On the Price of Privacy for Language Identification and Generation

Xiaoyu Li
University of New South Wales
xiaoyu.li2@unsw.edu.au

Andi Han
University of Sydney
andi.han@sydney.edu.au

Jiaojiao Jiang
University of New South Wales
jiaojiao.jiang@unsw.edu.au

Junbin Gao
University of Sydney
junbin.gao@sydney.edu.au

Abstract

As large language models (LLMs) are increasingly trained on sensitive user data, understanding the fundamental cost of privacy in language learning becomes essential. We initiate the study of differentially private (DP) language identification and generation in the agnostic statistical setting, establishing algorithms and lower bounds that quantify the cost of privacy. For both tasks, approximate (ε, δ)-DP with constant $\varepsilon > 0$ and $\delta \geq 1/\operatorname{poly}(n)$ recovers the non-private error rates: $\exp(-r(n))$ for identification (for any $r(n) = o(n)$) and, given a public witness window, $\exp(-\Omega(n))$ for generation. Under pure ε-DP, the exponents of our upper bounds degrade by a multiplicative factor of $\min\{1, \varepsilon\}$ (for generation, given also a public mass floor); conversely, a private two-point inequality shows that for every n, every ε-DP algorithm errs with probability $\exp(-O(\min\{1, \varepsilon\} \cdot n))$ on one of two fixed distributions. For generation with a public witness window and mass floor, the two bounds match on suitable two-language collections, establishing the optimal rate $\exp(-\Theta(\min\{1, \varepsilon\} \cdot n))$. Our results show that, given such public side information for generation, the cost of privacy in language learning is mild: absent under approximate DP, and at most a $\min\{1, \varepsilon\}$ factor in the exponent under pure DP.

1 Introduction

Differential privacy [Dwork et al., 2006b] provides a mathematically rigorous framework for learning from sensitive data, yet its cost is not uniform across tasks: some learning problems can be solved privately at no statistical penalty, while others suffer an unavoidable degradation. Language identification and generation, as fundamental tasks in language learning, are a natural setting in which to ask what privacy costs, especially as LLMs are increasingly trained on sensitive data: differentially private fine-tuning sometimes approaches non-private baselines [Li et al., 2022, Yu et al., 2022], and the compute-privacy-utility tradeoff of private training has been characterized empirically [McKenna et al., 2025]. Yet the fundamental theoretical cost of privacy for these tasks remains largely unexplored. We address this gap by quantifying the price of privacy for both language identification and generation, with upper and lower bounds.

The formal study of language learnability traces back to the seminal work of Gold [1967], who introduced the *identification in the limit* model, and to Angluin [1979, 1980], who characterized identifiability for several important language classes. These classical results revealed fundamental barriers: for instance, even the class of all regular languages is not identifiable in the limit from positive examples alone. Recently, Kleinberg and Mullainathan [2024] proposed *generation* as an

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

alternative objective: rather than naming the target language, the learner need only produce a novel string consistent with the underlying distribution. This relaxation turns out to be dramatically more powerful: generation is achievable for every countable language collection, sidestepping the classical impossibility results for identification.

The aforementioned results all operate in an *online* setting, where the learner receives examples one at a time from an adversarially chosen sequence and must eventually converge to a correct hypothesis. Kalavasis et al. [2025] initiated the study of language learning in the *statistical* setting, where the learner receives an i.i.d. sample of fixed size drawn from some unknown distribution, and Høgsgaard and Pabbaraju [2026] extended this framework to the agnostic case, where the target distribution need not be supported on any single language in the collection.

Our setup follows this *agnostic statistical* language learning: A learner receives n i.i.d. samples $S = (x_1, \ldots, x_n) \sim \mathcal{D}^n$ drawn from an unknown distribution $\mathcal{D}$ over a countable universe $\mathcal{U}$, together with a countable collection $\mathcal{C} = \{L_1, L_2, \ldots\}$ of languages, where each language L_i is a subset of $\mathcal{U}$. This setting gives rise to two fundamental tasks:

- **Language Identification:** Output a language $\widehat{L} \in \mathcal{C}$ whose population risk $\Pr_{x\sim\mathcal{D}}[x \notin \widehat{L}]$ is close to the best achievable within $\mathcal{C}$.
- **Language Generation:** Output a string $\widehat{x} \in \mathcal{U}$ that is both *valid* ($\widehat{x} \in \operatorname{supp}(\mathcal{D})$) and *novel* ($\widehat{x} \notin S$) with high probability.

Without privacy constraints, Høgsgaard and Pabbaraju [2026] showed that identification error decays at a nearly exponential rate $\exp(-r(n))$ where r is any sublinear function (i.e., $r(n) = o(n)$) under an attainability condition on the agnostic optimum, and that generation error decays at a fully exponential rate $\exp(-\Omega(n))$ under mild structural assumptions. This raises a natural question:

What is the price of differential privacy for language identification and generation?

1.1 Main Results

Our answer is optimistic and we show: under *approximate* (ε, δ)-DP with constant $\varepsilon > 0$ (and $\delta \geq n^{-O(1)}$), privacy is free (for generation, given a public witness window); under *pure* ε-DP, the convergence exponent of our algorithms degrades by a multiplicative factor of $\min\{1, \varepsilon\}$, and our lower bounds show that this scaling cannot be avoided for error exponents of order n. Table 1 summarizes the results.

Table 1: Summary of results. Approximate DP assumes a constant $\varepsilon > 0$ and $\delta \geq n^{-O(1)}$. §With a public witness window, and under pure DP also a public mass floor; with a window but no floor, pure DP gives $\exp(-\Omega(\min\{1, \varepsilon\}\, r(n)))$ for $r(n) = o(n)$, and without a window the private bounds are conditional on coverage (Section 4). ⋆Holds for every $r(n) = o(n)$; the algorithm may depend on r. †Non-private results are due to Høgsgaard and Pabbaraju [2026]. ‡The lower bounds hold for every n (Section 5); the non-private lower bound applies to any approximate DP algorithm.

	Non-private†	**Pure ε-DP**	**Approx. (ε, δ)-DP**
Identification (UB)⋆	$\exp\big(-r(n)\big)$	$\exp\big(-\min\{1, \varepsilon\} \cdot r(n)\big)$	$\exp\big(-r(n)\big)$
Identification (LB) ‡	$\exp\big(-O(n)\big)$	$\exp\big(-\min\{1, \varepsilon\} \cdot O(n)\big)$	$\exp\big(-O(n)\big)$
Generation (UB)	$\exp\big(-\Omega(n)\big)$	$\exp\big(-\min\{1, \varepsilon\} \cdot \Omega(n)\big)^{\S}$	$\exp\big(-\Omega(n)\big)^{\S}$
Generation (LB) ‡	$\exp\big(-O(n)\big)$	$\exp\big(-\min\{1, \varepsilon\} \cdot O(n)\big)$	$\exp\big(-O(n)\big)$

Two key takeaways stand out. First, approximate DP eliminates the privacy cost entirely for both tasks (for generation, given a public witness window), yielding a separation from pure DP that becomes qualitative as $\varepsilon \to 0$. Second, generation enjoys a *tighter* privacy-utility tradeoff than identification: with a public witness window and mass floor, the pure DP upper and lower bounds match up to constants in the exponent (Remark 5.6), whereas identification retains an $o(n)$ vs. $O(n)$ gap inherited from the non-private setting. Høgsgaard and Pabbaraju [2026] provide the non-private framework, algorithms, and rates, and the exponential and Gaussian mechanisms and group privacy are standard DP tools. Our contribution is the construction and analysis of low-sensitivity objectives that preserve the non-private decision logic, together with privacy-specific lower bounds:

