, Y. Nan, S. Zhang, Y. Fang, M. Roberts, C.-B. Schönlieb, J. Del Ser, et al. On the caveats of ai autophagy. *Nature Machine Intelligence*, pages 1–9, 2025.
- X. Xu, S. Diao, C. Yang, and Y. Wang. Can we verify step by step for incorrect answer detection? *arXiv preprint arXiv:2402.10528*, 2024.
- Q. Zhang, H. Qiu, D. Wang, H. Qian, Y. Li, T. Zhang, and M. Huang. Understanding the dark side of llms’ intrinsic self-correction. *arXiv preprint arXiv:2412.14959*, 2024a.
- Y. Zhang, F. Zhang, Z. Yang, and Z. Wang. What and how does in-context learning learn? bayesian model averaging, parameterization, and generalization. *arXiv preprint arXiv:2305.19420*, 2023.
- Y. Zhang, M. Khalifa, L. Logeswaran, J. Kim, M. Lee, H. Lee, and L. Wang. Small language models need strong verifiers to self-correct reasoning. *arXiv preprint arXiv:2404.17140*, 2024b.
- Y. Zhang, F. Zhang, Z. Yang, and Z. Wang. What and how does in-context learning learn? bayesian model averaging, parameterization, and generalization. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.
- F. Zhu, Z. Cheng, X.-Y. Zhang, and C.-L. Liu. Rethinking confidence calibration for failure prediction. In *European conference on computer vision*, pages 518–536. Springer, 2022.
- X. Zhu, D. Cheng, H. Li, K. Zhang, E. Hua, X. Lv, N. Ding, Z. Lin, Z. Zheng, and B. Zhou. How to synthesize text data without model collapse? *arXiv preprint arXiv:2412.14689*, 2024.

Appendix

A Overview

In this supplementary material, we provide additional details and complete proofs to support the theoretical developments presented in the main paper. The appendix is organized as follows:

- **Appendix B. Related Work: Self-Verification in LLMs.**
We situate our work within the emerging literature on self-verification for LLMs, highlighting key differences between our theoretically grounded, training-time framework and prior prompt-based or post hoc empirical approaches.
- **Appendix C. Limitations.**
We discuss a primary limitation of our framework—its reliance on a single verification strategy based on internal confidence scores—and outline possible extensions using richer or more adaptive verification methods.
- **Appendix D. Broader Impacts.**
We outline the potential positive societal implications of our work, such as reducing dependence on human-labeled data through self-verification, and note that our theoretical analysis poses no immediate negative societal risks.
- **Appendix E. Proof of Theorem 3: Error Bound with Self-Verification.**
We present a detailed finite-sample analysis of recursive training with self-verification. The proof introduces recursive coverage coefficients to quantify verification quality and demonstrates how confidence-guided filtering prevents error amplification across generations.
- **Appendix F. Proof of Theorem 1: Error Amplification in Naive Recursive Training.**
We formalize the failure mode of naive recursive training without any verification and prove that, under mild conditions, prediction error grows exponentially with the number of generations.
- **Appendix G. Proof of Theorem 2: Stabilization via Real Data Injection.**
We analyze how introducing a fixed proportion of real data at each generation mitigates exponential error growth, and derive sufficient conditions under which training remains stable.
- **Appendix H. Proof of Theorem 4: Theoretical Guarantees for Transformer-Based LLMs with Self-Verification.**
We extend our theoretical framework to Transformer-based language models, providing generalization bounds under recursive training with self-verification, accounting for model capacity and confidence calibration.
- **Appendix I. Exponential Error Growth in Naive Recursive Training of Transformer-Based LLMs.**
We instantiate Theorem 1 in the Transformer setting and analyze how parameter confidence bounds affect the rate of error amplification during unverified recursive training.
- **Appendix J. Mitigating Collapse via Real Data in Recursive Training of Transformer-Based LLMs.**
We apply Theorem 2 to the Transformer case, showing that even limited real data injection suffices to stabilize recursive training and suppress error accumulation in practical scenarios.
- **Appendix K. Additional Experimental Analysis and Future Work.**
We provide further empirical validation for our assumptions, present experimental results demonstrating how self-verification prevents model collapse, and discuss potential directions for future work.

This appendix provides the full mathematical foundations of our results and demonstrates their applicability to modern high-capacity models used in practice.

B Related Work: Self-Verification in LLMs

Recent research has explored self-verification as a means of improving LLM reasoning performance during inference [Weng et al., 2023, Stechly et al., 2024, Ma et al., 2025, Lightman et al., 2023, Xu et al., 2024, Kamoi et al., 2024, Tao et al., 2024]. Most existing approaches adopt a prompt-based paradigm, where LLMs are instructed at inference time to critique or revise their own outputs [Huang et al., 2023, Tyen et al., 2023, Zhang et al., 2024a], or are trained post hoc to mimic such behavior via curated demonstrations or reinforcement learning [Saunders et al., 2022, Jiang et al., 2024, Kumar et al., 2024, Zhang et al., 2024b, Ma et al., 2025]. While effective in certain settings, these methods largely focus on empirical improvements, and rely on externally crafted prompts or behavior cloning pipelines.

In contrast to these empirical efforts, our work develops a formal theoretical framework for self-verification in recursive training settings. Rather than prompting models to verify at inference time, we integrate self-verification directly into the training process via internal confidence-based filtering. This allows us to establish provable error bounds, showing that self-verification alone can prevent collapse under fully synthetic training. Furthermore, we extend these guarantees to Transformer-based LLMs, providing the first formal justification of stable self-evolution in high-capacity models without external supervision. Our contribution bridges a gap in the literature by grounding the concept of self-verification in statistical learning theory, rather than relying on prompt design or behavioral imitation.

C Limitations

A primary limitation of our framework lies in its reliance on a single verification strategy—confidence-based filtering using a fixed base model. While effective in our theoretical setting, this approach may be insufficient to capture semantic correctness in more complex tasks. Future work should explore richer and more adaptive verification methods, such as ensemble-based voting or task-specific evaluators, to enhance robustness.

D Broader Impacts

This work provides a theoretical foundation for stable and data-efficient training of language models through self-verification, potentially reducing the reliance on large-scale human-labeled datasets and enabling broader access to capable models. As a theoretical contribution, it does not pose any immediate negative societal impacts.

E Proof of Theorem 3: Error Bound with Self-Verification

In this section, we present the proof of Theorem 3, which provides a finite-sample error bound for recursive training with self-verification. Our analysis draws inspiration from the confidence framework introduced in Huang et al. [2025a], but extends it to a more general recursive setting that captures the dynamics of synthetic supervision over multiple generations.

The key insight lies in controlling the evolution of the *recursive coverage coefficient*, which measures how well the model concentrates probability mass on its own verified predictions across generations. While classical generalization bounds typically depend only on data size and model capacity, recursive training introduces a new challenge: maintaining reliable supervision when the training data is generated by the model itself. To formalize this, we define the recursive coverage coefficient as

$$C_{\gamma,t} := \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_t(\mathbf{y}_{\gamma}^+(x) \mid x)} \right],$$

which captures the expected inverse confidence over the high-quality, self-verified region. A smaller value of $C_{\gamma,t}$ implies that the model assigns higher likelihood to its verified outputs, indicating better internal consistency and semantic alignment.

By bounding the growth of this coefficient over generations, we are able to characterize how self-verification mitigates error amplification and ensures stability, even under fully synthetic training.

The proof proceeds by decomposing failure events and recursively controlling the probability mass over verified subsets.

Proof of Theorem 3. We begin by recalling that the initial model $\mathcal{G}_0$ is trained via maximum likelihood estimation on a ground-truth dataset $S_0 = \{(x_j, y_j^*)\}_{j=1}^n$, that is,

$$\mathcal{G}_0 = \arg \max_{\mathcal{G} \in \mathcal{H}} \sum_{j=1}^n \log \mathcal{G}(y_j^* | x_j),$$

where $\mathcal{H}$ denotes the model hypothesis class. The resulting model $\mathcal{G}_0$ encodes a set of inductive biases aligned with human-annotated supervision, and serves as a fixed and calibrated reference model in the recursive training pipeline.

Defining Confident Support Set $\mathcal{D}_1^+$

We define the subset of inputs with sufficient sample coverage as

$$\mathcal{D}_1^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) \geq \frac{\log \frac{2}{\tau}}{N} \right\},$$

where the high-quality candidate set $\mathbf{y}_\gamma^+(x)$ is defined based on the internal self-verification score function:

$$\mathbf{y}_\gamma^+ = \left\{ y \mid s_{\text{verify}}(y | x) \geq \max_{y' \in \mathcal{Y}} s_{\text{verify}}(y' | x) - \gamma \right\},$$

with self-verification score function defined as $s_{\text{verify}}(y | x) := \log \mathcal{G}_0(y | x)$. This gives:

$$\begin{aligned} \mathbf{y}_\gamma^+ &= \left\{ y \mid \log \mathcal{G}_0(y | x) \geq \max_{y' \in \mathcal{Y}} \log \mathcal{G}_0(y' | x) - \gamma \right\} \\ &= \left\{ y \mid \mathcal{G}_0(y | x) \geq \max_{y' \in \mathcal{Y}} \gamma' \mathcal{G}_0(y' | x) \right\}, \end{aligned} \quad (3)$$

where $\gamma' = e^{-\gamma}$.

Decomposing the Probability of Low Confidence Generation

We aim to bound:

$$\begin{aligned} &\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \\ &\leq \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_1^+] + \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+]. \end{aligned} \quad (4)$$

Bounding the First Term: Using Monte Carlo Estimate

For $x \in \mathcal{D}_1^+$, we aim to show

$$\mathcal{G}_{1,N}^*(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau/2.$$

Indeed, observe that $y \sim \mathcal{G}_{1,N}^*(\cdot | x) \notin \mathbf{y}_\gamma^+(x)$ if and only if $y_{1,1}, \dots, y_{1,N} \sim \mathcal{G}_0(\cdot)$ have $y_{1,j} \notin \mathbf{y}_\gamma^+(x)$ for all j , since $x \in \mathcal{D}_1^+$ which happens with probability

$$(1 - \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x))^N \leq (1 - \frac{\log \frac{2}{\tau}}{N})^N. \quad (5)$$

Then we show that:

$$(1 - \frac{\log \frac{2}{\tau}}{N})^N \leq \frac{\tau}{2},$$

To do so, we analyze its asymptotic behavior, and verify that it satisfies the above inequality. This is now a typical exponential approximation problem, where the base $(1 - \frac{1}{N})^N$ plays a key role. We notice that:

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = \frac{1}{e},$$

Thus, we get

$$\left(1 - \frac{\log \frac{2}{\tau}}{N}\right)^N = \left[\left(1 - \frac{\log \frac{2}{\tau}}{N}\right)^{\frac{N}{\log \frac{2}{\tau}}}\right]^{\log \frac{2}{\tau}} \leq \left(\frac{1}{e}\right)^{\log \frac{2}{\tau}} = \frac{\tau}{2}. \quad (6)$$

Then, we obtain

$$(1 - \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x))^N \leq \frac{\tau}{2},$$

Thus, we get

$$\mathcal{G}_{1,N}^*(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau/2.$$

Bounding Discrepancy via Total Variation Distance

The total variation (TV) distance between two probability measures P and Q is defined as

$$D_{\text{TV}}(P, Q) = \frac{1}{2} \sum_y |P(y) - Q(y)|.$$

Importantly, the TV distance provides a direct measure of the pointwise probability differences between P and Q . In particular, for any measurable event A , the following inequality holds:

$$|P(A) - Q(A)| \leq 2D_{\text{TV}}(P, Q).$$

Specializing to singleton events $\{y\}$, it follows that

$$|\mathcal{G}_1(y | x) - \mathcal{G}_{1,N}^*(y | x)| \leq 2D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)).$$

Equivalently, we can lower bound the TV distance by

$$D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)) \geq \frac{1}{2} |\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) - \mathcal{G}_{1,N}^*(\mathbf{y}_\gamma^+(x) | x)|.$$

From the definition of the set $\mathcal{D}_1^+$, we know that for any $x \in \mathcal{D}_1^+$,

$$\mathcal{G}_{1,N}^*(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \frac{\tau}{2}.$$

Suppose that $\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau$. Then it follows that

$$|\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) - \mathcal{G}_{1,N}^*(\mathbf{y}_\gamma^+(x) | x)| \geq \tau - \frac{\tau}{2} = \frac{\tau}{2}.$$

Substituting back into the previous inequality, we conclude that for all $x \in \mathcal{D}_1^+$,

$$D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)) \geq \frac{\tau}{4} \cdot \mathbb{I}\{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau\}.$$

Thus, we get:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_1^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)) \quad (7)$$

Bounding the Second Term: Using Markov's Inequality

Let $C_\gamma := \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} \right]$. Then:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+] &= \mathbb{P}_{x \sim \mathcal{X}} \left[\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) < \frac{\log \frac{2}{\tau}}{N} \right] \\ &= \mathbb{P}_{x \sim \mathcal{X}} \left[\frac{\log \frac{2}{\tau}}{N \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} > 1 \right] \\ &\leq \frac{\log \frac{2}{\tau}}{N} \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} \right] \\ &\leq \frac{C_\gamma}{N} \log \frac{2}{\tau}. \end{aligned} \quad (8)$$

Final Bound for $\mathcal{G}_1$

Substituting equalities 7 and 8 into 4, we obtain

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)) + \frac{C_\gamma}{N} \log \frac{2}{\tau}. \quad (9)$$

Recursive Bound for the Second Generation Model $\mathcal{G}_2$

Building upon the prior analysis, we now investigate how a self-verification mechanism can be systematically employed to refine model training across generations. Specifically, we focus on filtering the outputs generated by $\mathcal{G}_1$ through confidence-aware criteria, and subsequently utilizing only the verified outputs to train the next-generation model $\mathcal{G}_2$. This recursive training framework ensures that each successive model benefits from high-quality and semantically coherent supervision.