- **DP Identification (Section 3).** The non-private algorithm employs a margin-based selection rule that is discontinuous: changing a single sample can flip which indices satisfy the margin constraint. We replace it with a *smooth score function* that jointly encodes the preference for large language indices and the penalty for failing the margin test, and privatize it via the exponential mechanism (pure DP); for approximate DP, the Gaussian mechanism perturbs the empirical risks instead. The score has sensitivity at most $2f(n)^2/n$, where $f(n) \to \infty$ is a growing horizon, yet the score gap on good events is $\Omega(1)$. The function f must be carefully chosen to balance the resulting bias–variance–privacy tradeoff, which yields the rates in Table 1.
- **DP Generation (Section 4).** The non-private pointer-based rule has unbounded sensitivity: a single sample change can cause a language's pointer to jump arbitrarily. We replace it with a *thresholded prefix count*, defined as the minimum number of times each relevant string appears in the sample, whose sensitivity is a *constant*, independent of n and $f(n)$. We again privatize via the exponential mechanism (pure DP) and the Gaussian mechanism (approximate DP). This structural advantage is the key reason generation achieves strictly better rates: with a public witness window and mass floor, the score gap grows as $\Omega(n)$ while the sensitivity stays $O(1)$, yielding a fully exponential rate $\exp(-\min\{1,\varepsilon\} \cdot \Omega(n))$ under pure DP. We present the algorithm first with a public witness bound, then extend it to the setting where no such bound is available, where the guarantee is conditional on coverage.
- **Lower Bounds** (Section 5). We prove that, for error exponents of order n, the $\min\{1,\varepsilon\}$ scaling is necessary for both tasks. For each task, we construct a pair of hard distributions that are close in Hamming distance under a natural coupling, then apply a private two-point inequality, $\Pr[\mathcal{A}(S) \notin \mathcal{F}] + \Pr[\mathcal{A}(S') \in \mathcal{F}] \geq \mathbb{E}[e^{-\varepsilon\, \mathrm{d_{Ham}}(S,S')}]$, which follows from group privacy, to show that for every n any ε-DP algorithm must incur error at least $\exp(-\min\{1,\varepsilon\} \cdot O(n))$ on one of them. The two arguments differ in structure. For identification, the construction is symmetric: the two distributions swap the roles of two languages, and the inequality constrains the misidentification probability in both directions simultaneously. For generation, the construction is inherently asymmetric: the two distributions share a common high-probability element but have disjoint "private" supports, so that any private-support string witnessing success under one distribution necessarily witnesses failure under the other. For generation with a public window and mass floor, the resulting lower bound matches the upper bound up to constants in the exponent on suitable collections (Remark 5.6), establishing an optimal rate of $\exp(-\Theta(\min\{1,\varepsilon\} \cdot n))$.

Why pure DP pays $\min\{1,\varepsilon\}$ and approximate DP pays nothing. In each upper bound, the correct candidates win by a margin Γ under a score of sensitivity Δ, so the exponential mechanism errs with probability at most $|\mathcal{R}|\, e^{-\varepsilon\Gamma/(2\Delta)}$ (Lemma 2.3). Our scores make Γ/Δ of the order of the statistical exponent ($n/(2f(n)^2)$ against $n/(8f(n)^2)$ for identification; $g(n)$ against $np^\star_W/8$ for generation, with $g(n) = \Theta(n)$ given a mass floor), so pure DP keeps a fraction $\Theta(\min\{1,\varepsilon\})$ of it, while the quadratic tails of Gaussian noise make the approximate-DP cost negligible.

1.2 Related Work

Language Identification and Generation. The formal study of language learnability was initiated by Gold [1967] and characterized by Angluin [1979, 1980]. Kleinberg and Mullainathan [2024] introduced the weaker notion of *generation in the limit*, sparking a rich line of follow-up work on diversity–hallucination tradeoffs, noise robustness, and computational barriers [Kalavasis et al., 2025, 2026, Charikar and Pabbaraju, 2025, Kleinberg and Wei, 2025, 2026a, Raman and Raman, 2025, Bai et al., 2026, Mehrotra et al., 2026a, Arenas et al., 2026]. Kalavasis et al. [2025] and Høgsgaard and Pabbaraju [2026] transitioned the problem to the statistical setting: the former under realizability, the latter in the agnostic case that we adopt. Concurrent work by Mehrotra et al. [2026b] studies pure DP generation and identification in the limit for realizable sequences under continual release; this setting differs from our agnostic one-shot setting, so the two are complementary rather than directly comparable. We defer the more comprehensive review, and a rate-and-assumption comparison with other private learning tasks, to Appendix H.

Private Hypothesis Selection and PAC Learning. Our identification algorithms can be viewed as private model selection over a *countably infinite* class with a growing horizon $f(n) \to \infty$, extending the finite-class setting of Bun et al. [2019] and Gopi et al. [2020]; the growing horizon creates a bias–variance–privacy tradeoff, and our score has sensitivity at most $2f(n)^2/n$. Our generation

algorithms exploit a score with constant sensitivity, yielding tighter rates given a public witness window; our lower bounds refine the coupling and group-privacy framework of Acharya et al. [2021] into a private two-point inequality. A central question in private learning theory is whether privacy degrades statistical performance. Alon et al. [2019] and Bun et al. [2020] showed that approximate DP learnability is equivalent to online learnability, while pure DP imposes a strictly stronger requirement; at the level of rates, Smith [2011], Feldman and Xiao [2015], Bun et al. [2015] established that approximate DP often preserves non-private rates whereas pure DP can incur an arbitrarily larger sample cost. Separations of this kind have been demonstrated for private learning [Beimel et al., 2013], counting queries [Bun et al., 2014], and marginal estimation [Steinke and Ullman, 2017]. Our results contribute a new instance of this phenomenon in the language learning setting: approximate DP recovers non-private rates for both tasks (for generation, given a public witness window), while under pure DP our exponents lose a factor $\min\{1, \varepsilon\}$ that the lower bounds show cannot be avoided at the linear scale.

2 Preliminaries

Notation. Let $\mathbb{N}$ denote the positive integers and $[n] := \{1, \dots, n\}$ for an $n \in \mathbb{N}$. For $a \in \mathbb{R}$, we write $a_+ := \max\{a, 0\}$. We use $\|\cdot\|_2$ for the ℓ_2-norm, $\mathcal{N}(\mu, \Sigma)$ for the multivariate Gaussian, and $\mathbf{1}\{\cdot\}$ for the indicator function. Given a distribution $\mathcal{D}$ over $\mathcal{U}$, its support is $\mathrm{supp}(\mathcal{D}) := \{x \in \mathcal{U} : \mathcal{D}(x) > 0\}$. We use standard asymptotic notation $O, \Omega, \Theta, o, \omega$ and write $\lesssim, \gtrsim, \asymp$ as synonyms for O, Ω, Θ respectively.

2.1 Language Identification and Generation

Let $\mathcal{U} = \{u_1, u_2, \dots\}$ be a countable universe of strings. A *language* is any subset $L \subseteq \mathcal{U}$, and a *language collection* $\mathcal{C} \subseteq 2^{\mathcal{U}}$ is indexed as $\mathcal{C} = \{L_1, L_2, \dots\}$ when countable. Let $\mathcal{D}$ be an unknown distribution over $\mathcal{U}$. For any language L, its *population risk* is $\mathrm{err}_{\mathcal{D}}(L) := \Pr_{x\sim\mathcal{D}}[x \notin L]$, and its *empirical risk* on a sample $S = (x_1, \dots, x_n)$ is $\mathrm{err}_S(L) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \notin L\}$.

We work in the agnostic statistical setting of Høgsgaard and Pabbaraju [2026].

Language Identification. An identification algorithm $\mathcal{A}^{\mathrm{id}}$ (possibly using randomness r) observes $S = (x_1, \dots, x_n) \sim \mathcal{D}^n$ and outputs a language $\mathcal{A}^{\mathrm{id}}(S) \in \mathcal{C}$. Its *identification error* is the expected excess risk over the agnostic optimum:

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}, \mathcal{D}, \mathcal{C}, n) := \mathop{\mathbb{E}}_{S\sim\mathcal{D}^n, r}\big[\mathrm{err}_{\mathcal{D}}(\mathcal{A}^{\mathrm{id}}(S))\big] - \inf_{L\in\mathcal{C}} \mathrm{err}_{\mathcal{D}}(L).$$

The final version of Høgsgaard and Pabbaraju [2026] calls this objective *weak identification*.

Language Generation. A generation algorithm $\mathcal{A}^{\mathrm{gen}}$ (possibly using randomness r) observes $S \sim \mathcal{D}^n$ and outputs a string $\mathcal{A}^{\mathrm{gen}}(S) \in \mathcal{U}$ that should be both *valid* (in $\mathrm{supp}(\mathcal{D})$) and *novel* (not in S). Its *generation error* is the failure probability:

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}, \mathcal{D}, \mathcal{C}, n) := \Pr_{S\sim\mathcal{D}^n, r}\big[\mathcal{A}^{\mathrm{gen}}(S) \notin \mathrm{supp}(\mathcal{D}) \setminus S\big].$$

2.2 Differential Privacy

Two datasets $S, S' \in \mathcal{U}^n$ are *neighboring*, written $S \sim S'$, if they differ in exactly one entry.

Definition 2.1 (Differential Privacy [Dwork et al., 2006b])**.** A randomized algorithm $\mathcal{A}\colon \mathcal{U}^n \to \mathcal{R}$ is (ε, δ)-differentially private if for every pair of neighboring datasets $S \sim S'$ and every measurable $\mathcal{F} \subseteq \mathcal{R}$, $\Pr[\mathcal{A}(S) \in \mathcal{F}] \le e^{\varepsilon} \cdot \Pr[\mathcal{A}(S') \in \mathcal{F}] + \delta$. When $\delta = 0$ we say $\mathcal{A}$ is ε-DP (*pure* DP); when $\delta > 0$ we say (ε, δ)-DP (*approximate* DP), typically with $\delta \le 1/\mathrm{poly}(n)$.

Lemma 2.2 (Post-Processing [Dwork and Roth, 2014])**.** *If $\mathcal{A}$ is (ε, δ)-DP and h is any (possibly randomized) function, then $h \circ \mathcal{A}$ is (ε, δ)-DP.*

Exponential mechanism. The exponential mechanism [McSherry and Talwar, 2007] is the canonical tool for privately optimizing over a discrete set. Let $q\colon \mathcal{U}^n \times \mathcal{R} \to \mathbb{R}$ be a score function over a finite set $\mathcal{R}$, and let $\Delta q > 0$ bound its sensitivity: $|q(S, i) - q(S', i)| \le \Delta q$ for all $i \in \mathcal{R}$ and all $S \sim S'$. The exponential mechanism $\mathrm{EM}_{\varepsilon}(S, q, \mathcal{R})$ selects and outputs $i \in \mathcal{R}$ with probability $\propto \exp\Big(\frac{\varepsilon}{2\Delta q}\, q(S, i)\Big)$.

Lemma 2.3 (McSherry and Talwar [2007], Dwork and Roth [2014]). *The exponential mechanism* $\mathrm{EM}_\varepsilon(S, q, \mathcal{R})$ *is* ε*-DP. Moreover, fix* S *and a set* $\mathcal{G} \subseteq \mathcal{R}$ *of good outputs with* margin $\Gamma := \max_{i\in\mathcal{G}} q(S,i) - \max_{i\in\mathcal{R}\setminus\mathcal{G}} q(S,i) > 0$. *Then* $\Pr[\mathrm{EM}_\varepsilon(S,q,\mathcal{R}) \notin \mathcal{G}] \le |\mathcal{R}\setminus\mathcal{G}|\, e^{-\varepsilon\Gamma/(2\Delta q)}$ *(Lemma A.5).*

Gaussian Mechanism. We make use of Gaussian mechanism [Dwork et al., 2006a] to design our approximate-DP algorithms. Given $T\colon \mathcal{U}^n \to \mathbb{R}^d$ with ℓ_2-sensitivity $\Delta_2 := \max_{S\sim S'} \|T(S) - T(S')\|_2$, the Gaussian mechanism releases $T(S) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I_d)$.

Lemma 2.4 (Dwork et al. [2006a], Dwork and Roth [2014]). *For* $\varepsilon, \delta \in (0,1)$*, the Gaussian mechanism is* (ε,δ)*-DP if* $\sigma \ge \frac{\Delta_2}{\varepsilon}\sqrt{2\log(1.25/\delta)}$.

An (ε_0,δ)-DP algorithm is (ε,δ)-DP for every $\varepsilon \ge \varepsilon_0$, so for $\varepsilon \ge 1$ we run our approximate-DP algorithms with $\varepsilon_0 = 1/2$.

3 Differentially Private Language Identification

We design differentially private identification algorithms that, given $S \sim \mathcal{D}^n$, output a language $L_{\hat{i}} \in \mathcal{C}$ whose population risk is nearly as small as the agnostic optimum. We work under the following assumption, necessary for fast convergence even without privacy [Høgsgaard and Pabbaraju, 2026, Theorem 2.3].

Assumption 3.1 (Agnostic optimum is attainable). *There exists an index* $i^\star \in \mathbb{N}$ *such that* $\mathrm{err}_\mathcal{D}(L_{i^\star}) = \inf_{L\in\mathcal{C}} \mathrm{err}_\mathcal{D}(L)$. *Let* $i^\star$ *denote the smallest such index.*

In the realizable case some language has risk 0, so Assumption 3.1 holds and all results of this section apply.

Score function design. The non-private algorithm of Høgsgaard and Pabbaraju [2026] selects the largest index $i \in [f(n)]$ whose empirical risk beats all predecessors by more than $2/f(n)$, where $f(n) \to \infty$ is a growing horizon. A natural first attempt would use the empirical risk directly as the score of the exponential mechanism. However, the rule must select the *largest* index that passes the margin test, not the one with minimum empirical risk, and the test is discontinuous in the data: changing a single sample can flip an index from feasible to infeasible.

We resolve this by replacing the hard feasibility test with a *soft score* that continuously encodes both objectives. For each $i \in [f(n)]$, define the empirical margin

$$M_S(i) := \begin{cases} 1, & i = 1, \\ \min_{j\in[i-1]} \big(\mathrm{err}_S(L_j) - \mathrm{err}_S(L_i)\big), & i \ge 2, \end{cases}$$

the deficit $d_S(i) := \big(\frac{2}{f(n)} - M_S(i)\big)_+$, and the score $q(S,i) := i - f(n)^2 d_S(i)$. The term i rewards larger indices, and the penalty $-f(n)^2 d_S(i)$ continuously downweights indices that fail the margin test; the coefficient $f(n)^2$ makes the penalty exceed every positional reward once the deficit is of order $1/f(n)$.

3.1 Pure DP Identification

Algorithm 1 samples $\hat{i}$ from the exponential mechanism with score q over $[f(n)]$.