Definition of Verified Input Set for $\mathcal{G}_2$

We define the subset of inputs with sufficient verified coverage as:

$$\mathcal{D}_2^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \geq \frac{\log \frac{2}{\tau}}{N} \right\},$$

where the high-quality candidate set $\mathbf{y}_\gamma^+(x)$ is defined based on the internal verification score function:

$$\mathbf{y}_\gamma^+ = \left\{ y \mid s_{\text{verify}}(y | x) \geq \max_{y' \in \mathcal{Y}} s_{\text{verify}}(y' | x) - \gamma \right\}.$$

Bounding the Low-Confidence Region for $\mathcal{G}_2$

Following the same logic as in Eq. (7), we derive the following bound on the probability that $\mathcal{G}_2$ assigns low confidence to the verified candidate region:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_2^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_{2,N}^*(\cdot | x)) \quad (10)$$

Bounding the Failure Rate of Verified Samples

Similarly, we follow the derivation of Eq. (8) to obtain an upper bound on the probability that an input falls outside the region $\mathcal{D}_2^+$:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_2^+] &= \mathbb{P}_{x \sim \mathcal{X}} \left[\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) < \frac{\log \frac{2}{\tau}}{N} \right] \\ &= \mathbb{P}_{x \sim \mathcal{X}} \left[\frac{\log \frac{2}{\tau}}{N \mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} > 1 \right] \\ &\leq \frac{\log \frac{2}{\tau}}{N} \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right] \end{aligned} \quad (11)$$

The quantity $\mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]$ captures the expected inverse confidence in the high-quality region across all inputs, and plays a central role in the recursion.

To proceed, we introduce the variable:

$$Z(x) = \frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)}.$$

Its expectation is computed via the tail-integral identity for nonnegative random variables:

$$\mathbb{E}_{x \sim \mathcal{X}} [Z(x)] = \int_0^\infty \mathbb{P}(Z(x) \geq t) dt. \quad (12)$$

Since the posterior satisfies $C_0 \leq \mathcal{G}_1(\mathbf{y}_\gamma^+(x) \mid x) \leq 1$, it follows that

$$1 \leq Z(x) = \frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) \mid x)} \leq \frac{1}{C_0}.$$

Hence,

$$\mathbb{P}(Z(x) \geq t) = \begin{cases} 1, & 0 \leq t < 1, \\ 0, & t > \frac{1}{C_0}. \end{cases}$$

Thus the integral in (12) reduces to

$$\begin{aligned} \mathbb{E}[Z(x)] &= \int_0^1 1 \, dt + \int_1^{\frac{1}{C_0}} \mathbb{P}(Z(x) \geq t) \, dt \\ &= 1 + \int_1^{\frac{1}{C_0}} \mathbb{P}(Z(x) \geq t) \, dt. \end{aligned}$$

Recall that Eq (9), for $t \geq 1$, then we can deduce that:

$$\mathbb{P}_{x \sim \mathcal{X}}[Z(x) \geq t] \leq \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_{1,N}^*(\cdot \mid x)) + \frac{C_\gamma}{N} \log \frac{2t}{t-1}$$

The rest of the proof proceeds by upper bounding this integral through a sharp relaxation using $B(t)$, characterizing tail probabilities of inverse confidence values, and leading to a recursive generalization bound.

Explicit Upper Bound on Recursive Confidence

We define:

$$B(t) = \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_{1,N}^*(\cdot \mid x)) + \frac{C_\gamma}{N} \log \frac{2t}{t-1}.$$

Then we Upper bound for the tail probability via $\min\{1, B(t)\}$. However, probabilities cannot exceed 1. Hence a tighter statement is

$$\mathbb{P}(Z(x) \geq t) \leq \min\{1, B(t)\}.$$

and therefore

$$\mathbb{E}[Z(x)] = 1 + \int_1^{\frac{1}{C_0}} \mathbb{P}(Z(x) \geq t) \, dt \leq 1 + \int_1^{\frac{1}{C_0}} \min\{1, B(t)\} \, dt.$$

Since $B(t)$ is monotonically decreasing for $t \in (1^+, +\infty)$, we focus on the non-trivial case where $B\left(\frac{1}{C_0}\right) < 1$. In this setting, there exists a *unique* point $t_0 \in \left(1, \frac{1}{C_0}\right)$ such that

$$B(t_0) = 1 \quad \text{and} \quad B(t) \begin{cases} \geq 1, & 1 \leq t \leq t_0, \\ \leq 1, & t_0 \leq t \leq \frac{1}{C_0}. \end{cases}$$

Consequently,

$$\min\{1, B(t)\} = \begin{cases} 1, & t \in [1, t_0], \\ B(t), & t \in [t_0, 1/C_0]. \end{cases}$$

Hence,

$$\int_1^{\frac{1}{C_0}} \min\{1, B(t)\} \, dt = \int_1^{t_0} 1 \, dt + \int_{t_0}^{\frac{1}{C_0}} B(t) \, dt = (t_0 - 1) + \int_{t_0}^{\frac{1}{C_0}} B(t) \, dt. \quad (13)$$

Then, we compute $\int B(t) \, dt$ explicitly. Let us denote

$$B(t) = A \frac{t}{t-1} + C \log\left(\frac{2t}{t-1}\right),$$

where

$$A = 4D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_{1,N}^*(\cdot | x)), \quad C = \frac{C_\gamma}{N}.$$

Then we decompose and integrate each part:

$$\begin{aligned} \int \frac{t}{t-1} dt &= \int \left(1 + \frac{1}{t-1}\right) dt = t + \ln|t-1| + \text{constant}, \\ \int \log\left(\frac{2t}{t-1}\right) dt &= \int \log(2t) dt - \int \log(t-1) dt \\ &= \left[t \log(2t) - t\right] - \left[(t-1) \log(t-1) - (t-1)\right] + \text{constant} \\ &= t \log(2t) - (t-1) \log(t-1) + \text{constant}. \end{aligned}$$

Therefore,

$$\int B(t) dt = A \left[t + \ln|t-1|\right] + C \left[t \log(2t) - (t-1) \log(t-1)\right] + \text{constant}.$$

We shall denote a particular antiderivative by $F(t)$. Then

$$\int_{t_0}^{\frac{1}{C_0}} B(t) dt = F\left(\frac{1}{C_0}\right) - F(t_0),$$

where t_0 is the unique solution to $B(t_0) = 1$. Then we conclude the final upper bound. Combining (13) with the preceding results, we see that whenever $B(\frac{1}{C_0}) < 1$, we obtain

$$\int_1^{\frac{1}{C_0}} \min\{1, B(t)\} dt = (t_0 - 1) + \left[F\left(\frac{1}{C_0}\right) - F(t_0)\right],$$

and thus

$$\mathbb{E}[Z(x)] = 1 + \int_1^{\frac{1}{C_0}} P(Z(x) \geq t) dt \leq 1 + (t_0 - 1) + \left[F\left(\frac{1}{C_0}\right) - F(t_0)\right].$$

Recalling that $Z(x) = 1/\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)$, we finally arrive at

$$\mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right] \leq 1 + (t_0 - 1) + F\left(\frac{1}{C_0}\right) - F(t_0),$$

which provides an explicit upper bound on the coverage coefficient in the regime $B(\frac{1}{C_0}) < 1$. Here, $t_0 \in (1, \frac{1}{C_0})$ is the unique point where $B(t_0) = 1$, and $F(\cdot)$ is any antiderivative of $B(\cdot)$.

Then we further derive the explicit coverage coefficient upper bound. First evaluate $F(\frac{1}{C_0}) - F(t_0)$. From the explicit antiderivative, ignoring any additive constant (which cancels in the difference), we get

$$\begin{aligned} F\left(\frac{1}{C_0}\right) - F(t_0) &= A \left[\left(\frac{1}{C_0} - t_0\right) + \ln\left(\frac{\frac{1}{C_0} - 1}{t_0 - 1}\right) \right] \\ &+ C \left[\left(\frac{1}{C_0} \log\left(\frac{2}{C_0}\right) - t_0 \log(2t_0)\right) - \left(\left(\frac{1}{C_0} - 1\right) \log\left(\frac{1}{C_0} - 1\right) - (t_0 - 1) \log(t_0 - 1)\right) \right]. \end{aligned}$$

Thus,

$$(t_0 - 1) + F\left(\frac{1}{C_0}\right) - F(t_0) = (t_0 - 1) + A \left[\frac{1}{C_0} - t_0 + \ln\left(\frac{\frac{1}{C_0} - 1}{t_0 - 1}\right) \right] + C [\dots],$$

where the bracket $[\dots]$ in the C -term is the obvious difference from above.

Then we incorporate the condition $B(t_0) = 1$. Recall

$$B(t_0) = A \frac{t_0}{t_0 - 1} + C \log\left(\frac{2t_0}{t_0 - 1}\right) = 1.$$

This equation confines t_0 to the interval $(1, \frac{1}{C_0})$. Although t_0 generally cannot be expressed in an elementary closed form, we can appropriately scale the equation to obtain an approximate range for t_0 .

Moreover, elementary inequalities for $\log t$ imply

$$\frac{t-1}{t} \leq \log t \leq t-1 \quad \text{for all } t > 0,$$

and setting $t = \frac{2t_0}{t_0-1}$ yields

$$\frac{1}{2} + \frac{1}{2t_0} \leq \log\left(\frac{2t_0}{t_0-1}\right) \leq 1 + \frac{2}{t_0-1}. \quad (14)$$

Hence, from $B(t_0) = 1$ we deduce two bracketing inequalities:

$$A \frac{t_0}{t_0-1} + C \left(1 + \frac{2}{t_0-1}\right) \geq 1, \quad (15)$$

$$A \frac{t_0}{t_0-1} + C \left(\frac{1}{2} + \frac{1}{2t_0}\right) \leq 1. \quad (16)$$

Next, we consider the non-trivial case that aligns more closely with practical scenarios, A and C are small enough that $A + C < 1$. Under that assumption, (15) and (16) is valid and yields an *upper* and *lower* bound on t_0 . Thus

$$\frac{1 + \sqrt{1 - 2C + C^2 + 2AC}}{2 - (2A + C)} \leq t_0 \leq \frac{1 + C}{1 - (A + C)}.$$

Thus,

$$\begin{aligned} t_0 - 1 &\geq \frac{1 + \sqrt{1 - 2C + C^2 + 2AC}}{2 - (2A + C)} - 1 \\ &\geq \frac{\sqrt{1 - 2C + C^2 + 2AC} - (1 - C) + 2A}{2 - (2A + C)} \\ &\geq \frac{2A}{2 - (2A + C)}. \end{aligned} \quad (17)$$

Our goal is to convert these into simpler *numerical* bounds on t_0 , which can then be substituted into the final coverage coefficient bound

$$C_{\gamma,1} = \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) \mid x)} \right] \leq 1 + (t_0 - 1) + [F(\frac{1}{C_0}) - F(t_0)].$$

Thus, we get

$$\begin{aligned} C_{\gamma,1} &= \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) \mid x)} \right] \leq t_0 + [F(\frac{1}{C_0}) - F(t_0)]. \\ &\leq t_0 + A \left[\left(\frac{1}{C_0} - t_0 \right) + \ln \left(\frac{\frac{1}{C_0} - 1}{t_0 - 1} \right) \right] \end{aligned} \quad (18)$$

$$\begin{aligned} &+ C \left[\left(\frac{1}{C_0} \log \left(\frac{2}{C_0} \right) - t_0 \log(2t_0) \right) - \left(\left(\frac{1}{C_0} - 1 \right) \log \left(\frac{1}{C_0} - 1 \right) - (t_0 - 1) \log(t_0 - 1) \right) \right]. \\ &\lesssim t_0 + A \frac{1}{C_0} \log \left(\frac{1}{t_0 - 1} \right) + C \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\ &\lesssim \frac{1 + C}{1 - (A + C)} + A \frac{1}{C_0} \log \left(\frac{2 - (2A + C)}{2A} \right) + C \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\ &\lesssim 1 + C + A \frac{1}{C_0} \log \left(\frac{1}{A} \right) + C \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\ &\lesssim 1 + A \frac{1}{C_0} \log \left(\frac{1}{A} \right) + C \frac{1}{C_0} \log \left(\frac{1}{C_0} \right). \end{aligned} \quad (19)$$

Final Bound for $\mathcal{G}_2$

Combining the bounds from (10) and (11), we obtain:

$$\begin{aligned} & \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \\ & \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_{2,N}^*(\cdot | x)) + \frac{\log \frac{2}{\tau}}{N} \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right] \\ & \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_{2,N}^*(\cdot | x)) + \frac{C_{\gamma,1}}{N} \log \frac{2}{\tau}. \end{aligned} \quad (20)$$

where $C_{\gamma,1}$ is the bound derived above. This completes the recursive error bound analysis for the second generation model.

Recursive Generalization Bounds for Arbitrary Generation t

We now extend our recursive analysis to arbitrary generation steps beyond $t = 2$. To simplify the discussion and maintain consistency, we make the following assumption:

Assumption: Uniform Total Variation Distance

Without loss of generality, we assume that the total variation distance across all generations $1 \leq t \leq T$ remains approximately the same, that is,

$$D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t,N}^*(\cdot | x)) \asymp D_{\text{TV}}.$$

This allows us to write for the second generation:

$$\begin{aligned} A_2 &= D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_{2,N}^*(\cdot | x)) \asymp D_{\text{TV}}, \\ C_2 &= \frac{C_{\gamma,1}}{N} \lesssim \frac{1}{N} + \frac{1}{N} A \frac{1}{C_0} \log \left(\frac{1}{A} \right) + \frac{C}{N} \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right). \end{aligned} \quad (21)$$

Continuing this recursion, we arrive at:

$$C_{\gamma,2} = \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x)} \right] \lesssim 1 + A_2 \frac{1}{C_0} \log \left(\frac{1}{A_2} \right) + C_2 \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \quad (22)$$

leading to:

$$A_3 \asymp A_2, \quad (23)$$

$$C_3 = \frac{C_{\gamma,2}}{N} \lesssim \frac{1}{N} + \frac{1}{N} A_2 \frac{1}{C_0} \log \left(\frac{1}{A_2} \right) + \frac{C_2}{N} \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right). \quad (24)$$

This recurrence generalizes to all future generations:

$$A_t \asymp A, \quad (25)$$

$$C_t \lesssim \frac{1}{N} + \frac{1}{N} A \frac{1}{C_0} \log \left(\frac{1}{A} \right) + \frac{C_t}{N} \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right). \quad (26)$$

Solving the Recurrence Relation

Given the following definitions:

$$\alpha = \frac{1}{NC_0} \log \left(\frac{1}{C_0} \right), \quad \beta = \frac{1}{N} + \frac{A}{NC_0} \log \left(\frac{1}{A} \right),$$

we have the recurrence relation:

$$C_t \lesssim \beta + \alpha C_t,$$

which can be solved as:

$$C_t \lesssim (\alpha)^t C + \frac{\beta}{1 - \alpha} (1 - (\alpha)^t).$$

Thus, we have:

$$C_t \lesssim \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t C + \frac{\frac{1}{N} + \frac{A}{NC_0} \log \left(\frac{1}{A} \right)}{1 - \frac{1}{NC_0} \log \left(\frac{1}{C_0} \right)} \left[1 - \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^i \right].$$