Algorithm 1 Pure DP Identification $\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}$

Require: A dataset $S = (x_1,\dots,x_n) \in \mathcal{U}^n$, a privacy parameter $\varepsilon > 0$, a function f, a language collection $\mathcal{C} = \{L_1, L_2, \dots\}$

1: Compute $\mathrm{err}_S(L_i)$, $M_S(i)$, $d_S(i)$, and $q(S,i)$ for every $i \in [f(n)]$

2: **return** $\hat{i} \sim \mathrm{EM}_\varepsilon(S, q, [f(n)])$ with sensitivity bound $\Delta q = 2f(n)^2/n$

Theorem 3.2 (Pure DP Identification). *Let* $\mathcal{C}$ *be a countable language collection,* $\mathcal{D}$ *any distribution over* $\mathcal{U}$ *satisfying Assumption 3.1,* $\varepsilon > 0$*, and* $f\colon \mathbb{N} \to \mathbb{N}$ *with* $f(n) \to \infty$. *Then Algorithm 1 is*

ε-differentially private and, for all $n \geq n_0$, where n_0 depends on $\mathcal{C}$, $\mathcal{D}$ and f but not on ε,

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) \leq \underbrace{2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big)}_{\textit{statistical term}} + \underbrace{f(n)\exp\Big(-\frac{\varepsilon n}{4f(n)^2}\Big)}_{\textit{privacy term}}.$$

Proof idea. Let $\mathcal{E}$ be the event that $|\mathrm{err}_S(L_i) - \mathrm{err}_{\mathcal{D}}(L_i)| \leq 1/(4f(n))$ for every $i \in [f(n)]$; Hoeffding's inequality and a union bound give $\Pr[\mathcal{E}^c] \leq 2f(n)e^{-n/(8f(n)^2)}$. On $\mathcal{E}$, the index $i^\star$ passes the margin test, while every $i > i^\star$ fails it with deficit at least $3/(2f(n))$; hence $q(S, i^\star) = i^\star$ and every other index scores at most $i^\star - 1$. The score has sensitivity at most $2f(n)^2/n$, so Lemma 2.3 with margin $\Gamma = 1$ selects some $i \neq i^\star$ with probability at most $f(n)e^{-\varepsilon n/(4f(n)^2)}$. The full proof is in Appendix C.

When $\varepsilon \geq 1/2$, the privacy term is at most the statistical term and privacy is essentially free; when $\varepsilon < 1/2$, the privacy term dominates and the exponent shrinks by the factor 2ε. Choosing $f(n) = \lceil \sqrt{c \log n} \rceil$ for a constant $c > 0$ gives $\mathrm{IdErr} \leq \exp(-\Omega(\min\{1, \varepsilon\} \cdot n/\log n))$, and for every $r(n) = o(n)$ a suitable horizon gives $\mathrm{IdErr} \leq \exp(-\min\{1, \varepsilon\} \cdot r(n))$ for all large n (Corollary C.9).

3.2 Approximate DP Identification

For approximate DP, we replace the exponential mechanism with the Gaussian mechanism: in Algorithm 2, we privately release a noisy empirical error vector $(\widetilde{\mathrm{err}}_S(L_i))_{i \in [f(n)]}$ by adding independent Gaussian noise to each coordinate, then apply the deterministic margin-selection rule to the noisy margin $\widetilde{M}_S(i) := \min_{j \in [i-1]}(\widetilde{\mathrm{err}}_S(L_j) - \widetilde{\mathrm{err}}_S(L_i))$, with $\widetilde{M}_S(1) := 1$. By post-processing, the algorithm is differentially private.

Algorithm 2 Approximate DP Identification $\mathcal{A}^{\mathrm{id}}_{\varepsilon,\delta,f}$

Require: A dataset $S \in \mathcal{U}^n$, privacy parameters $\varepsilon, \delta \in (0,1)$, a function $f : \mathbb{N} \to \mathbb{N}$, a language collection $\mathcal{C}$

1: $\sigma \leftarrow \frac{\sqrt{f(n)}}{\varepsilon n}\sqrt{2\log(1.25/\delta)}$
2: $\widetilde{\mathrm{err}}_S(L_i) \leftarrow \mathrm{err}_S(L_i) + Z_i$ for every $i \in [f(n)]$, with independent $Z_i \sim \mathcal{N}(0, \sigma^2)$
3: $\widetilde{M}_S(1) \leftarrow 1$ and $\widetilde{M}_S(i) \leftarrow \min_{j \in [i-1]}\big(\widetilde{\mathrm{err}}_S(L_j) - \widetilde{\mathrm{err}}_S(L_i)\big)$ for $2 \leq i \leq f(n)$
4: **return** $\hat{i} \leftarrow \max\big\{i \in [f(n)] : \widetilde{M}_S(i) > 2/f(n)\big\}$ (or $\hat{i} \leftarrow 1$ if no index qualifies)

The empirical error vector has ℓ_2-sensitivity $\sqrt{f(n)}/n$. Correctness on the event $\mathcal{E} \cap \mathcal{F}$, where $\mathcal{E}$ is the concentration event of Section 3.1 and $\mathcal{F}$ is the event that all Gaussian perturbations are at most $1/(4f(n))$, is deterministic: the noisy errors deviate from population values by at most $1/(2f(n))$, ensuring $i^\star$ passes the noisy margin test and all $i > i^\star$ fail.

Theorem 3.3 (Approximate DP Identification)**.** *Under the same conditions as Theorem 3.2, let $\varepsilon, \delta \in (0,1)$. Then Algorithm 2 is (ε, δ)-differentially private and, for all $n \geq n_0$ with n_0 as in Theorem 3.2,*

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,\delta,f}, \mathcal{D}, \mathcal{C}, n) \leq 2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big) + 2f(n)\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big).$$

Choosing $f(n) = \lceil \sqrt{c \log n} \rceil$ gives $\mathrm{IdErr} \leq \exp(-\Omega(n/\log n))$ for any constant $\varepsilon \in (0,1)$ and $\delta \geq n^{-O(1)}$, and for every $r(n) = o(n)$ a suitable horizon gives $\mathrm{IdErr} \leq \exp(-r(n))$ for all large n (Corollary C.16), matching the non-private rate of Høgsgaard and Pabbaraju [2026]. Hence, under approximate DP with constant privacy parameters, identification incurs no asymptotic cost relative to its non-private counterpart. Moreover, the privacy term is at most the statistical term whenever $\varepsilon^2 n \geq 8f(n)\log(1.25/\delta)$, so the approximate DP rate is unaffected even for slowly vanishing ε, whereas the pure DP exponent scales linearly with ε. The proof of Theorem 3.3 is in Appendix C.

4 Differentially Private Language Generation

We design differentially private generation algorithms that, given $S \sim \mathcal{D}^n$, output a novel string from $\mathrm{supp}(\mathcal{D}) \setminus S$. We fix an enumeration $\mathcal{U} = \{u_i\}_{i\in\mathbb{N}}$ and assume every language $L \in \mathcal{C}$ is infinite.

Definition 4.1 (Witness Index)**.** The *witness index* is defined as

$$i(\mathcal{C}, \mathcal{D}) := \inf\{i \in \mathbb{N} : \forall\, L \in \mathcal{C} \text{ with } L \not\subseteq \mathrm{supp}(\mathcal{D}),\ \exists\, k \le i \text{ s.t. } u_k \in L \setminus \mathrm{supp}(\mathcal{D})\},$$

with the convention $\inf \varnothing = \infty$.

In words, $i(\mathcal{C}, \mathcal{D})$ is the smallest prefix length such that every language not fully contained in $\mathrm{supp}(\mathcal{D})$ has a *witness* (a string in the language but outside the support) within the first $i(\mathcal{C}, \mathcal{D})$ elements of $\mathcal{U}$.

We work under the following assumptions [Høgsgaard and Pabbaraju, 2026].

Assumption 4.2 (Finite witness index)**.** *The witness index is finite, i.e.,* $i(\mathcal{C}, \mathcal{D}) < \infty$.

Assumption 4.3 (Support contains a reference language)**.** *There exists* $i^\star \in \mathbb{N}$ *such that* $L_{i^\star} \subseteq \mathrm{supp}(\mathcal{D})$. *Let* $i^\star$ *denote the smallest such index.*

These are the assumptions under which Høgsgaard and Pabbaraju [2026] prove the non-private rate. Assumption 4.2 lets every language not contained in $\mathrm{supp}(\mathcal{D})$ be ruled out by a finite prefix of the enumeration, and Assumption 4.3 provides a language in $\mathcal{C}$ from which every output is valid. Under exact-support realizability ($\mathrm{supp}(\mathcal{D}) = L^\star \in \mathcal{C}$), Assumption 4.3 holds automatically, but Assumption 4.2 need not.

Score function design. The non-private algorithm in [Høgsgaard and Pabbaraju, 2026] maintains, for each language L_i, a *pointer*: the index of the smallest string of L_i not yet seen in the sample. For good languages ($L_i \subseteq \mathrm{supp}(\mathcal{D})$), the pointer eventually advances past the witness window; for bad languages ($L_i \not\subseteq \mathrm{supp}(\mathcal{D})$), it gets stuck at their witness. However, this pointer has *unbounded sensitivity*.

We replace the pointer with a *thresholded prefix count*. For each language L_i and a witness window $[W]$, let $I_i(W) := \{k \in [W] : u_k \in L_i\}$ and define the minimum prefix count

$$a_i^W(S) := \begin{cases} \min_{k\in I_i(W)} N_S(u_k), & \text{if } I_i(W) \neq \varnothing, \\ n, & \text{if } I_i(W) = \varnothing, \end{cases}$$

where $N_S(u_k) := \sum_{t=1}^n \mathbf{1}\{x_t = u_k\}$. Rather than asking whether every relevant prefix string has appeared at least once, we ask whether these strings have appeared at least $g(n)$ times for a threshold $g(n) \to \infty$. The *deficit* is $d_i^W(S) := (g(n) - a_i^W(S))_+$ and the *score* is $q^W(S, i) := -d_i^W(S)$. The score is 0 when the prefix is well-covered (good language) and $-g(n)$ when a witness string is never observed (bad language).

4.1 Pure DP Generation with a Public Witness Bound

We first consider the setting where a public upper bound $W \ge i(\mathcal{C}, \mathcal{D})$ is available (given to the algorithm before observing S and therefore not protected by DP).

Algorithm 3 Pure DP Generation with Public Witness Bound $\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}$

Require: A dataset $S = (x_1, \ldots, x_n) \in \mathcal{U}^n$, a privacy parameter $\varepsilon > 0$, a public witness bound W, functions $f, g\colon \mathbb{N} \to \mathbb{N}$, a language collection $\mathcal{C} = \{L_1, L_2, \ldots\}$

Ensure: A string $\widehat{x} \in \mathcal{U}$

1: Compute $N_S(u_k)$ for $k \in [W]$, and $I_i(W)$, $a_i^W(S)$, $d_i^W(S)$, $q^W(S, i)$ for $i \in [f(n)]$

2: Sample $\widehat{i} \sim \mathrm{EM}_\varepsilon(S, q^W, [f(n)])$ with $\Delta q = 1$ ▷ Privately select a language

3: Sample $\widehat{j} \sim \mathrm{Unif}([2^n])$ ▷ Public randomness

4: **return** the $\widehat{j}$-th smallest-indexed string in $L_{\widehat{i}} \cap \{u_{W+1}, u_{W+2}, \ldots\}$

Having selected a language $L_{\widehat{i}}$, the algorithm outputs a uniformly random string among the first 2^n elements of $L_{\widehat{i}}$ beyond the witness window; when $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$, all of them are valid.

Theorem 4.4 (Pure DP Generation with Public Bound)**.** *Let $\mathcal{C}$ be a countable collection of infinite languages, $\mathcal{D}$ any distribution over $\mathcal{U}$. Suppose that Assumptions 4.2 and 4.3 hold and that a public integer $W \geq i(\mathcal{C},\mathcal{D})$ is given. Define $I_W^\star := \{k \in [W] : u_k \in L_{i^\star}\}$ and $p_W^\star := \min_{k\in I_W^\star} \Pr_{x\sim\mathcal{D}}[x = u_k]$ (with $p_W^\star := 1$ if $I_W^\star = \varnothing$). Let $\varepsilon > 0$ and let $f, g\colon \mathbb{N} \to \mathbb{N}$ satisfy $f(n) \to \infty$, $g(n) \to \infty$, $f(n) \geq i^\star$, and $g(n) \leq np_W^\star/2$ for all large n. Then Algorithm 3 is ε-differentially private and, for all sufficiently large n,*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) \leq \underbrace{|I_W^\star| \exp\Big(-\frac{np_W^\star}{8}\Big)}_{\textit{coverage}} + \underbrace{f(n)\exp\Big(-\frac{\varepsilon\, g(n)}{2}\Big)}_{\textit{private selection}} + \underbrace{\exp\Big(-\frac{n}{2}\Big)}_{\textit{collision}}.$$

Proof idea. Let $\mathcal{E}_W$ be the event that every u_k with $k \in I_W^\star$ appears at least $g(n)$ times in S; a Chernoff bound gives $\Pr[\mathcal{E}_W^c] \leq |I_W^\star| e^{-np_W^\star/8}$. Every bad language contains a witness u_k with $k \leq W$ that never appears, so it scores $-g(n)$ on every sample from $\mathcal{D}$, while on $\mathcal{E}_W$ the language $L_{i^\star}$ scores 0. With sensitivity 1 and margin $g(n)$, Lemma 2.3 selects a bad language with probability at most $f(n)e^{-\varepsilon g(n)/2}$, and the output is novel except with probability $n2^{-n} \leq e^{-n/2}$ for $n \geq 14$ (Lemma A.6). Corollary 4.6 below turns this bound into explicit rates, with and without a public mass floor. The full proof is in Appendix D.

4.2 Pure DP Generation without a Public Witness Bound

When no public bound on $i(\mathcal{C},\mathcal{D})$ is available, we jointly search over both the language index i and a candidate witness threshold t. For each pair $(i,t) \in [f(n)] \times [h(n)]$, define the prefix count $a_{i,t}(S)$ and deficit $d_{i,t}(S)$ analogously to the public-bound case with t replacing W, and the *pair score*

$$q_{\mathrm{pair}}(S,(i,t)) := t - \frac{h(n)}{g(n)}\, d_{i,t}(S). \tag{1}$$

The term t rewards larger thresholds (indicating progress past the witness window), while the penalty $\frac{h(n)}{g(n)} d_{i,t}(S)$ suppresses languages that fail the witness test at threshold t. The coefficient $h(n)/g(n)$ is chosen so that a bad language with full deficit $g(n)$ incurs a penalty of exactly $h(n)$, ensuring its score is at most $t - h(n) \leq 0$.

The algorithm applies the exponential mechanism over the product range $[f(n)] \times [h(n)]$ and outputs a string from the selected language beyond the selected threshold (see Algorithm 6 in the Appendix).

Sensitivity and score gap. The pair score has sensitivity $h(n)/g(n)$. On the coverage event, the good pair $(i^\star, h(n))$ scores $h(n)$, while every pair with a bad language scores at most $i(\mathcal{C},\mathcal{D}) - 1 \leq h(n)/2 - 1$ when $h(n) \geq 2\, i(\mathcal{C},\mathcal{D})$; so the margin exceeds $h(n)/2$, and the ratio of margin to sensitivity is at least $g(n)/2$, as in the public-bound case up to a factor 2. The cost of not knowing the witness bound lies elsewhere: in the larger range $[f(n)] \times [h(n)]$, and in a coverage term that depends on $p_h^\star$, which tends to 0 as $h(n) \to \infty$.