Now recall $C = \frac{C_\gamma}{N}$. Under the assumption $N > \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)$, we conclude:

$$C_t \leq \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1 + \frac{A}{C_0} \log \left(\frac{1}{A} \right)}{N - \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)}. \quad (27)$$

Substituting this into the bound on $C_{\gamma,t}$, we get:

$$\begin{aligned} C_{\gamma,t} &\lesssim 1 + A_t \frac{1}{C_0} \log \left(\frac{1}{A_t} \right) + C_t \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\ &\lesssim 1 + \frac{1}{C_0} D_{\text{TV}} \log \left(\frac{1}{D_{\text{TV}}} \right) \end{aligned} \quad (28)$$

$$+ \left(\left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1 + \frac{A}{C_0} \log \left(\frac{1}{A} \right)}{N - \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)} \right) \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right). \quad (29)$$

Final Error Bound After T Recursive Generations

Thus, after T generations of the recursive training loop, we obtain

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_\gamma^+(x) \mid x) \leq 1 - \tau] &\lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{C_{\gamma,t}}{N} \log \left(\frac{1}{\tau} \right) \\ &\lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1}{N} \log \left(\frac{1}{\tau} \right) + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \\ &\quad + \left(\left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1 + \frac{A}{C_0} \log \left(\frac{1}{A} \right)}{N - \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)} \right) \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \frac{1}{N} \log \left(\frac{1}{\tau} \right) \\ &\lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1}{N} \log \left(\frac{1}{\tau} \right) + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \\ &\quad + \left(\left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \frac{1}{N^2} \right) \log \left(\frac{1}{\tau} \right) \\ &\lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} + \frac{1}{N^2} \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \right) \log \left(\frac{1}{\tau} \right) \\ &\quad + \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right). \end{aligned} \quad (30)$$

Furthermore, we obtain:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_\gamma^+(x) \mid x) \geq 1 - \tau] &\gtrsim 1 - \frac{1}{\tau} \cdot D_{\text{TV}} - \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} + \frac{1}{N^2} \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \right) \log \left(\frac{1}{\tau} \right) \\ &\quad - \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right). \end{aligned} \quad (31)$$

Final Accuracy and Error Decomposition

Define $y_T^+(x) = \arg \max_y \mathcal{G}_T(y | x)$. Then:

$$\begin{aligned}
& \text{Acc}(\mathcal{G}_T) \\
&= \mathbb{P}_{x \sim \mathcal{X}}[y_T^+(x) = y^*(x)] \\
&= \mathbb{P}_{x \sim \mathcal{X}}[y_T^+(x) = y^*(x) \mid \mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau] \cdot \mathbb{P}_{x \sim \mathcal{X}}[\mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau] \\
&\gtrsim 1 - \epsilon(\gamma) - \frac{1 - \epsilon(\gamma)}{\tau} \cdot D_{\text{TV}} - \left(\frac{1}{NC_0} \log\left(\frac{1}{C_0}\right) \right)^t \frac{C_\gamma(1 - \epsilon(\gamma))}{N} \log\left(\frac{1}{\tau}\right) \\
&\quad - (1 - \epsilon(\gamma)) \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} + \frac{1}{N^2} \frac{1}{C_0} \log\left(\frac{1}{C_0}\right) \right) \log\left(\frac{1}{\tau}\right). \tag{32}
\end{aligned}$$

Furthermore, we get

$$\begin{aligned}
& \text{Err}(\mathcal{G}_T) \\
&= \mathbb{P}_{x \sim \mathcal{X}}[y_T^+(x) \neq y^*(x)] \\
&\lesssim \epsilon(\gamma) + \frac{1 - \epsilon(\gamma)}{\tau} \cdot D_{\text{TV}} + \left(\frac{1}{NC_0} \log\left(\frac{1}{C_0}\right) \right)^t \frac{C_\gamma(1 - \epsilon(\gamma))}{N} \log\left(\frac{1}{\tau}\right) \\
&\quad + (1 - \epsilon(\gamma)) \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} + \frac{1}{N^2} \frac{1}{C_0} \log\left(\frac{1}{C_0}\right) \right) \log\left(\frac{1}{\tau}\right) \\
&\lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \left(\frac{1}{N} + \frac{1}{NC_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} + \frac{1}{N^2} \frac{1}{C_0} \log\left(\frac{1}{C_0}\right) \right) \log\left(\frac{1}{\tau}\right) \\
&\quad + \left(\frac{1}{NC_0} \log\left(\frac{1}{C_0}\right) \right)^t \frac{C_\gamma}{N} \log\left(\frac{1}{\tau}\right) + \epsilon(\gamma). \tag{33}
\end{aligned}$$

The proof is completed. $\square$

F Proof of Theorem 1: Error Amplification in Naive Recursive Training

In this section, we prove Theorem 1, which characterizes the failure mode of naive recursive training—where each model is trained solely on synthetic data generated by its predecessor, without any verification or quality control. The key result shows that prediction errors amplify exponentially across generations, ultimately leading to model collapse.

This phenomenon stems from the accumulation of distributional shifts at each generation, which progressively degrade supervision quality. Our analysis formalizes this effect by bounding the total error in terms of the cumulative divergence between successive models. Even when per-step errors are small, their unregulated propagation drives the model away from the original data distribution.

This negative result underscores the need for intervention—whether via real data, external feedback, or self-verification—to stabilize recursive training. We now present the detailed proof.

Proof of Theorem 1. We establish our results by carefully analyzing the propagation of errors in a recursive training paradigm without verification mechanisms. Our analytical approach parallels the methodology used in Theorem 3, with critical distinctions reflecting the absence of quality control measures.

Foundational Definitions and Decomposition

We begin by defining the subset of inputs with sufficient confidence under the base model:

$$\mathcal{D}_1^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \frac{\tau}{2} \right\},$$

where the high-quality candidate set $\mathbf{y}_\gamma^+(x)$ is defined via the verification score function:

$$\mathbf{y}_\gamma^+ = \left\{ y \mid s_{\text{verify}}(y | x) \geq \max_{y' \in \mathcal{Y}} s_{\text{verify}}(y' | x) - \gamma \right\},$$

with verification score function defined as $s_{\text{verify}}(y \mid x) := \log \mathcal{G}_0(y \mid x)$.

To establish our primary results, we decompose the probability of low-confidence generation:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(y_\gamma^+(x) \mid x) \leq 1 - \tau] \\ \leq \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(y_\gamma^+(x) \mid x) \leq 1 - \tau, x \in \mathcal{D}_1^+] + \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+]. \end{aligned} \quad (34)$$

Analysis of Model Divergence via Total Variation Distance

To quantify the disagreement between consecutive model generations, we leverage the total variation distance. For any input x and output y :

$$|\mathcal{G}_1(y \mid x) - \mathcal{G}_0(y \mid x)| \leq 2D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_0(\cdot \mid x)).$$

This inequality can be reformulated to establish a lower bound on the TV distance:

$$D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_0(\cdot \mid x)) \geq \frac{1}{2} |\mathcal{G}_1(y_\gamma^+(x) \mid x) - \mathcal{G}_0(y_\gamma^+(x) \mid x)|.$$

For inputs in the high-confidence region $\mathcal{D}_1^+$, we have:

$$\mathcal{G}_0(y_\gamma^+(x) \mid x) \geq 1 - \frac{\tau}{2}.$$

When $\mathcal{G}_1$ assigns insufficient confidence to the optimal region, i.e., $\mathcal{G}_1(y_\gamma^+(x) \mid x) \leq 1 - \tau$, the discrepancy between models is bounded below:

$$|\mathcal{G}_1(y_\gamma^+(x) \mid x) - \mathcal{G}_0(y_\gamma^+(x) \mid x)| \geq \tau - \frac{\tau}{2} = \frac{\tau}{2}.$$

This allows us to express the probability of low confidence for inputs in $\mathcal{D}_1^+$ in terms of the TV distance:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(y_\gamma^+(x) \mid x) \leq 1 - \tau, x \in \mathcal{D}_1^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_0(\cdot \mid x)). \quad (35)$$

Characterizing the Complement Set via Inverse Confidence

Let $C_\gamma := \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(y_\gamma^+(x) \mid x)} \right]$ denote the expected inverse confidence under the base model. Applying Markov's inequality, we derive:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+] &= \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_0(y_\gamma^+(x) \mid x) < 1 - \frac{\tau}{2}] \\ &= \mathbb{P}_{x \sim \mathcal{X}} \left[\frac{2 - \tau}{2\mathcal{G}_0(y_\gamma^+(x) \mid x)} > 1 \right] \\ &\leq \left(1 - \frac{\tau}{2}\right) \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(y_\gamma^+(x) \mid x)} \right] \\ &\leq \left(1 - \frac{\tau}{2}\right) C_\gamma. \end{aligned} \quad (36)$$

Integrating equations (35) and (36) into (34), we establish the comprehensive bound for the first generation:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(y_\gamma^+(x) \mid x) \leq 1 - \tau] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_0(\cdot \mid x)) + \left(1 - \frac{\tau}{2}\right) C_\gamma. \quad (37)$$

Recursive Analysis for the Second Generation Model $\mathcal{G}_2$

Following an analogous approach, we define the high-confidence region for the second generation:

$$\mathcal{D}_2^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_1(y_\gamma^+(x) \mid x) \geq 1 - \frac{\tau}{2} \right\},$$

By symmetric reasoning, we obtain:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_2^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_1(\cdot | x)). \quad (38)$$

For the complement set:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_2^+] &= \mathbb{P}_{x \sim \mathcal{X}} \left[\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) < 1 - \frac{\tau}{2} \right] \\ &\leq \left(1 - \frac{\tau}{2} \right) \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]. \end{aligned} \quad (39)$$

Advanced Analysis of Expected Inverse Confidence

The expected inverse confidence $\mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]$ represents a critical quantity for understanding error propagation across generations. We analyze this quantity by introducing the random variable:

$$Z(x) = \frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)}.$$

Utilizing the tail-integral identity for non-negative random variables:

$$\begin{aligned} \mathbb{E}[Z(x)] &= \int_0^\infty \mathbb{P}(Z(x) \geq t) dt \\ &= \int_0^1 1 dt + \int_1^{\frac{1}{C_0}} \mathbb{P}(Z(x) \geq t) dt \\ &= 1 + \int_1^{\frac{1}{C_0}} \mathbb{P}(Z(x) \geq t) dt. \end{aligned}$$

From equation (37), for $t \geq 1$, we derive:

$$\mathbb{P}_{x \sim \mathcal{X}} [Z(x) \geq t] \leq \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_0(\cdot | x)) + \left(\frac{1}{2} + \frac{1}{2t} \right) C_\gamma.$$

Threshold-based Analysis of Inverse Confidence Distribution

We define the bounding function:

$$B(t) = \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_0(\cdot | x)) + \left(\frac{1}{2} + \frac{1}{2t} \right) C_\gamma.$$

Since $B(t)$ is monotonically decreasing for $t > 1$, we focus on the non-trivial case where $B\left(\frac{1}{C_0}\right) < 1$, then there exists a unique threshold $t_0 \in \left(1, \frac{1}{C_0}\right)$ such that:

$$B(t_0) = 1 \quad \text{and} \quad B(t) \begin{cases} \geq 1, & 1 \leq t \leq t_0, \\ \leq 1, & t_0 \leq t \leq \frac{1}{C_0}. \end{cases}$$

This threshold partitioning enables us to refine our expectation bound:

$$\mathbb{E}[Z(x)] \leq 1 + (t_0 - 1) + \int_{t_0}^{\frac{1}{C_0}} B(t) dt.$$

Explicit Evaluation of the Integral Term

For analytical tractability, we represent $B(t)$ in the form:

$$B(t) = A \frac{t}{t-1} + C \frac{1}{t} + \frac{C_\gamma}{2},$$

where:

$$A = 4D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_0(\cdot | x)), \quad C = \frac{C_\gamma}{2}.$$

This decomposition yields:

$$\mathbb{E}[Z(x)] \leq t_0 + \frac{C_\gamma}{2} \left(\frac{1}{C_0} - t_0 \right) + \int_{t_0}^{\frac{1}{C_0}} \left(A \frac{t}{t-1} + C \frac{1}{t} \right) dt.$$

Through careful integration and algebraic manipulation:

$$\begin{aligned} \int_{t_0}^{\frac{1}{C_0}} A \cdot \frac{t}{t-1} + C \cdot \frac{1}{t} dt &= A \left(\frac{1}{C_0} - t_0 + \log \left| \frac{\frac{1}{C_0} - 1}{t_0 - 1} \right| \right) + C \log \left(\frac{1}{C_0 t_0} \right) \\ &\lesssim A \left(\frac{1}{C_0} + \log \frac{1}{t_0 - 1} \right) + C \log \left(\frac{1}{C_0} \right). \end{aligned}$$

Determining the Threshold Value t_0

The threshold t_0 is characterized by the equation $B(t_0) = 1$. Under the reasonable assumption that $C_\gamma + A < 2$, we can express t_0 explicitly:

$$t_0 = \frac{\sqrt{(C_\gamma - 1)^2 + 2AC_\gamma} + 1}{2 - (C_\gamma + 2A)}. \quad (40)$$

This allows us to derive the lower bound:

$$t_0 - 1 \geq \frac{2A}{2 - (2A + C_\gamma)}. \quad (41)$$

Analytical Expression for Expected Inverse Confidence

After substituting the threshold expressions and simplifying, we arrive at:

$$\mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right] \leq \frac{\frac{1}{2}C_\gamma + A}{C_0} + A \log \frac{1}{A} + \frac{C_\gamma}{2} \log \frac{1}{C_0}. \quad (42)$$

Consolidated Bound for $\mathcal{G}_2$

Denoting $C_{\gamma,1} = \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]$ and combining equations (38) and (39), we establish:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_1(\cdot | x)) + \left(1 - \frac{\tau}{2}\right) C_{\gamma,1}. \quad (43)$$

Generalization to Arbitrary Generations

To extend our analysis beyond the second generation, we assume uniform distributional shift across generations:

$$D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t-1}(\cdot | x)) \asymp D_{\text{TV}},$$

This assumption enables us to derive the recursive relation:

$$\begin{aligned} C_{\gamma,t} &\leq \frac{C_{\gamma,t-1} + A}{C_0} + A \log \frac{1}{A} + C_{\gamma,t-1} \log \frac{1}{C_0} \\ &\leq \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right) \frac{C_{\gamma,t-1}}{2} + A \log \frac{1}{A}. \end{aligned} \quad (44)$$

Solution to the Recursive Relation

We reframe this as a standard first-order linear recurrence relation:

$$C_{\gamma,t} \leq \alpha C_{\gamma,t-1} + \beta,$$

where $\alpha := \frac{1}{2} \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right) > 0$ and $\beta := A \log \frac{1}{A}$. Crucially, $\alpha > 1$, indicating exponential growth.