Theorem 4.5 (Pure DP Generation without Public Bound)**.** *Let $\mathcal{C}$ be a countable collection of infinite languages, $\mathcal{D}$ any distribution over $\mathcal{U}$, and suppose Assumptions 4.2 and 4.3 hold. Let $\varepsilon > 0$ and let $f, g, h\colon \mathbb{N} \to \mathbb{N}$ satisfy $f(n) \geq i^\star$, $h(n) \geq 2\, i(\mathcal{C},\mathcal{D})$, $g(n) \to \infty$, and $g(n) \leq np_h^\star/2$ for all large n, where $p_h^\star := \min_{k\in I_h^\star} \Pr_{x\sim\mathcal{D}}[x = u_k]$ (or 1 if $I_h^\star = \varnothing$) and $I_h^\star := \{k \in [h(n)] : u_k \in L_{i^\star}\}$. Then there exists an algorithm (Algorithm 6) that is ε-differentially private and, for all sufficiently large n,*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,h}, \mathcal{D}, \mathcal{C}, n) \leq |I_h^\star| \exp\Big(-\frac{np_h^\star}{8}\Big) + f(n)\, h(n) \exp\Big(-\frac{\varepsilon\, g(n)}{4}\Big) + \exp\Big(-\frac{n}{2}\Big).$$

The proof is in Appendix D.

Rates by available side information. With a fixed public window, $p_W^\star > 0$ does not depend on n; without public information, $h(n) \to \infty$ forces $p_h^\star \to 0$, so no constant mass floor can persist and coverage becomes a condition.

Corollary 4.6 (Fixed public window)**.** *Suppose a window $W \geq i(\mathcal{C},\mathcal{D})$ is public, as in Theorem 4.4, $\varepsilon > 0$ is fixed, and $f(n) \to \infty$ with $f(n) \geq i^\star$. (a) If a constant $p_0 \in (0, p_W^\star]$ is also public, then $g(n) = \max\{1, \lfloor np_0/2 \rfloor\}$ and $\log f(n) = o(n)$ give $\mathrm{GenErr} \leq \exp(-\Omega(\min\{1,\varepsilon\}\cdot n))$, matching Theorem 5.5 up to constants in the exponent. (b) Without a mass floor, for any $r(n) \to \infty$ with $r(n) = o(n)$, $g(n) = \lfloor r(n) \rfloor$ and $\log f(n) = o(r(n))$ give $\mathrm{GenErr} \leq \exp(-\Omega(\min\{1,\varepsilon\} \cdot r(n)))$.*

Corollary 4.7 (No public information)**.** *Under the conditions of Theorem 4.5 with $\varepsilon > 0$ fixed, let $r(n) \to \infty$ with $r(n) = o(n)$, $g(n) = \lfloor r(n) \rfloor$, and $\log(f(n)\,h(n)) = o(r(n))$. If $r(n) \le np_h^\star/2$ for all large n, then* $\mathrm{GenErr} \le \exp(-\Omega(\min\{1,\varepsilon\} \cdot r(n)))$. *This guarantee is conditional on the coverage requirement; it is not an unconditional adaptive rate.*

4.3 Approximate DP Generation

For approximate DP, we replace the exponential mechanism with the Gaussian mechanism: independent Gaussian noise is added to each score before taking the argmax (Algorithms 7 and 8 in Appendix D). With a public window, q^W has sensitivity 1 per coordinate, so its ℓ_2-sensitivity over $[f(n)]$ is at most $\sqrt{f(n)}$; the pair score has sensitivity $h(n)/g(n)$ per coordinate, so its ℓ_2-sensitivity is at most $\frac{h(n)}{g(n)}\sqrt{f(n)h(n)}$.

Theorem 4.8 (Approximate DP Generation)**.** *Let $\varepsilon, \delta \in (0,1)$. (a)* Fixed public window. *Under the conditions of Theorem 4.4, the Gaussian version of Algorithm 3 (Algorithm 7) is (ε,δ)-differentially private and, for all sufficiently large n,*

$$\mathrm{GenErr} \le |I_W^\star| \exp\Big(-\frac{np_W^\star}{8}\Big) + 2f(n)\exp\Big(-\frac{\varepsilon^2 g(n)^2}{16 f(n)\log(1.25/\delta)}\Big) + \exp\Big(-\frac{n}{2}\Big).$$

In particular, for constant ε, the choice $g(n) = \lfloor n/\log n \rfloor$, which needs no knowledge of $p_W^\star$, gives $\mathrm{GenErr} \le \exp(-\Omega(n))$ *whenever $f(n)\log(1.25/\delta) = o(n/\log^2 n)$, matching the non-private rate. (b)* No public information. *Under the conditions of Theorem 4.5, Algorithm 8 is (ε,δ)-differentially private and satisfies the bound of Theorem 4.5 with the privacy term replaced by $2f(n)h(n)\exp\big(-\varepsilon^2 g(n)^2/(64 f(n)h(n)\log(1.25/\delta))\big)$.*

Thus, with a fixed public window, privacy is essentially free under approximate DP for generation; part (b) keeps the coverage term $|I_h^\star|\exp(-np_h^\star/8)$, which is not $\exp(-\Omega(n))$ in general. Unlike identification (Section 3.2), the public-window rate is fully exponential, because the noise $\sigma = O(\sqrt{f(n)\log(1.25/\delta)}/\varepsilon)$ is negligible relative to the gap $g(n)$. The proof is in Appendix D.

5 Lower Bounds for DP Identification and Generation

We show that, for error exponents of order n, the $\min\{1,\varepsilon\}$ scaling is necessary for both tasks. Both constructions couple two distributions that share an atom of mass $3/4$, so that coupled samples differ in $\mathrm{Bin}(n,1/4)$ coordinates, and then apply group privacy; identification is symmetric, while generation uses a one-sided event. We first state the main tool underlying both proofs, a refinement of Lemma 19 of Acharya et al. [2021]; write $\mathrm{d_{Ham}}(S,S')$ for the number of coordinates in which S and S' differ, and $\theta_\varepsilon := 1 - e^{-\varepsilon}$.

Lemma 5.1 (Private Two-Point Inequality)**.** *Let $\varepsilon \in (0,\infty]$, and let $\mathcal{A} : \mathcal{U}^n \to \mathcal{R}$ be ε-differentially private using randomness r (for $\varepsilon = \infty$ no privacy constraint is imposed). Let (S,S') be a coupling of two distributions over $\mathcal{U}^n$. Then for every measurable $\mathcal{F} \subseteq \mathcal{R}$,*

$$\Pr_{S,r}[\mathcal{A}(S) \notin \mathcal{F}] + \Pr_{S',r}[\mathcal{A}(S') \in \mathcal{F}] \;\ge\; \mathbb{E}\big[e^{-\varepsilon\,\mathrm{d_{Ham}}(S,S')}\big],$$

with $e^{-\infty\cdot 0} = 1$. If the coordinate pairs are i.i.d. and differ with probability p, the right side equals $(1 - p\,\theta_\varepsilon)^n$.