Unrolling this recurrence yields:

$$\begin{aligned} C_{\gamma,t} &\leq \alpha C_{\gamma,t-1} + \beta \\ &\leq \alpha (\alpha C_{\gamma,t-2} + \beta) + \beta = \alpha^2 C_{\gamma,t-2} + \alpha\beta + \beta \\ &\leq \alpha^3 C_{\gamma,t-3} + \alpha^2\beta + \alpha\beta + \beta \\ &\vdots \\ &\leq \alpha^t C_{\gamma} + \beta \sum_{k=0}^{t-1} \alpha^k. \end{aligned}$$

Applying the formula for finite geometric series:

$$\sum_{k=0}^{t-1} \alpha^k = \frac{1 - \alpha^t}{1 - \alpha}.$$

This gives:

$$C_{\gamma,t} \leq \alpha^t C_{\gamma} + \frac{\beta}{1 - \alpha} (1 - \alpha^t).$$

With simplification and substitution of our specific parameters:

$$C_{\gamma,t} \lesssim (C_{\gamma} + A \log \frac{1}{A}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^t. \quad (45)$$

Asymptotic Error Analysis After T Generations

Extending our analysis to the terminal generation T , we establish:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_{\gamma}^+(x) \mid x) \leq 1 - \tau] \lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + (C_{\gamma} + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T. \quad (46)$$

Correspondingly, for the complementary event:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_{\gamma}^+(x) \mid x) \geq 1 - \tau] \gtrsim 1 - \frac{1}{\tau} \cdot D_{\text{TV}} - (C_{\gamma} + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T. \quad (47)$$

Quantification of Model Accuracy and Error

For the final model $\mathcal{G}_T$, we define $y_T^+(x) = \arg \max_y \mathcal{G}_T(y \mid x)$ as its preferred prediction. The accuracy is then:

$$\begin{aligned} \text{Acc}(\mathcal{G}_T) &= \mathbb{P}_{x \sim \mathcal{X}} [y_T^+(x) = y^*(x)] \\ &= \mathbb{P}_{x \sim \mathcal{X}} [y_t^+(x) = y^*(x) \mid \mathcal{G}_t(\mathbf{y}_{\gamma}^+(x) \mid x) \geq 1 - \tau] \cdot \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_{\gamma}^+(x) \mid x) \geq 1 - \tau] \\ &\gtrsim (1 - \epsilon(\gamma)) \cdot \left(1 - \frac{1}{\tau} \cdot D_{\text{TV}} - (C_{\gamma} + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T \right) \\ &\gtrsim 1 - \epsilon(\gamma) - \frac{1 - \epsilon(\gamma)}{\tau} \cdot D_{\text{TV}} - (1 - \epsilon(\gamma))(C_{\gamma} + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T. \end{aligned} \quad (48)$$

The corresponding error rate is:

$$\begin{aligned} \text{Err}(\mathcal{G}_T) &= \mathbb{P}_{x \sim \mathcal{X}}[y_T^+(x) \neq y^*(x)] \\ &\lesssim \epsilon(\gamma) + \frac{1}{\tau} \cdot D_{\text{TV}} + (C_\gamma + D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}) \left(\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} \right)^T. \end{aligned} \quad (49)$$

The critical observation is that if $\frac{1}{2C_0} + \frac{1}{2} \log \frac{1}{C_0} > 1$, the error term grows exponentially with T , inexorably leading to model collapse as $T \rightarrow \infty$. This establishes the theorem. $\square$

G Proof of Theorem 2: Stabilization via Real Data Injection

In this section, we prove Theorem 2, which establishes that incorporating a sufficient proportion of real data at each generation can prevent model collapse in recursive training. Unlike naive training that relies exclusively on synthetic data, this hybrid strategy anchors the training distribution to the ground-truth data manifold, thereby mitigating the accumulation of distributional shift across generations.

The proof shows that even modest amounts of real data are sufficient to dominate the recursive error dynamics and ensure bounded generalization error over time. This result provides a theoretical justification for widely adopted empirical practices that mix real and synthetic data to maintain model quality. We now present the formal derivation.

Proof of Theorem 2. We establish our results by analyzing the error propagation in a recursive training paradigm that incorporates a proportion α of real data at each generation. The remaining $1 - \alpha$ proportion of data is generated synthetically by the previous model.

Foundational Definitions and Decomposition

Let the training dataset for generation t be a mixture of real data and synthetic data:

$$\mathcal{D}_t = \alpha \cdot \mathcal{D}_{\text{real}} + (1 - \alpha) \cdot \mathcal{D}_{\text{synthetic}},$$

where $\mathcal{D}_{\text{real}} = \{(x, y^*) \mid x \sim \mathcal{X}, y^* = \text{ground-truth label}\}$ and $\mathcal{D}_{\text{synthetic}} = \{(x, \hat{y}) \mid x \sim \mathcal{X}, \hat{y} = \mathcal{G}_{t-1}(x)\}$.

We define the mixed distribution as $\mathcal{G}_t^{\text{mix}} = \alpha \mathcal{G}^* + (1 - \alpha) \mathcal{G}_{t-1}$, where $\mathcal{G}^*$ represents the ground truth real data distribution.

We define the subset of inputs with sufficient real-data confidence under the base model $\mathcal{G}_0$:

$$\mathcal{D}_t^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_0(y^* \mid x) \geq 1 - \frac{\tau}{2(1 - \alpha)} \right\},$$

where y^* is the ground-truth label, and τ is the confidence threshold.

The probability of low-confidence generation at generation t can be decomposed as:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}}[\mathcal{G}_t(y^* \mid x) \leq 1 - \tau] &\leq \mathbb{P}_{x \sim \mathcal{X}}[\mathcal{G}_t(y^* \mid x) \leq 1 - \tau, x \in \mathcal{D}_t^+] \\ &\quad + \mathbb{P}_{x \sim \mathcal{X}}[x \notin \mathcal{D}_t^+]. \end{aligned} \quad (50)$$

Analysis of Model Divergence via Total Variation Distance

To quantify the disagreement between consecutive model generations, we leverage the total variation distance. For any input x and output y :

$$|\mathcal{G}_1(y \mid x) - \mathcal{G}_1^{\text{mix}}(y \mid x)| \leq 2D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_1^{\text{mix}}(\cdot \mid x)).$$

This inequality can be reformulated to establish a lower bound on the TV distance:

$$D_{\text{TV}}(\mathcal{G}_1(\cdot \mid x), \mathcal{G}_1^{\text{mix}}(\cdot \mid x)) \geq \frac{1}{2} |\mathcal{G}_1(y_\gamma^+(x) \mid x) - \mathcal{G}_1^{\text{mix}}(y_\gamma^+(x) \mid x)|.$$

For inputs in the high-confidence region $\mathcal{D}_1^+$, we have:

$$\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \frac{\tau}{2(1-\alpha)}.$$

When $\mathcal{G}_1$ assigns insufficient confidence to the optimal region, i.e., $\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau$, the discrepancy between models is bounded below:

$$|\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) - \mathcal{G}_1^{\text{mix}}(\mathbf{y}_\gamma^+(x) | x)| \geq \alpha + (1 - \alpha) \left(1 - \frac{\tau}{2(1-\alpha)}\right) - (1 - \tau) = \frac{\tau}{2}.$$

This allows us to express the probability of low confidence for inputs in $\mathcal{D}_1^+$ in terms of the TV distance:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_1^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x)). \quad (51)$$

Characterizing the Complement Set via Inverse Confidence

Let $C_\gamma := \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} \right]$ denote the expected inverse confidence under the base model. Applying Markov's inequality, we derive:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_1^+] &= \mathbb{P}_{x \sim \mathcal{X}} \left[\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x) < 1 - \frac{\tau}{2(1-\alpha)} \right] \\ &= \mathbb{P}_{x \sim \mathcal{X}} \left[\left(1 - \frac{\tau}{2(1-\alpha)}\right) \frac{1}{\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} > 1 \right] \\ &\leq \left(1 - \frac{\tau}{2(1-\alpha)}\right) \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_0(\mathbf{y}_\gamma^+(x) | x)} \right] \\ &\leq \left(1 - \frac{\tau}{2(1-\alpha)}\right) C_\gamma. \end{aligned} \quad (52)$$

Integrating equations (51) and (52) into (50), we establish the comprehensive bound for the first generation:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x)) + \left(1 - \frac{\tau}{2(1-\alpha)}\right) C_\gamma. \quad (53)$$

Recursive Analysis for the Second Generation Model $\mathcal{G}_2$

Following an analogous approach, we define the high-confidence region for the second generation:

$$\mathcal{D}_2^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \frac{\tau}{2(1-\alpha)} \right\},$$

Then, we obtain:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau, x \in \mathcal{D}_2^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_2^{\text{mix}}(\cdot | x)). \quad (54)$$

For the complement set:

$$\begin{aligned} \mathbb{P}_{x \sim \mathcal{X}} [x \notin \mathcal{D}_2^+] &= \mathbb{P}_{x \sim \mathcal{X}} \left[\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x) < 1 - \frac{\tau}{2(1-\alpha)} \right] \\ &\leq \left(1 - \frac{\tau}{2(1-\alpha)}\right) \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]. \end{aligned} \quad (55)$$

Advanced Analysis of Expected Inverse Confidence

The expected inverse confidence $\mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]$ represents a critical quantity for understanding error propagation across generations. We analyze this quantity by introducing the random variable:

$$Z(x) = \frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)}.$$

Utilizing the tail-integral identity for non-negative random variables:

$$\begin{aligned} \mathbb{E}[Z(x)] &= \int_0^\infty \mathbb{P}(Z(x) \geq t) dt \\ &= \int_0^1 1 dt + \int_1^{\frac{1}{c_0}} \mathbb{P}(Z(x) \geq t) dt \\ &= 1 + \int_1^{\frac{1}{c_0}} \mathbb{P}(Z(x) \geq t) dt. \end{aligned}$$

From equation (53), for $t \geq 1$, we derive:

$$\mathbb{P}_{x \sim \mathcal{X}} [Z(x) \geq t] \leq \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x)) + \left(1 - \frac{1}{2(1-\alpha)} + \frac{1}{2(1-\alpha)t} \right) C_\gamma.$$

Threshold-based Analysis of Inverse Confidence Distribution

We define the bounding function:

$$B(t) = \frac{4t}{t-1} D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x)) + \left(1 - \frac{1}{2(1-\alpha)} + \frac{1}{2(1-\alpha)t} \right) C_\gamma.$$

Since $B(t)$ is monotonically decreasing for $t > 1$, we focus on the non-trivial case where $B\left(\frac{1}{c_0}\right) < 1$, then there exists a unique threshold $t_0 \in \left(1, \frac{1}{c_0}\right)$ such that:

$$B(t_0) = 1 \quad \text{and} \quad B(t) \begin{cases} \geq 1, & 1 \leq t \leq t_0, \\ \leq 1, & t_0 \leq t \leq \frac{1}{c_0}. \end{cases}$$

This threshold partitioning enables us to refine our expectation bound:

$$\mathbb{E}[Z(x)] \leq 1 + (t_0 - 1) + \int_{t_0}^{\frac{1}{c_0}} B(t) dt.$$

Explicit Evaluation of the Integral Term

For analytical tractability, we represent $B(t)$ in the form:

$$B(t) = A \frac{t}{t-1} + C \frac{1}{t} + \left(1 - \frac{1}{2(1-\alpha)} \right) C_\gamma,$$

where:

$$A = 4D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x)), \quad C = \frac{C_\gamma}{2(1-\alpha)}.$$

This decomposition yields:

$$\mathbb{E}[Z(x)] \leq t_0 + \left(1 - \frac{1}{2(1-\alpha)} \right) C_\gamma \left(\frac{1}{C_0} - t_0 \right) + \int_{t_0}^{\frac{1}{c_0}} \left(A \frac{t}{t-1} + C \frac{1}{t} \right) dt.$$

Through careful integration and algebraic manipulation:

$$\begin{aligned} \int_{t_0}^{\frac{1}{C_0}} A \cdot \frac{t}{t-1} + C \cdot \frac{1}{t} dt &= A \left(\frac{1}{C_0} - t_0 + \log \left| \frac{\frac{1}{C_0} - 1}{t_0 - 1} \right| \right) + C \log \left(\frac{1}{C_0 t_0} \right) \\ &\lesssim A \left(\frac{1}{C_0} + \log \frac{1}{t_0 - 1} \right) + C \log \left(\frac{1}{C_0} \right). \end{aligned}$$

Determining the Threshold Value t_0

The threshold t_0 is characterized by the equation $B(t_0) = 1$. Then, we can express t_0 explicitly:

$$\begin{aligned} t_0 - 1 &= \frac{2A + C_\gamma - 1 + \sqrt{\left(\frac{C_\gamma \alpha}{1-\alpha} + 1\right)^2 - 4 \left(-A - C_\gamma + \frac{C_\gamma}{2(1-\alpha)} + 1\right) \cdot \frac{C_\gamma}{2(1-\alpha)}}}{-2A - 2C_\gamma + \frac{C_\gamma}{1-\alpha} + 2} \\ &= \frac{2A + C_\gamma - 1 + \sqrt{\left(\frac{C_\gamma \alpha}{1-\alpha} + 1\right)^2 + \left(2A + 2C_\gamma - \frac{C_\gamma}{1-\alpha} - 2\right) \cdot \frac{C_\gamma}{(1-\alpha)}}}{-2A - 2C_\gamma + \frac{C_\gamma}{1-\alpha} + 2}. \end{aligned} \quad (56)$$

This allows us to derive the lower bound:

$$t_0 - 1 \geq \frac{2A}{2 - (2A + C_\gamma)}. \quad (57)$$

Analytical Expression for Expected Inverse Confidence

Let $A = D_{\text{TV}}(\mathcal{G}_1(\cdot | x), \mathcal{G}_1^{\text{mix}}(\cdot | x))$. After substituting the threshold expressions and simplifying, we arrive at:

$$\begin{aligned} \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right] &\leq 1 + \int_1^{\frac{1}{C_0}} \frac{4t}{t-1} A + \left(1 - \frac{1}{2(1-\alpha)} + \frac{1}{2(1-\alpha)t} \right) C_\gamma dt \\ &\lesssim \left(1 - \frac{1}{2(1-\alpha)} \right) C_\gamma \frac{1}{C_0} + \frac{A}{C_0} + A \log \frac{1}{A} + \frac{C_\gamma}{2(1-\alpha)} \log \frac{1}{C_0}. \end{aligned} \quad (58)$$