Proof. Fix (S,S') and let φ, φ' be the probabilities over r that $\mathcal{A}(S) \in \mathcal{F}$ and $\mathcal{A}(S') \in \mathcal{F}$. Group privacy (Lemma A.4; trivial for $\varepsilon = \infty$) gives $\varphi' \ge e^{-\varepsilon\,\mathrm{d_{Ham}}(S,S')}\varphi$, hence $(1-\varphi) + \varphi' \ge e^{-\varepsilon\,\mathrm{d_{Ham}}(S,S')}$. Take expectations over the coupling; in the i.i.d. case, $\mathrm{d_{Ham}}(S,S')$ is binomial with parameters n and p. □

Each coordinate in which the two samples differ "survives" the privacy constraint with probability θ_ε, and $(1-p\theta_\varepsilon)^n$ is the probability that none survives. Since $(1-e^{-1})\min\{1,\varepsilon\} \le \theta_\varepsilon \le \min\{1,\varepsilon\}$ (Lemma A.6), this is the source of the factor $\min\{1,\varepsilon\}$: the regimes $\varepsilon < 1$ and $\varepsilon \ge 1$ need no separate argument, and $\varepsilon = \infty$ recovers Le Cam's two-point method [Le Cam, 1973, Yu, 1997], so the theorems below use no non-private lower bound of Høgsgaard and Pabbaraju [2026] (Appendix E checks both instances).

Definition 5.2 (IPP Condition)**.** A collection $\mathcal{C}$ satisfies the *Intersecting Private-Pair* (IPP) condition if there exist two distinct languages $L, L' \in \mathcal{C}$ with $L \cap L' \neq \varnothing$, such that both L and L' contain an element not belonging to any other language in $\mathcal{C}$. We call such an element *private* to that language.

Theorem 5.3 (Lower Bound for DP Identification)**.** *Let $\mathcal{C}$ satisfy the IPP condition. There are two distributions $\mathcal{D}, \mathcal{D}'$ satisfying Assumption 3.1 such that for every $\varepsilon \in (0, \infty]$, every ε-DP identification algorithm $\mathcal{A}$, and every $n \in \mathbb{N}$,*

$$\max\{\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n),\ \mathrm{IdErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)\} \ \geq\ \frac{1}{8}\Big(1 - \frac{\theta_\varepsilon}{4}\Big)^n \ \geq\ \frac{1}{8}\, e^{-\min\{1,\varepsilon\}\, n/3}.$$

In particular, one of the two distributions has $\mathrm{IdErr} \geq \frac{1}{8} e^{-\min\{1,\varepsilon\}\, n/3}$ for infinitely many n.

Proof. Let $s_0 \in L \cap L'$, let s_1 be private to L, and let s_2 be private to L'. Let $\mathcal{D}$ (resp. $\mathcal{D}'$) place mass $3/4$ on s_0 and $1/4$ on s_1 (resp. s_2). Under $\mathcal{D}$, the language L has risk 0 and every other language misses s_1, so $\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) \geq \frac{1}{4} \Pr[\mathcal{A}(S) \neq L]$; in the same way, $\mathrm{IdErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n) \geq \frac{1}{4} \Pr[\mathcal{A}(S') = L]$ for $S' \sim (\mathcal{D}')^n$. Couple the coordinates independently, with $(x_t, x_t') = (s_0, s_0)$ with probability $3/4$ and (s_1, s_2) otherwise. Lemma 5.1 with $\mathcal{F} = \{L\}$ and $p = 1/4$ shows that the two misidentification probabilities sum to at least $(1 - \theta_\varepsilon/4)^n$, which gives the first inequality. The second uses $\theta_\varepsilon \leq \min\{1, \varepsilon\}$ and $1 - x \geq e^{-4x/3}$ for $x \in [0, 1/4]$ (Lemma A.6), and the last claim follows by pigeonhole. □

Definition 5.4 (IIDP Condition)**.** A collection $\mathcal{C}$ satisfies the *Intersecting Infinite-Difference Pair* (IIDP) condition if there exist two distinct languages $L, L' \in \mathcal{C}$, an element $s_0 \in L \cap L'$, and two infinite sequences of distinct elements $(a_k)_{k \geq 1} \subseteq L \setminus L'$ and $(b_k)_{k \geq 1} \subseteq L' \setminus L$.

Theorem 5.5 (Lower Bound for DP Generation)**.** *Let $\mathcal{C}$ satisfy the IIDP condition. There are two distributions $\mathcal{D}, \mathcal{D}'$ such that for every $\varepsilon \in (0, \infty]$, every ε-DP generation algorithm $\mathcal{A}$, and every $n \in \mathbb{N}$,*

$$\max\{\mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n),\ \mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)\} \ \geq\ \frac{1}{3}\Big(1 - \frac{\theta_\varepsilon}{4}\Big)^n \ \geq\ \frac{1}{3}\, e^{-\min\{1,\varepsilon\}\, n/3}.$$

In particular, one of the two distributions has $\mathrm{GenErr} \geq \frac{1}{3} e^{-\min\{1,\varepsilon\}\, n/3}$ for infinitely many n.

Proof. Let $\mathcal{D}$ place mass $3/4$ on s_0 and $2^{-k}/4$ on a_k, let $\mathcal{D}'$ do the same with b_k, and let $\mathcal{F} := \{a_k : k \geq 1\}$. Under $\mathcal{D}$, the string s_0 is in S except with probability 4^{-n}, and then every valid novel output lies in $\mathcal{F}$; so $\Pr[\mathcal{A}(S) \notin \mathcal{F}] \leq \mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) + 4^{-n}$. No string of $\mathcal{F}$ lies in $\mathrm{supp}(\mathcal{D}') \subseteq L'$, so $\Pr[\mathcal{A}(S') \in \mathcal{F}] \leq \mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)$. Couple $(x_t, x_t') = (s_0, s_0)$ with probability $3/4$, and (a_K, b_K) with $\Pr[K = k] = 2^{-k}$ otherwise. Lemma 5.1 with $p = 1/4$ shows that the two errors sum to at least $(1 - \theta_\varepsilon/4)^n - 4^{-n} \geq \frac{2}{3}(1 - \theta_\varepsilon/4)^n$, because $(1 - \theta_\varepsilon/4)^n \geq (3/4)^n \geq 3 \cdot 4^{-n}$. The rest is as in the proof of Theorem 5.3. □

Remark 5.6 (Where the Bounds Meet)**.** If $\mathcal{C} = \{L, L'\}$ satisfies IIDP with $L \cap L' = \{s_0\}$, and (a_k), (b_k) enumerate $L \setminus L'$, $L' \setminus L$ in the order of $\mathcal{U}$, then $\mathrm{supp}(\mathcal{D}) = L$ and $\mathrm{supp}(\mathcal{D}') = L'$. Both hard distributions of Theorem 5.5 then satisfy Assumptions 4.2 and 4.3 with a common public window W and mass floor p_0 (Remark E.17), so Corollary 4.6(a) applies, and over the distributions on $\mathcal{C}$ that admit these public values (in particular over $\{\mathcal{D}, \mathcal{D}'\}$) the optimal pure-DP generation rate is $\exp(-\Theta(\min\{1, \varepsilon\} \cdot n))$. For example, $L = \{1\} \cup 3\mathbb{N}$ and $L' = \{1\} \cup (3\mathbb{N} + 1)$ in $\mathcal{U} = \mathbb{N}$ give $W = 4$ and $p_0 = 1/8$, and this $\mathcal{C}$ also satisfies IPP.

6 Conclusion

We initiated the study of differentially private language identification and generation in the agnostic statistical setting. Approximate DP recovers the non-private rates of both tasks (for generation, given a public witness window), and pure DP costs our algorithms a factor $\min\{1, \varepsilon\}$ in the exponent (for generation, given also a public mass floor), which a private two-point inequality shows cannot be avoided at the linear scale. Without public side information, our generation guarantees are conditional on a coverage requirement.

Acknowledgments and Disclosure of Funding

This work was supported by the Australian Research Council (ARC) under the DECRA Fellowship DE240101049. We thank the anonymous reviewers and the area chair for their helpful comments. We used an AI assistant (Claude, Anthropic) to check the proofs and to draft parts of this revision: the scope corrections of Section 4, Lemmas 5.1 and A.5 with the results derived from them, and parts (c) of Corollaries C.9 and C.16; the authors verified all statements and proofs. The authors declare no competing interests.