Consolidated Bound for $\mathcal{G}_2$

Denoting $C_{\gamma,1} = \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_1(\mathbf{y}_\gamma^+(x) | x)} \right]$ and combining equations (38) and (39), we establish:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_2(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_2(\cdot | x), \mathcal{G}_1(\cdot | x)) + \left(1 - \frac{\tau}{2(1-\alpha)} \right) C_{\gamma,1}. \quad (59)$$

Generalization to Arbitrary Generations

To extend our analysis beyond the second generation, we assume uniform distributional shift across generations:

$$D_{\text{TV}}(\mathcal{G}_t(\cdot | x), \mathcal{G}_{t-1}(\cdot | x)) \asymp D_{\text{TV}},$$

This assumption enables us to derive the recursive relation:

$$\begin{aligned} C_{\gamma,t} &\leq \frac{C_{\gamma,t-1} + A}{C_0} + A \log \frac{1}{A} + \frac{C_{\gamma,t-1}}{1-\alpha} \log \frac{1}{C_0} \\ &\leq \left(\left(1 - \frac{1}{2(1-\alpha)} \right) \frac{1}{C_0} + \frac{1}{2(1-\alpha)} \log \frac{1}{C_0} \right) C_{\gamma,t-1} + A \log \frac{1}{A}. \end{aligned} \quad (60)$$

Assume $\alpha > 1 - \frac{1}{2(\frac{1}{C_0}-1)}$, $T \rightarrow \infty$, then we obtain:

$$C_{\gamma,T} \lesssim \frac{1-\alpha}{C_0} A \log \frac{1}{A} \quad (61)$$

Asymptotic Error Analysis After T Generations

Extending our analysis to the terminal generation T , we establish:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1 - \alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}. \quad (62)$$

Quantification of Model Accuracy and Error

For the final model $\mathcal{G}_T$, we define $y_T^+(x) = \arg \max_y \mathcal{G}_T(y | x)$ as its preferred prediction. The accuracy is then:

$$\begin{aligned} \text{Acc}(\mathcal{G}_T) &= \mathbb{P}_{x \sim \mathcal{X}} [y_T^+(x) = y^*(x)] \\ &= \mathbb{P}_{x \sim \mathcal{X}} [y_T^+(x) = y^*(x) | \mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau] \cdot \mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \geq 1 - \tau] \\ &\gtrsim (1 - \epsilon(\gamma)) \cdot \left(1 - \frac{1}{\tau} \cdot D_{\text{TV}} - \frac{1 - \alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}} \right) \\ &\gtrsim 1 - \epsilon(\gamma) - \frac{1 - \epsilon(\gamma)}{\tau} \cdot D_{\text{TV}} - (1 - \epsilon(\gamma)) \frac{1 - \alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}. \end{aligned} \quad (63)$$

The corresponding error rate is:

$$\begin{aligned} \text{Err}(\mathcal{G}_T) &= \mathbb{P}_{x \sim \mathcal{X}} [y_T^+(x) \neq y^*(x)] \\ &\lesssim \epsilon(\gamma) + \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1 - \alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}. \end{aligned} \quad (64)$$

Conclusion

The inclusion of real data at each generation mitigates error accumulation, ensuring that the model maintains stability over recursive generations. This completes the proof. $\square$

H Proof of Theorem 4: Theoretical Guarantees for Transformer-Based LLMs with Self-Verification

In this section, we prove Theorem 4, which extends our self-verification framework to autoregressive Transformer-based LLMs. This result establishes that, even in high-capacity architectures, self-verification can effectively prevent error amplification across recursive training generations.

To enable this analysis, we assume that the Transformer model parameters are uniformly bounded, a standard condition in prior theoretical studies [Li et al., 2023, Hu et al., 2024, Zhang et al., 2025, Fu et al., 2025] as follows:

Assumption 2 (Bounded Parameters [Zhang et al., 2025]). The parameter space Θ of the Transformer model is constrained as:

$$\Theta = \left\{ \theta \left| \begin{aligned} &\left\| A^{(D+1), \top} \right\|_{1,2} \leq B_A, \max \{ |\gamma_1^{(t)}|, |\gamma_2^{(t)}| \} \leq 1, \left\| A_1^{(t)} \right\|_F \leq B_{A,1}, \left\| A_2^{(t)} \right\|_F \leq B_{A,2}, \\ &\left\| W_{Q,i}^{(t)} \right\|_F \leq B_Q, \left\| W_{K,i}^{(t)} \right\|_F \leq B_K, \left\| W_{V,i}^{(t)} \right\|_F \leq B_V \text{ for all } t \in [D], i \in [h] \end{aligned} \right\},$$

where h denotes the number of attention heads, and $B_A, B_{A,1}, B_{A,2}, B_Q, B_K, B_V > 1$ are pre-specified constants.

Under this assumption, we then present the complete theorem statement characterizing the final test error after T recursive steps.

Theorem 5 (Error Bound for Transformer-based LLMs with Self-Verification). *Let $\mathcal{G}_{\theta_t}^{LLM}(x | S)$ denote the conditional generation distribution of a Transformer-based LLM at the t -th recursive generation, where $S \in \mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$ is a token prefix sequence and $x \in \mathcal{X}$ is the next token. Then, after*

T recursive updates, the final prediction error is bounded with probability at least $1 - \delta$:

$$\begin{aligned} \text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) &\lesssim \epsilon(\gamma) + \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) \right) \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \\ &\quad + \left(\frac{1}{N} + \frac{1}{N^2 C_0} \right) \log \left(\frac{1}{\tau} \right) + \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^T \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right). \end{aligned} \quad (65)$$

Furthermore, if $N > \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)$, then as $T \rightarrow \infty$, the exponential decay term vanishes, and we obtain the bound:

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) \lesssim \epsilon(\gamma) + \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \frac{1}{\tau} \right) \frac{\log(n\tilde{B}) \tilde{D}}{\sqrt{n}} \log \frac{1}{\delta} + \frac{1}{N} \log \frac{1}{\tau},$$

where $n = m\ell$ is the total number of training samples, and $\tilde{D} = D^2 \cdot d \cdot (d_F + d_h + d) + d \cdot d_y$ and $\tilde{B} = \beta^{-1} R h B_A B_{A,1} B_{A,2} B_Q B_K B_V$ reflect the model capacity and prior radius respectively.

Our proof draws inspiration from the PAC-Bayesian analysis developed in Zhang et al. [2025], but introduces new techniques to handle recursive training dynamics. In particular, we leverage a recursive coverage coefficient to track how the model’s confidence evolves across generations, allowing us to rigorously control generalization error under fully synthetic training. We now present the formal derivation.

Proof of Theorem 5. We present the proof as follows.

MLE Initialization and Sequence-Based Setup

We assume that all token prefix sequences $S_t = (x_1, \dots, x_t)$ are elements of a structured sequence space $\mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$, where $\mathcal{X}$ denotes the token vocabulary and ℓ is the maximum sequence length. This ensures that all context representations during both training and generation reside within a bounded and well-defined domain.

The initial autoregressive language model $\mathcal{G}_{\theta_0}^{\text{LLM}}$ is trained via maximum likelihood estimation (MLE) on a corpus of sequences $\{(x_1^{(j)}, \dots, x_\ell^{(j)})\}_{j=1}^m$. The training objective is given by:

$$\theta_0 = \arg \max_{\theta \in \Theta_{\text{LLM}}} \frac{1}{n} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \mathcal{G}_{\theta}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)}),$$

where $S_t^{(j)} = (x_1^{(j)}, \dots, x_t^{(j)}) \in \mathcal{S}$ and $n = m\ell$ is the total number of prefix-conditioned samples. The resulting model $\mathcal{G}_{\theta_0}^{\text{LLM}}$ serves as a fixed reference distribution encoding inductive biases from human supervision.

Defining the Confident Support Set $\mathcal{S}_1^+$

To identify reliable contexts for self-verification, we define the verification score function as:

$$s_{\text{verify}}(x \mid S) := \log \mathcal{G}_{\theta_0}^{\text{LLM}}(x \mid S),$$

and the corresponding high-quality candidate set:

$$\mathcal{S}_\gamma^+ := \left\{ x \in \mathcal{X} \mid s_{\text{verify}}(x \mid S) \geq \max_{x' \in \mathcal{X}} s_{\text{verify}}(x' \mid S) - \gamma \right\}.$$

We then define the confident support set $\mathcal{S}_1^+$ as the subset of prefixes where the reference model assigns sufficient mass to the high-reward region:

$$\mathcal{S}_1^+ := \left\{ S \in \mathcal{S} \mid \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathcal{S}_\gamma^+ \mid S) \geq \frac{\log \frac{2}{\tau}}{N} \right\}.$$

Decomposing the Probability of Low-Confidence Generation

We aim to bound the probability that the updated model assigns low probability mass to verified candidates:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau] &\leq \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau, S \in \mathcal{S}_1^+] \\ &\quad + \mathbb{P}_{S \sim \mathcal{S}} [S \notin \mathcal{S}_1^+]. \end{aligned} \quad (66)$$

Bounding the First Term: Monte Carlo Estimate for Verified Support

Let $\mathcal{G}_{1,N}^*$ be the empirical estimator from N samples drawn from $\mathcal{G}_{\theta_0}^{\text{LLM}}$. For $S \in \mathcal{S}_1^+$, the probability that none of the samples falls into $\mathbf{x}_\gamma^+(S)$ is at most:

$$(1 - \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S))^N \leq \left(1 - \frac{\log \frac{2}{\tau}}{N}\right)^N \leq \frac{\tau}{2}.$$

Hence,

$$\mathcal{G}_{1,N}^*(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \frac{\tau}{2}.$$

Bounding Discrepancy via Total Variation Distance

The total variation distance between two distributions P and Q over $\mathcal{X}$ is:

$$D_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|.$$

For any event $A \subseteq \mathcal{X}$, it satisfies:

$$|P(A) - Q(A)| \leq 2D_{\text{TV}}(P, Q).$$

In particular, if $\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau$ and $\mathcal{G}_{1,N}^*(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \frac{\tau}{2}$, then

$$|\mathcal{G}_{\theta_1}^{\text{LLM}} - \mathcal{G}_{1,N}^*| \geq \frac{\tau}{2} \Rightarrow D_{\text{TV}} \geq \frac{\tau}{4}.$$

Taking expectation:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau, S \in \mathcal{S}_1^+] \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}, \mathcal{G}_{1,N}^*)]. \quad (67)$$

Bounding the Second Term via Markov's Inequality

Define the expected inverse support quantity:

$$C_\gamma := \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S)} \right].$$

Then, by Markov's inequality:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} [S \notin \mathcal{S}_1^+] &= \mathbb{P} \left[\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) < \frac{\log \frac{2}{\tau}}{N} \right] \\ &= \mathbb{P} \left[\frac{\log \frac{2}{\tau}}{N \cdot \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S)} > 1 \right] \\ &\leq \frac{C_\gamma}{N} \log \frac{2}{\tau}. \end{aligned} \quad (68)$$

Final Bound for the First-Step Generation Error

Combining inequalities (7) and (8), and substituting into the decomposition, we obtain the following upper bound:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau] \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{1,N}^*(\cdot \mid S))] + \frac{C_\gamma}{N} \log \frac{2}{\tau}. \quad (69)$$

Since $\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S)$ is the maximum likelihood estimator of the conditional distribution $\mathcal{G}_{1,N}^*(\cdot \mid S)$ based on sampled data, we now focus on analyzing the expected total variation distance $\mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{1,N}^*(\cdot \mid S))]$.

Step 1: Error Decomposition under Token Prefix Sequence Modeling

We begin by establishing a framework for analyzing the error. Consider our dataset consisting of m token trajectories, each of length ℓ :

$$\mathcal{D} = \left\{ \left(x_1^{(j)}, x_2^{(j)}, \dots, x_\ell^{(j)} \right) \right\}_{j=1}^m,$$

where each token prefix $S_t^{(j)} := (x_1^{(j)}, \dots, x_t^{(j)}) \in \mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$.

To measure how closely our model approximates the reference distribution, we define the log-likelihood discrepancy function:

$$L(\theta, \mathcal{D}) := -\frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{1,N}^*(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_\theta^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})}.$$

For applying the PAC-Bayes framework, we introduce a ghost sample $\tilde{\mathcal{D}} = \left\{ (\tilde{S}_t^{(j)}, \tilde{x}_{t+1}^{(j)}) \right\}_{j=1, t=1}^{m, \ell}$, where $\tilde{S}_t^{(j)} = S_t^{(j)}$ and $\tilde{x}_{t+1}^{(j)} \sim \mathcal{G}_{1,N}^*(\cdot \mid \tilde{S}_t^{(j)})$. We then define:

$$g(\theta) := L(\theta, \mathcal{D}) - \log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(L(\theta, \tilde{\mathcal{D}}) \right) \mid \mathcal{D} \right].$$

PAC-Bayes Bound under Token Prefix Modeling

By the standard PAC-Bayes lemma, for any distributions $Q, P \in \Delta(\Theta)$ where P may depend on $\mathcal{D}$, we have:

$$\mathbb{E}_{\theta \sim P} [g(\theta)] \leq \text{KL}(P \parallel Q) + \log \mathbb{E}_{\theta \sim Q} [\exp(g(\theta))].$$

Substituting the definition of $g(\theta)$ and taking expectation over $\mathcal{D}$:

$$\mathbb{E}_{\mathcal{D}} \left[\exp \left\{ \mathbb{E}_{\theta \sim P} \left[L(\theta, \mathcal{D}) - \log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp(L(\theta, \tilde{\mathcal{D}})) \mid \mathcal{D} \right] \right] - \text{KL}(P \parallel Q) \right\} \right] \leq 1.$$

Applying Chernoff bounds, with probability at least $1 - \delta$, we obtain:

$$-\mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp(L(\theta, \tilde{\mathcal{D}})) \mid \mathcal{D} \right] \right] \leq -\mathbb{E}_{\theta \sim P} [L(\theta, \mathcal{D})] + \text{KL}(P \parallel Q) + \log \frac{1}{\delta}.$$

We can lower-bound the left-hand side using a two-step approach involving Jensen's inequality and total variation distance:

$$\begin{aligned}
& -\mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp(L(\theta, \tilde{\mathcal{D}})) \mid \mathcal{D} \right] \right] \\
& \geq -\frac{1}{2} \log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(-\frac{1}{2} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{1,N}^*(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_{\theta_1}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})} \right) \mid \mathcal{D} \right] \\
& \quad - \frac{1}{2} \mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(-\frac{1}{2} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{\theta_1}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_{\theta}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})} \right) \mid \mathcal{D} \right] \right] \\
& \geq \frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}^2(\mathcal{G}_{1,N}^*(\cdot \mid S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S_t^{(j)})) \\
& \quad - \frac{1}{2} \mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(-\frac{1}{2} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{\theta_1}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_{\theta}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})} \right) \mid \mathcal{D} \right] \right].
\end{aligned}$$

Next, for any reference parameter $\theta^* \in \Theta$, we decompose the expected loss:

$$\begin{aligned}
-\mathbb{E}_{\theta \sim P} [L(\theta, \mathcal{D})] &= \mathbb{E}_{\theta \sim P} [L(\theta^*, \mathcal{D}) + (L(\theta_1, \mathcal{D}) - L(\theta^*, \mathcal{D})) + (L(\theta, \mathcal{D}) - L(\theta_1, \mathcal{D}))] \\
&\leq L(\theta^*, \mathcal{D}) + \mathbb{E}_{\theta \sim P} [L(\theta, \mathcal{D}) - L(\theta_1, \mathcal{D})],
\end{aligned}$$

where we use the fact that θ_1 minimizes $L(\theta, \mathcal{D})$. In the realizable setting, there exists θ^* such that $\mathcal{G}_{\theta^*}^{\text{LLM}}(\cdot \mid S) = \mathcal{G}_{1,N}^*(\cdot \mid S)$.

Combining these bounds, we obtain:

$$\frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}^2(\mathcal{G}_{1,N}^*(\cdot \mid S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S_t^{(j)})) \tag{70}$$

$$\begin{aligned}
& \leq \frac{1}{2} \mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(-\frac{1}{2} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{\theta_1}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_{\theta}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})} \right) \mid \mathcal{D} \right] \right] \\
& \quad + \frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} \mathbb{E}_{\theta \sim P} \left(\log \frac{\mathcal{G}_{\theta_1}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})}{\mathcal{G}_{\theta}^{\text{LLM}}(x_{t+1}^{(j)} \mid S_t^{(j)})} \right) + \text{KL}(P \parallel Q) + \log \frac{1}{\delta}.
\end{aligned} \tag{71}$$

The first two terms capture the fluctuation under the posterior distribution $\theta \sim P$, which we will address next.

Step 2: Controlling the Fluctuation via Structured Posterior Sampling

To control the fluctuation terms, we define the posterior distribution $P \in \Delta(\Theta)$ as a product distribution over each layer of the transformer:

$$P = \prod_{t=1}^{D+1} \mathcal{L}_P(\theta^{(t)}) \tag{72}$$

$$\begin{aligned}
\mathcal{L}_P(\theta^{(D+1)}) &= \text{Unif} \left(\mathbb{B} \left(\hat{A}^{(D+1)}, r^{(D+1)}, \|\cdot\|_{1,2} \right) \right) \\
\mathcal{L}_P(\theta^{(t)}) &= \text{Unif} \left(\mathbb{B}(\hat{\gamma}_1^{(t)}, r_{\gamma,1}^{(t)}, |\cdot|) \right) \cdot \text{Unif} \left(\mathbb{B}(\hat{\gamma}_2^{(t)}, r_{\gamma,2}^{(t)}, |\cdot|) \right) \cdot \mathcal{L}_P(A^{(t)}) \cdot \mathcal{L}_P(W^{(t)}) \\
\mathcal{L}_P(A^{(t)}) &= \text{Unif} \left(\mathbb{B}(\hat{A}_1^{(t)}, r_{A,1}^{(t)}, \|\cdot\|_F) \right) \cdot \text{Unif} \left(\mathbb{B}(\hat{A}_2^{(t)}, r_{A,2}^{(t)}, \|\cdot\|_F) \right) \\
\mathcal{L}_P(W^{(t)}) &= \prod_{i=1}^h \text{Unif} \left(\mathbb{B}(\hat{W}_i^{Q,(t)}, r_Q^{(t)}, \|\cdot\|_F) \right) \cdot \text{Unif} \left(\mathbb{B}(\hat{W}_i^{K,(t)}, r_K^{(t)}, \|\cdot\|_F) \right) \cdot \text{Unif} \left(\mathbb{B}(\hat{W}_i^{V,(t)}, r_V^{(t)}, \|\cdot\|_F) \right)
\end{aligned} \tag{73}$$

for all $t \in [D]$, where $\text{Unif}(\mathbb{B}(a, r, \|\cdot\|))$ denotes the uniform distribution over a ball centered at a with radius r under norm $\|\cdot\|$.

The radii are carefully calibrated as follows:

$$\begin{aligned} r_{\gamma,1}^{(t)} &= R^{-1}(1 + B_{A,1}B_{A,2})^{-1}\alpha_t^{-1}/(m\ell), & r_{\gamma,2}^{(t)} &= R^{-1}\alpha_t^{-1}/(m\ell), \\ r_{A,1}^{(t)} &= R^{-1}B_{A,2}^{-1}\alpha_t^{-1}/(m\ell), & r_{A,2}^{(t)} &= R^{-1}B_{A,1}^{-1}\alpha_t^{-1}/(m\ell), \\ r_V^{(t)} &= R^{-1}h^{-1}(1 + B_{A,1}B_{A,2})^{-1}\alpha_t^{-1}/(m\ell), & r_Q^{(t)} &= R^{-1}h^{-1}(1 + B_{A,1}B_{A,2})^{-1}B_V^{-1}B_K^{-1}\alpha_t^{-1}/(m\ell), \\ r_K^{(t)} &= R^{-1}h^{-1}(1 + B_{A,1}B_{A,2})^{-1}B_V^{-1}B_Q^{-1}\alpha_t^{-1}/(m\ell), & r^{(D+1)} &= \tau B_A^{-1}/(m\ell). \end{aligned}$$

This carefully structured posterior allows us to prove the following key result:

Lemma 6. *Zhang et al. [2025] With the posterior distribution P defined in (72), which is a uniform distribution over a neighborhood around θ_1 with radius proportional to $1/(m\ell)$, we have:*

$$\begin{aligned} \frac{1}{2}\mathbb{E}_{\theta \sim P} \left[\log \mathbb{E}_{\tilde{\mathcal{D}}} \left[\exp \left(-\frac{1}{2} \sum_{j=1}^m \sum_{t=1}^{\ell} \log \frac{\mathcal{G}_{\theta_1}^{LLM}(x_{t+1}^{(j)} | S_t^{(j)})}{\mathcal{G}_{\theta}^{LLM}(x_{t+1}^{(j)} | S_t^{(j)})} \right) \mid \mathcal{D} \right] \right] \\ + \frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} \mathbb{E}_{\theta \sim P} \left(\log \frac{\mathcal{G}_{\theta_1}^{LLM}(x_{t+1}^{(j)} | S_t^{(j)})}{\mathcal{G}_{\theta}^{LLM}(x_{t+1}^{(j)} | S_t^{(j)})} \right) = \mathcal{O}(1). \end{aligned} \quad (74)$$

Proof. A detailed proof can be found in Appendix F.2 of Zhang et al. [2025]. The key insight is that the construction of P with specific radii ensures that the model outputs remain stable under small parameter perturbations, allowing us to bound the fluctuation terms by a constant. $\square$

Bounding the KL Divergence Term

To complete our bound, we need to quantify the KL divergence term in (71). We define the reference prior distribution $Q \in \Delta(\Theta)$ as:

$$Q = \prod_{t=1}^{D+1} \mathcal{L}_Q(\theta^{(t)}),$$

with the following component-wise structure:

- **Final linear layer:**

$$\mathcal{L}_Q(\theta^{(D+1)}) = \text{Unif}(\mathbb{B}(0, B_A, \|\cdot\|_{1,2})).$$

- **Transformer layers $t \in [D]$:**

$$\mathcal{L}_Q(\theta^{(t)}) = \text{Unif}(\mathbb{B}(1/2, 1/2, |\cdot|)) \cdot \text{Unif}(\mathbb{B}(1/2, 1/2, |\cdot|)) \cdot \mathcal{L}_Q(A^{(t)}) \cdot \mathcal{L}_Q(W^{(t)}),$$

$$\mathcal{L}_Q(A^{(t)}) = \text{Unif}(\mathbb{B}(0, B_{A,1}, \|\cdot\|_F)) \cdot \text{Unif}(\mathbb{B}(0, B_{A,2}, \|\cdot\|_F)),$$

$$\mathcal{L}_Q(W^{(t)}) = \prod_{i=1}^h \text{Unif}(\mathbb{B}(0, B_Q, \|\cdot\|_F)) \cdot \text{Unif}(\mathbb{B}(0, B_K, \|\cdot\|_F)) \cdot \text{Unif}(\mathbb{B}(0, B_V, \|\cdot\|_F)).$$

The KL divergence between this prior Q and our posterior P can be bounded as:

$$\begin{aligned} \text{KL}(P\|Q) &= \mathcal{O} \left((D^2 \cdot d \cdot (d_F + d_h + d) + d \cdot d_y) \cdot \log \left(1 + \frac{m\ell}{\beta} R h B_A B_{A,1} B_{A,2} B_Q B_K B_V \right) \right) \\ &= \mathcal{O} \left(\tilde{D} \cdot \log \left(1 + m\ell \tilde{B} \right) \right) \end{aligned} \quad (75)$$

This term represents the complexity of the posterior family relative to the prior and controls the flexibility allowed in posterior fluctuations under the PAC-Bayes objective.

Deriving the Final Bound

By combining (74) and (75) and substituting into (71), we obtain:

$$\frac{1}{4} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}^2(\mathcal{G}_{1,N}^*(\cdot | S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S_t^{(j)})) \lesssim \mathcal{O}(1) + \mathcal{O}\left(\tilde{D} \cdot \log\left(1 + m\ell\tilde{B}\right)\right) + \log \frac{1}{\delta}$$

Taking the square root and using Jensen's inequality, we get:

$$\begin{aligned} & \frac{1}{m\ell} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}(\mathcal{G}_{1,N}^*(\cdot | S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S_t^{(j)})) \\ & \leq \sqrt{\frac{1}{(m\ell)^2} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}^2(\mathcal{G}_{1,N}^*(\cdot | S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S_t^{(j)}))} \\ & \lesssim \sqrt{\frac{\tilde{D} \cdot \log\left(1 + m\ell\tilde{B}\right) + \log \frac{1}{\delta}}{m\ell}} \end{aligned}$$

Based on Lemma F.4 in Hu et al. [2024], we can further bound the difference between the empirical and expected total variation distances:

$$\begin{aligned} & \frac{1}{m\ell} \sum_{j=1}^m \sum_{t=1}^{\ell} \mathbb{E}_{S_t^{(j)}} \left[D_{\text{TV}}(\mathcal{G}_{1,N}^*(\cdot | S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S_t^{(j)})) \right] - \frac{1}{m\ell} \sum_{j=1}^m \sum_{t=1}^{\ell} D_{\text{TV}}(\mathcal{G}_{1,N}^*(\cdot | S_t^{(j)}), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S_t^{(j)})) \\ & \lesssim \sqrt{\frac{1}{m\ell}} \left(\tilde{D} \cdot \log\left(1 + m\ell\tilde{B}\right) + \log \frac{1}{\delta} \right) \end{aligned} \quad (76)$$

Therefore, with probability at least $1 - \delta$, we can bound the expected total variation distance:

$$\mathbb{E}_{S \sim \mathcal{S}} \left[D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S), \mathcal{G}_{1,N}^*(\cdot | S)) \right] \lesssim \sqrt{\frac{\tilde{D} \cdot \log\left(1 + m\ell\tilde{B}\right) + \log \frac{1}{\delta}}{m\ell}} \quad (77)$$

Final Bound for $\mathcal{G}_{\theta_1}^{\text{LLM}}$

Returning to our original objective in (69), we can now derive the final bound:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} \left[\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_{\gamma}^+(S) | S) \leq 1 - \tau \right] & \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} \left[D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S), \mathcal{G}_{1,N}^*(\cdot | S)) \right] + \frac{C_{\gamma}}{N} \log \frac{2}{\tau} \\ & \lesssim \frac{4}{\tau} \cdot \sqrt{\frac{\tilde{D} \cdot \log\left(1 + m\ell\tilde{B}\right) + \log \frac{1}{\delta}}{m\ell}} + \frac{C_{\gamma}}{N} \log \frac{2}{\tau} \\ & \lesssim \frac{4}{\tau} \cdot \sqrt{\frac{\tilde{D} \cdot \log\left(m\ell\tilde{B}\right)}{m\ell}} \log \frac{1}{\delta} + \frac{C_{\gamma}}{N} \log \frac{2}{\tau} \end{aligned} \quad (78)$$

Let $n = m\ell$ denote the total number of samples. Thus, with probability at least $1 - \delta$, we have:

$$\mathbb{P}_{S \sim \mathcal{S}} \left[\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_{\gamma}^+(S) | S) \leq 1 - \tau \right] \lesssim \frac{1}{\tau} \cdot \sqrt{\frac{\tilde{D} \log\left(n\tilde{B}\right)}{n}} \log \frac{1}{\delta} + \frac{C_{\gamma}}{N} \log \frac{1}{\tau} \quad (79)$$

This final bound characterizes the probability that our trained LLM fails to assign sufficient probability to the desired completion. The bound decays with the square root of the sample size n , but depends on model complexity $\tilde{D}$ and scales inversely with the threshold parameter τ .

Recursive Bound for the Second Generation Model $\mathcal{G}_{\theta_2}^{\text{LLM}}$

Definition of Verified Prefix Set for $\mathcal{G}_{\theta_2}^{\text{LLM}}$

We now extend our analysis to the second generation model by defining a subset of prefix-conditioned inputs that have sufficient verified coverage from the first model:

$$\mathcal{D}_2^+ := \left\{ S \in \mathcal{S} \mid \mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S) \geq \frac{\log \frac{2}{\tau}}{N} \right\},$$

where the verified high-reward candidate set $\mathbf{y}_\gamma^+(S)$ contains tokens with near-optimal self-reward:

$$\mathbf{y}_\gamma^+(S) = \left\{ y \in \mathcal{X} \mid s_{\text{verify}}(y \mid S) \geq \max_{y' \in \mathcal{X}} s_{\text{verify}}(y' \mid S) - \gamma \right\}.$$

Bounding the Low-Confidence Region for $\mathcal{G}_{\theta_2}^{\text{LLM}}$

Applying the same total variation-based analysis as for the first-step model, we can bound the probability that $\mathcal{G}_{\theta_2}^{\text{LLM}}$ assigns low confidence to verified candidate regions:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_2}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S) \leq 1 - \tau, S \in \mathcal{D}_2^+] \leq \frac{4}{\tau} \cdot D_{\text{TV}}(\mathcal{G}_{\theta_2}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{2,N}^*(\cdot \mid S)).$$

Bounding the Failure Rate of Verified Prefixes

To complete our analysis, we need to bound the probability that a prefix falls outside the verified region $\mathcal{D}_2^+$. Using Markov's inequality, we have:

$$\mathbb{P}_{S \sim \mathcal{S}}[S \notin \mathcal{D}_2^+] = \mathbb{P}_{S \sim \mathcal{S}} \left[\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S) < \frac{\log \frac{2}{\tau}}{N} \right] \quad (80)$$

$$= \mathbb{P}_{S \sim \mathcal{S}} \left[\frac{\log \frac{2}{\tau}}{N \cdot \mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S)} > 1 \right] \quad (81)$$

$$\leq \frac{\log \frac{2}{\tau}}{N} \cdot \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S)} \right]. \quad (82)$$

This leads us to define the reciprocal of the model confidence as a random variable:

$$Z(S) := \frac{1}{\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S)}.$$

Using the tail-integral identity for expectations:

$$\mathbb{E}_{S \sim \mathcal{S}}[Z(S)] = \int_0^\infty \mathbb{P}(Z(S) \geq t) dt,$$

and combining with our PAC-Bayes bounds on the total variation distance between $\mathcal{G}_{\theta_1}^{\text{LLM}}$ and $\mathcal{G}_{1,N}^*$, we can derive:

$$C_{\gamma,1} := \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{y}_\gamma^+(S) \mid S)} \right] \lesssim 1 + A \frac{1}{C_0} \log \left(\frac{1}{A} \right) + C \frac{1}{C_0} \log \left(\frac{1}{C_0} \right),$$

where:

- $A = \sqrt{\frac{1}{n}} \log(n\tilde{B})\tilde{D} \log \frac{1}{\delta}$ represents the statistical error from PAC-Bayes bounds
- $C = \frac{C_\gamma}{N}$ is the approximation error from the confidence threshold
- C_0 represents the minimum prediction confidence of $\mathcal{G}_{\theta_1}^{\text{LLM}}$ across verified high-reward candidates

This bound quantifies how low-confidence predictions propagate into the next generation model.

Solving the Recurrence Relation

By extending our analysis recursively, similar to the proof of Theorem 1, we can derive a recurrence relation for the expected reciprocal of model confidence after t generations. The solution to this recurrence is:

$$\begin{aligned}
C_{\gamma,t} &\lesssim 1 + A_t \frac{1}{C_0} \log \left(\frac{1}{A_t} \right) + C_t \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\
&\lesssim 1 + \frac{1}{C_0} D_{\text{TV}} \log \left(\frac{1}{D_{\text{TV}}} \right) \\
&\quad + \left(\left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1 + \frac{D_{\text{TV}}}{C_0} \log \left(\frac{1}{D_{\text{TV}}} \right)}{N - \frac{1}{C_0} \log \left(\frac{1}{C_0} \right)} \right) \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \\
&\lesssim 1 + \frac{1}{C_0} \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \log(\sqrt{n}) \\
&\quad + \left(\left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^t \frac{C_\gamma}{N} + \frac{1}{N} \right) \cdot \frac{1}{C_0} \log \left(\frac{1}{C_0} \right) \tag{83}
\end{aligned}$$

This expression shows how the error compounds over iterations, with three main components: 1. A statistical error term that decreases with sample size n 2. An approximation error term that decreases with the number of samples N used in self-training 3. A recursively propagating error term that decays exponentially with the number of generations t

Final Low-Confidence Bound After T Recursive Generations

After T recursive updates, with probability at least $1 - \delta$, the probability that the final model assigns low confidence to its verified top candidates is bounded by:

$$\begin{aligned}
\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau] &\lesssim \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) \right) \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \\
&\quad + \left(\frac{1}{N} + \frac{1}{N^2 C_0} \right) \log \left(\frac{1}{\tau} \right) + \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^T \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right). \tag{84}
\end{aligned}$$

This bound reveals that as T increases, the third term vanishes exponentially, demonstrating that self-training eventually eliminates the initial confidence errors, provided that $\frac{1}{NC_0} < 1$, which is typically satisfied in practice.

Final Accuracy and Error Decomposition

To translate these confidence bounds into concrete accuracy guarantees, we define:

$$x_T^+(S) := \arg \max_{x \in \mathcal{X}} \mathcal{G}_{\theta_T}^{\text{LLM}}(x \mid S), \quad x^*(S) := \text{ground-truth token following } S.$$

The accuracy of generation after T steps is:

$$\text{Acc}(\mathcal{G}_{\theta_T}^{\text{LLM}}) := \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) = x^*(S)],$$

By decomposing this probability and using our previous bounds, we obtain:

$$\begin{aligned}
\text{Acc}(\theta_T) &= \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) = x^*(S) \mid \mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \tau] \cdot \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \tau] \\
&\gtrsim 1 - \epsilon(\gamma) - (1 - \epsilon(\gamma)) \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) \right) \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \\
&\quad - (1 - \epsilon(\gamma)) \left(\frac{1}{N} + \frac{1}{N^2 C_0} \right) \log \left(\frac{1}{\tau} \right) + \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^T \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right) \tag{85}
\end{aligned}$$

The corresponding test error is therefore:

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) := 1 - \text{Acc}(\mathcal{G}_{\theta_T}^{\text{LLM}}) = \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) \neq x^*(S)],$$

Which satisfies:

$$\begin{aligned} \text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) &\lesssim \epsilon(\gamma) + \left(\frac{1}{\tau} + \frac{1}{NC_0} \log \left(\frac{1}{\tau} \right) \right) \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \\ &\quad + \left(\frac{1}{N} + \frac{1}{N^2 C_0} \right) \log \left(\frac{1}{\tau} \right) + \left(\frac{1}{NC_0} \log \left(\frac{1}{C_0} \right) \right)^T \frac{C_\gamma}{N} \log \left(\frac{1}{\tau} \right). \end{aligned} \quad (86)$$

This error bound consists of four components: 1. An inherent approximation error $\epsilon(\gamma)$ from the γ -approximation of optimal reward 2. A statistical error from finite training data that scales as $O(1/\sqrt{n})$ 3. An approximation error from finite self-training samples that scales as $O(1/N)$ 4. A rapidly diminishing initial error term that decays as $O((1/NC_0)^{T+1})$

This analysis demonstrates how recursive self-training progressively improves model performance, with the error bound decreasing as both the pre-training dataset size n and the self-training dataset size N increase. Moreover, the benefit of additional recursive generations T is most pronounced in early iterations, with diminishing returns as T increases.

□

I Exponential Error Growth in Naive Recursive Training of Transformer-Based LLMs

In this section, we analyze the failure mode of naive recursive training applied to Transformer-based large language models (LLMs), where models are trained exclusively on synthetic data generated by their predecessors without any verification or external supervision. This setting, though increasingly common due to the scarcity of real data, introduces significant risk: even small inaccuracies in early generations can compound over time, causing the model to drift progressively further from the original data distribution.

We formalize this phenomenon by characterizing how prediction errors propagate across generations. Specifically, we show that the absence of quality control leads to exponential error growth, ultimately resulting in model collapse. This result highlights the inherent instability of unverified recursive training and underscores the necessity of intervention strategies such as real data anchoring or self-verification. We now present the formal theorem and provide its proof.

Theorem 7 (Error Propagation for Transformer-based LLMs in Naive Recursive Training). *Let $\mathcal{G}_{\theta_t}^{\text{LLM}}(x | S)$ denote the conditional generation distribution of an autoregressive Transformer-based LLM at recursive step t , where $S \in \mathcal{S} \subseteq \mathcal{X}^{\leq \ell}$ is a prefix sequence. Suppose the model is trained via naive recursive generation without any self-verification mechanism. Then, with probability at least $1 - \delta$, the prediction error of $\mathcal{G}_{\theta_T}^{\text{LLM}}$ is bounded as:*

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} + \left(C_\gamma + \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T,$$

where $C_\gamma := \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_T^+(S) | S)} \right]$ denotes the expected inverse confidence assigned to top-reward candidates under the initial model.

Note that since $C_\gamma > 1$, the error bound includes an exponentially growing term of the form

$$\left(C_\gamma + \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T.$$

As the number of recursive steps $T \rightarrow \infty$, this term can dominate the total error, reflecting the compounding effect of unfiltered distributional shift. Without confidence-based filtering, the model may gradually amplify its own errors, eventually leading to model collapse.

Proof of Theorem 7. We establish the following result by analyzing how prediction errors propagate across recursive training iterations in the absence of verification mechanisms. While our proof strategy parallels the one used in Theorem 5, it introduces key modifications to account for the lack of confidence-based filtering or quality control during training.

Foundational Definitions and Error Decomposition

We begin by defining the subset of prefix-token inputs for which the base model $\mathcal{G}_{\theta_0}^{\text{LLM}}$ assigns sufficient probability mass to its top predictions:

$$\mathcal{S}_1^+ := \left\{ S \in \mathcal{S} \mid \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \frac{\tau}{2} \right\},$$

where the high-reward candidate set is defined via an internal self-reward mechanism:

$$\mathbf{x}_\gamma^+(S) := \left\{ x \in \mathcal{X} \mid r_{\text{self}}(x \mid S) \geq \max_{x' \in \mathcal{X}} r_{\text{self}}(x' \mid S) - \gamma \right\},$$

and the self-reward function is given by:

$$r_{\text{self}}(x \mid S) := \log \mathcal{G}_{\theta_0}^{\text{LLM}}(x \mid S).$$

We aim to analyze the likelihood that the model $\mathcal{G}_{\theta_1}^{\text{LLM}}$, trained recursively without any verification mechanism, assigns insufficient probability mass to its own top candidates. We decompose this event as:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau] &\leq \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau, S \in \mathcal{S}_1^+] \\ &\quad + \mathbb{P}_{S \sim \mathcal{S}} [S \notin \mathcal{S}_1^+]. \end{aligned} \quad (87)$$

Analysis of Model Divergence via Total Variation Distance

To quantify the discrepancy between consecutive model generations in the absence of verification, we analyze their divergence using the total variation (TV) distance. For any input prefix $S \in \mathcal{S}$ and output token $x \in \mathcal{X}$, we have:

$$|\mathcal{G}_{\theta_1}^{\text{LLM}}(x \mid S) - \mathcal{G}_{\theta_0}^{\text{LLM}}(x \mid S)| \leq 2D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{\theta_0}^{\text{LLM}}(\cdot \mid S)).$$

This inequality can be inverted to obtain a lower bound on the TV distance between generations:

$$D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{\theta_0}^{\text{LLM}}(\cdot \mid S)) \geq \frac{1}{2} |\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) - \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S)|.$$

For prefix contexts in the high-confidence region $\mathcal{S}_1^+$, the base model assigns:

$$\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \geq 1 - \frac{\tau}{2}.$$

If the updated model assigns insufficient mass to this set, i.e.,

$$\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau,$$

then the discrepancy is bounded below by:

$$|\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) - \mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S)| \geq \frac{\tau}{2}.$$

Consequently, the probability that the updated model fails to retain high confidence on verified inputs can be bounded by the expected TV divergence:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) \mid S) \leq 1 - \tau, S \in \mathcal{S}_1^+] \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot \mid S), \mathcal{G}_{\theta_0}^{\text{LLM}}(\cdot \mid S))]. \quad (88)$$

Characterizing the Complement Set via Inverse Confidence

Let $C_\gamma := \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S)} \right]$ denote the expected inverse likelihood mass assigned to high-reward candidate sets under the base model. By applying Markov's inequality, we obtain:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}}[S \notin \mathcal{S}_1^+] &= \mathbb{P}_{S \sim \mathcal{S}} \left[\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) < 1 - \frac{\tau}{2} \right] \\ &= \mathbb{P}_{S \sim \mathcal{S}} \left[\frac{1 - \frac{\tau}{2}}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S)} > 1 \right] \\ &\leq \left(1 - \frac{\tau}{2}\right) \cdot \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S)} \right] \\ &= \left(1 - \frac{\tau}{2}\right) C_\gamma. \end{aligned} \quad (89)$$

Combining equations (35) and (36) into the decomposition (34), we arrive at the following bound for the first recursive generation step:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \leq 1 - \tau] \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S), \mathcal{G}_{\theta_0}^{\text{LLM}}(\cdot | S))] + \left(1 - \frac{\tau}{2}\right) C_\gamma. \quad (90)$$

Furthermore, from the proof of theorem 3, we know that:

$$\mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S), \mathcal{G}_{\theta_0}^{\text{LLM}}(\cdot | S))] \lesssim \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \quad (91)$$

Thus, we have

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \leq 1 - \tau] \lesssim \frac{1}{\tau} \cdot \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} + \left(1 - \frac{\tau}{2}\right) C_\gamma. \quad (92)$$

Recursive Analysis for the Second Generation Model $\mathcal{G}_{\theta_2}^{\text{LLM}}$

Following an analogous approach, we define the high-confidence support region for the second generation model:

$$\mathcal{S}_2^+ := \left\{ S \in \mathcal{S} \mid \mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \geq 1 - \frac{\tau}{2} \right\}.$$

For inputs $S \in \mathcal{S}_2^+$, a discrepancy in predicted probability mass between $\mathcal{G}_{\theta_2}^{\text{LLM}}$ and $\mathcal{G}_{\theta_1}^{\text{LLM}}$ on high-reward candidates leads to the bound:

$$\mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_2}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \leq 1 - \tau, S \in \mathcal{S}_2^+] \leq \frac{4}{\tau} \cdot \mathbb{E}_{S \sim \mathcal{S}} [D_{\text{TV}}(\mathcal{G}_{\theta_2}^{\text{LLM}}(\cdot | S), \mathcal{G}_{\theta_1}^{\text{LLM}}(\cdot | S))] . \quad (93)$$

For the complement region $\mathcal{S} \setminus \mathcal{S}_2^+$, we apply Markov's inequality as before:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} [S \notin \mathcal{S}_2^+] &= \mathbb{P}_{S \sim \mathcal{S}} \left[\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) < 1 - \frac{\tau}{2} \right] \\ &\leq \left(1 - \frac{\tau}{2}\right) \cdot \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_1}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S)} \right]. \end{aligned} \quad (94)$$

Generalization to Arbitrary Generations

We now generalize the inverse-confidence recurrence to arbitrary recursive steps. Let $C_{\gamma,t} := \mathbb{E}_{S \sim \mathcal{S}} \left[\frac{1}{\mathcal{G}_{\theta_t}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S)} \right]$ denote the expected inverse confidence at generation t . Then the recurrence

relation is given by:

$$\begin{aligned} C_{\gamma,t} &\leq \frac{C_{\gamma,t-1} + A}{C_0} + A \log \frac{1}{A} + C_{\gamma,t-1} \log \frac{1}{C_0} \\ &\leq \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right) C_{\gamma,t-1} + A \log \frac{1}{A}, \end{aligned} \quad (95)$$

where $A := \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta}$ reflects the TV-induced generalization noise, and C_0 is the minimum confidence over all prefix-conditioned top predictions in earlier generations.

Solving this recurrence yields the following bound:

$$C_{\gamma,t} \lesssim \left(C_\gamma + A \log \frac{1}{A} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^t. \quad (96)$$

Asymptotic Error Analysis After T Generations

Finally, applying this bound to the error probability after T recursive generations, we obtain:

$$\begin{aligned} \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \leq 1 - \tau] \\ &\lesssim \frac{1}{\tau} \cdot A + (1 - \frac{\tau}{2}) \left(C_\gamma + A \log \frac{1}{A} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T \\ &\lesssim \frac{1}{\tau} \cdot \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} + \left(C_\gamma + \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T \end{aligned} \quad (97)$$

This bound highlights exponential sensitivity to both the minimum confidence C_0 and the number of recursive steps T , in the absence of verification mechanisms.

Final Accuracy and Error Decomposition

We define:

$$x_T^+(S) := \arg \max_{x \in \mathcal{X}} \mathcal{G}_{\theta_T}^{\text{LLM}}(x | S), \quad x^*(S) := \text{ground-truth token following } S.$$

Then the accuracy of generation after T steps is:

$$\text{Acc}(\mathcal{G}_{\theta_T}^{\text{LLM}}) := \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) = x^*(S)],$$

and satisfies:

$$\begin{aligned} \text{Acc}(\theta_T) &= \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) = x^*(S) | \mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \geq 1 - \tau] \cdot \mathbb{P}_{S \sim \mathcal{S}} [\mathcal{G}_{\theta_T}^{\text{LLM}}(\mathbf{x}_\gamma^+(S) | S) \geq 1 - \tau] \\ &\gtrsim 1 - \epsilon(\gamma) - (1 - \epsilon(\gamma)) \frac{1}{\tau} \cdot \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \\ &\quad - (1 - \epsilon(\gamma)) \left(C_\gamma + \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T \end{aligned} \quad (98)$$

The corresponding test error is:

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) := 1 - \text{Acc}(\mathcal{G}_{\theta_T}^{\text{LLM}}) = \mathbb{P}_{S \sim \mathcal{S}} [x_T^+(S) \neq x^*(S)],$$

and satisfies:

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \cdot \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} + \left(C_\gamma + \sqrt{\frac{1}{n}} \log(n\tilde{B}) \tilde{D} \log \frac{1}{\delta} \right) \left(\frac{1}{C_0} + \log \frac{1}{C_0} \right)^T. \quad (99)$$

The proof is completed. $\square$

J Mitigating Collapse via Real Data in Recursive Training of Transformer-Based LLMs

In the previous section, we demonstrated that naive recursive training of transformer-based LLMs—without any quality control or external supervision—leads to exponential error growth across generations. This failure arises from unfiltered distributional drift, which causes the model to gradually reinforce and amplify its own mistakes.

To mitigate this collapse, one natural strategy is to incorporate real data into each generation of the training loop. In this section, we formally analyze how the inclusion of ground-truth supervised examples affects recursive training dynamics. Our key result shows that introducing even a moderate proportion of real data can effectively suppress exponential error amplification, ensuring bounded prediction error across generations.

Specifically, we consider a setting where each generation $\mathcal{G}_t$ is trained on a mixture of synthetic and real data. Let $\alpha \in (0, 1)$ denote the proportion of real data retained at each step. We show that this mixing strategy induces a recursive contraction effect, reducing the sensitivity of the model to its own erroneous generations.

Theorem 8 (Error Bound for Transformer-based LLMs with Real Data). *Let $\mathcal{G}_{\theta_t}^{\text{LLM}}(x | S)$ denote the conditional distribution of a Transformer-based LLM at recursive step t , trained on a mixture of real and synthetic data with real data proportion $\alpha > 0$. Then, under the same notation and assumptions as in Theorem 7, and assuming $\alpha > 1 - \frac{1}{2(1/C_0 - 1)}$ and $T \rightarrow \infty$, the prediction error of $\mathcal{G}_{\theta_T}^{\text{LLM}}$ is, with probability at least $1 - \delta$, bounded as follows:*

$$\text{Err}(\mathcal{G}_T) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \sqrt{\frac{D \log(n\bar{B})}{n}} + \frac{1 - \alpha}{C_0} \sqrt{\frac{D}{n}} \log(n\bar{B}) \quad (100)$$

Compared to the exponential growth in Theorem 7, the presence of real data introduces a damping effect, effectively anchoring the model and limiting the accumulation of distributional shift.

We now present the proof of this result.

Proof of Theorem 8. This proof builds upon the key techniques established in the proofs of Theorem 2 and Theorem 7, particularly the decomposition of error into high-confidence and low-confidence regions, and the recursive control of inverse confidence via total variation distance. We present a concise yet complete version adapted to Transformer-based LLMs.

We analyze a recursive training process where the model $\mathcal{G}_{\theta_t}^{\text{LLM}}$ at generation t is trained on a mixture of real and synthetic data:

$$\mathcal{D}_t = \alpha \cdot \mathcal{D}_{\text{real}} + (1 - \alpha) \cdot \mathcal{D}_{\text{synthetic}},$$

and the induced mixture distribution is:

$$\mathcal{G}_t^{\text{mix}} = \alpha \cdot \mathcal{G}^* + (1 - \alpha) \cdot \mathcal{G}_{\theta_{t-1}}^{\text{LLM}}.$$

Decomposition of prediction error

We define the high-confidence region:

$$\mathcal{D}_t^+ := \left\{ x \in \mathcal{X} \mid \mathcal{G}_{\theta_t}^{\text{LLM}}(y^* | x) \geq 1 - \frac{\tau}{2(1 - \alpha)} \right\},$$

and decompose the total error:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_{\theta_t}^{\text{LLM}}(y^* | x) \leq 1 - \tau] \leq \mathbb{P} [\mathcal{G}_{\theta_t}^{\text{LLM}}(y^* | x) \leq 1 - \tau \mid x \in \mathcal{D}_t^+] + \mathbb{P} [x \notin \mathcal{D}_t^+].$$

Bounding the first term via total variation

Using the standard TV bound:

$$|\mathcal{G}_{\theta_t}^{\text{LLM}}(y | x) - \mathcal{G}_t^{\text{mix}}(y | x)| \leq 2D_{\text{TV}}(\mathcal{G}_{\theta_t}^{\text{LLM}}, \mathcal{G}_t^{\text{mix}}),$$

we get:

$$\mathbb{P} [\mathcal{G}_{\theta_t}^{\text{LLM}}(y_\gamma^+(x) | x) \leq 1 - \tau \mid x \in \mathcal{D}_t^+] \leq \frac{4}{\tau} D_{\text{TV}}(\mathcal{G}_{\theta_t}^{\text{LLM}}, \mathcal{G}_t^{\text{mix}}).$$

Bounding the complement using inverse confidence

Let $C_\gamma := \mathbb{E}_{x \sim \mathcal{X}} \left[\frac{1}{\mathcal{G}_{\theta_0}^{\text{LLM}}(\mathbf{y}_\gamma^+(x) | x)} \right]$. By Markov's inequality:

$$\mathbb{P}[x \notin \mathcal{D}_t^+] \leq \left(1 - \frac{\tau}{2(1-\alpha)}\right) C_\gamma.$$

Recursive coverage coefficient control

Assuming $\alpha > 1 - \frac{1}{2(1/C_0-1)}$, we have:

$$C_{\gamma,T} \lesssim \frac{1-\alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}.$$

Finite-sample generalization bounds

In Transformer-based LLMs, the finite-sample deviation under PAC-Bayes analysis is:

$$\mathbb{P}_{x \sim \mathcal{X}} [\mathcal{G}_T(\mathbf{y}_\gamma^+(x) | x) \leq 1 - \tau] \lesssim \frac{1}{\tau} \cdot D_{\text{TV}} + \frac{1-\alpha}{C_0} D_{\text{TV}} \log \frac{1}{D_{\text{TV}}}. \quad (101)$$

Final Results

Combining all terms, we derive the final error bound:

$$\text{Err}(\mathcal{G}_{\theta_T}^{\text{LLM}}) \lesssim \epsilon(\gamma) + \frac{1}{\tau} \sqrt{\frac{\bar{D} \log(n\bar{B})}{n}} + \frac{1-\alpha}{C_0} \sqrt{\frac{\bar{D}}{n}} \log(n\bar{B}),$$

where $\bar{D}$ is the model's effective dimension, $\bar{B}$ its prior radius, and n the total sample count. The proof is completed. $\square$

K Further Discussion on Assumptions and Empirical Validation

K.1 On the Realism of Assumption 1

This section aims to clarify both the intuition behind Assumption 1 and its alignment with real world LLM behavior.

At its core, Assumption 1 formalizes a confidence calibrated agreement condition: high confidence model predictions are likely to be correct. This is consistent with standard assumptions in self training (e.g., Assumption 2 on page 14 of Huang et al. [2025b]) and aligns with active research on confidence calibration in LLMs, which is a well studied area (see Liu et al. [2025] for a survey).

Critical to its realism, empirical work shows modern LLMs often exhibit strong calibration [Achiam et al., 2023]. For example, Luo et al. [2025] reports small expected calibration errors for pre trained models like Llama-3-8B (3.52%), Qwen-2.5-7B (5.41%), and DeepSeek-V2-Lite (3.39%) on MMLU. This demonstrates a close alignment between confidence and accuracy.

To further validate this, we conducted experiments on the MATH dataset with Phi3.5-Mini, analyzing the log probabilities of correct versus incorrect responses. The results are summarized in Table 1.

As the data shows, correct responses consistently exhibit higher log probabilities, and their distribution stochastically dominates that of incorrect responses. This observation confirms the positive correlation between model confidence and correctness, directly supporting the validity of Assumption 1.

K.2 Empirical Validation of Self-Verification

To empirically test our theoretical findings on model collapse, we designed an experiment following the setting from Fu et al. [2025]. Specifically, we trained a 12 layer, 8 head GPT-2 model (with a hidden size of 256) to recursively perform in context learning of linear functions from the class:

$$\mathcal{F} = \{f \mid f(x) = w^\top x, w \in \mathbb{R}^5\}$$

Table 1: Distribution of Correct and Incorrect Responses by Log Probability on the MATH dataset with Phi3.5-Mini.

Log Probability Range	Number of Correct	Number of Incorrect
0 to -0.6	1032	83
-0.6 to -1.2	413	462
-1.2 to -1.8	73	386
-1.8 to -2.4	24	95
-2.4 to -3.0	6	18

For each prompt, we sampled $x_1, \dots, x_k, x_{\text{query}}$ and w independently from $\mathcal{N}(0, I_d)$. The model’s task was to predict $y_{\text{query}} = w^\top x_{\text{query}}$. We compared three distinct training strategies:

- **Full Synthetic:** The model was trained solely on synthetic data generated by its predecessor.
- **Mixed:** A combination of fresh real data and synthetic data was used, mixed in a 0.5 ratio.
- **Verification-Based Filtering:** For each training instance, we sampled 20 candidate responses, selected the one with the highest confidence (corresponding to $\gamma = 0$), and trained the model exclusively on this verified data.

The results of these experiments, measured by prediction error over recursive training rounds, are summarized in Table 2.

Table 2: Prediction Error Across Recursive Training Rounds for Different Strategies.

Strategy	Round 0	Round 1	Round 2	Round 3
Full Synthetic	0.2418	1.1625	1.4245	1.9523
Mixed	0.2418	0.2824	0.3178	0.3235
Verification	0.2418	0.2672	0.2949	0.3105

As observed, the error accumulates progressively with more generations of recursive training. This degradation is particularly severe in the full synthetic case, where the error grows rapidly. In contrast, both incorporating real data and applying the self verification mechanism effectively mitigate the increase in loss. This outcome is consistent with our theoretical findings.

K.3 Directions for Future Experimental Work

To further validate and extend the theoretical contributions of this paper, several experimental directions could be pursued for future work:

1. **Calibration Manipulation Experiment:** An initial experiment could involve selecting a model for recursive training and systematically adjusting its calibration via temperature scaling or other established methods. This process would produce model versions with "low," "moderate," and "high" calibration, where the quality of calibration is quantified by the Expected Calibration Error (ECE).
2. **Performance Correlation Verification:** Following the calibration manipulation, a second experiment could investigate the relationship between calibration and performance degradation. The differently calibrated models would undergo recursive training with self-verification, allowing for a comparison of their error accumulation over successive generations. Such an analysis would provide empirical insights into how a verifier’s calibration quality (measured by ECE) correlates with the final model’s test error under a self-verification framework.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction clearly state three main claims: (1) that recursive training without intervention leads to exponential error growth, (2) that self-verification using internal confidence scores can prevent collapse without relying on real data, and (3) that this theoretical framework extends to Transformer-based LLMs. These claims are rigorously justified through formal theorems and finite-sample error bounds in Sections 4 and 5, and their limitations are discussed in Appendix C.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Appendix C discusses the main limitation of relying on a single confidence-based verification strategy and suggests directions for improving verification robustness.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theoretical results include clearly stated assumptions and are supported by complete formal proofs provided in Appendices.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper is purely theoretical and does not include any experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer:[NA]

Justification: The paper is purely theoretical and does not include any experimental results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: The paper is purely theoretical and does not include any experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: The paper is purely theoretical and does not include any experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: The paper is purely theoretical and does not include any experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The work is purely theoretical and adheres to all principles outlined in the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: Appendix discusses positive impacts, such as reducing reliance on human-labeled data through self-verification, and clarifies that the theoretical nature of the work poses no immediate negative societal risks.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not release any models or datasets and does not involve components that pose risks of misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use any existing code, datasets, or model assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not introduce or release any new datasets, code, or models.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve any crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve any research with human subjects and therefore does not require IRB approval.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The paper does not involve the use of LLMs as part of its core methods; all results are derived through theoretical analysis.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.