References

Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 308–318, 2016.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private Assouad, Fano, and Le Cam. In *Algorithmic Learning Theory (ALT)*, volume 132, pages 48–78. PMLR, 2021.

Ishaq Aden-Ali, Hassan Ashtiani, and Gautam Kamath. On the sample complexity of privately learning unbounded high-dimensional Gaussians. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory (ALT)*, volume 132, pages 185–216. PMLR, 2021.

Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private PAC learning implies finite Littlestone dimension. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC)*, pages 852–860, 2019.

Antonios Anastasopoulos, Giuseppe Ateniese, and Evgenios M. Kornaropoulos. Safe language generation in the limit. *arXiv preprint arXiv:2601.08648*, 2026.

Dana Angluin. Finding patterns common to a set of strings. In *Proceedings of the eleventh annual ACM Symposium on Theory of Computing*, pages 130–141, 1979.

Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

Rohan Anil, Badih Ghazi, Vineet Gupta, Ravi Kumar, and Pasin Manurangsi. Large-scale differentially private BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6481–6491. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.findings-emnlp.484.

Marcelo Arenas, Pablo Barceló, Luis Cofré, and Alexander Kozachinskiy. Language generation in the limit: Complexity barriers and implications for learning. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, volume 306 of *Proceedings of Machine Learning Research*, pages 3423–3433. PMLR, 2026.

Yannan Bai, Debmalya Panigrahi, and Ian Zhang. Language generation in the limit: Noise, loss, and feedback. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 794–816. SIAM, 2026. doi: 10.1137/1.9781611978971.31.

Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In *Approximation, Randomization, and Combinatorial Optimization (APPROX/RANDOM)*, volume 8096 of *Lecture Notes in Computer Science*, pages 363–378. Springer, 2013.

Alex Bie, Gautam Kamath, and Vikrant Singhal. Private estimation with public data. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 18653–18666, 2022.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. ISBN 9780199535255.

Mark Bun. A computational separation between private learning and online learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 20732–20743, 2020.

Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1–10. ACM, 2014.

Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially private release and learning of threshold functions. In *Proceedings of the 56th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 634–649, 2015.

Mark Bun, Gautam Kamath, Thomas Steinke, and Zhiwei Steven Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, pages 156–167, 2019.

Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402, 2020.

Moses Charikar and Chirag Pabbaraju. Exploring facets of language generation in the limit. In *Proceedings of the Thirty-Eighth Conference on Learning Theory (COLT)*, volume 291 of *Proceedings of Machine Learning Research*, pages 854–887. PMLR, 2025.

Moses Charikar and Chirag Pabbaraju. Pareto-optimal non-uniform language generation. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, volume 313 of *Proceedings of Machine Learning Research*, pages 1–27. PMLR, 2026.

Moses Charikar, Chirag Pabbaraju, and Ambuj Tewari. A characterization of list language identification in the limit. In *Proceedings of the Thirty-Ninth Conference on Learning Theory (COLT)*, volume 336 of *Proceedings of Machine Learning Research*, pages 1022–1053. PMLR, 2026.

Kamalika Chaudhuri and Daniel Hsu. Convergence rates for differentially private statistical estimation. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 2012.

Edith Cohen, Xin Lyu, Jelani Nelson, Tamás Sarlós, and Uri Stemmer. Optimal differentially private learning of thresholds and quasi-concave optimization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 472–482. ACM, 2023. doi: 10.1145/3564246.3585148.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–487, 2014.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology – EUROCRYPT 2006*, volume 4004 of *Lecture Notes in Computer Science*, pages 486–503. Springer, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference (TCC)*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006b. doi: 10.1007/11681878_14.

Vitaly Feldman and David Xiao. Sample complexity bounds on differentially private learning via communication complexity. *SIAM Journal on Computing*, 44(6):1740–1764, 2015.

Fengyu Gao, Ruida Zhou, Tianhao Wang, Cong Shen, and Jing Yang. Data-adaptive differentially private prompt synthesis for in-context learning. In *The Thirteenth International Conference on Learning Representations (ICLR)*, pages 60152–60180, 2025.

Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Raghu Meka, and Chiyuan Zhang. User-level differential privacy with few examples per user. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 19263–19290, 2023.

E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

Sivakanth Gopi, Gautam Kamath, Janardhan Kulkarni, Aleksandar Nikolov, Zhiwei Steven Wu, and Huanyu Zhang. Locally private hypothesis selection. In *Proceedings of the Thirty-Third Conference on Learning Theory (COLT)*, volume 125, pages 1785–1816. PMLR, 2020.

Steve Hanneke, Amin Karbasi, Anay Mehrotra, and Grigoris Velegkas. On union-closedness of language generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, pages 99502–99525, 2025.

Mikael Møller Høgsgaard and Chirag Pabbaraju. Agnostic language identification and generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2026. To appear; arXiv:2601.23258.

Alkis Kalavasis, Anay Mehrotra, and Grigoris Velegkas. On the limits of language generation: Trade-offs between hallucination and mode-collapse. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1732–1743. ACM, 2025. doi: 10.1145/3717823.3718108.

Alkis Kalavasis, Anay Mehrotra, and Grigoris Velegkas. On characterizations for language generation: Interplay of hallucinations, breadth, and stability. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, volume 313 of *Proceedings of Machine Learning Research*, pages 1–49. PMLR, 2026.

Amin Karbasi, Omar Montasser, John Sous, and Grigoris Velegkas. (Im)possibility of automated hallucination detection in large language models. In *Conference on Language Modeling (COLM)*, 2025.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Jon Kleinberg and Sendhil Mullainathan. Language generation in the limit. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 66058–66079, 2024.

Jon Kleinberg and Fan Wei. Density measures for language generation. In *Proceedings of the 66th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 620–658. IEEE, 2025. doi: 10.1109/FOCS63196.2025.00034.

Jon Kleinberg and Fan Wei. Language generation and identification from partial enumeration: Tight density bounds and topological characterizations. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1684–1691. ACM, 2026a. doi: 10.1145/3798129.3800876.

Jon Kleinberg and Fan Wei. Validity, sparse holes, and breadth in language generation: Banach density, topology, and geometry. *arXiv preprint arXiv:2604.02385*, 2026b.

Steffen Lange, Thomas Zeugmann, and Sandra Zilles. Learning indexed families of recursive languages from positive data: A survey. *Theoretical Computer Science*, 397(1-3):194–232, 2008.

Lucien Le Cam. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, 1(1):38–53, 1973. doi: 10.1214/aos/1193342380.

Daniel Levy, Ziteng Sun, Kareem Amin, Satyen Kale, Alex Kulesza, Mehryar Mohri, and Ananda Theertha Suresh. Learning with user-level privacy. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 12466–12479, 2021.

Aaron Li and Ian Zhang. Characterizing the effect of noise in language generation in the limit. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, volume 306 of *Proceedings of Machine Learning Research*, pages 72277–72289. PMLR, 2026.

Jiaxun Li, Vinod Raman, and Ambuj Tewari. Generation through the lens of learning theory. In *Proceedings of the Thirty-Eighth Conference on Learning Theory (COLT)*, volume 291 of *Proceedings of Machine Learning Research*, pages 4740–4776. PMLR, 2025.

Jiaxun Li, Vinod Raman, and Ambuj Tewari. On generation in metric spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2026. To appear; arXiv:2602.07710.

Xuechen Li, Florian Tramèr, Percy Liang, and Tatsunori Hashimoto. Large language models can be strong differentially private learners. In *International Conference on Learning Representations (ICLR)*, 2022.

Ruixuan Liu and Zhiqi Bu. Towards hyperparameter-free optimization with differential privacy. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

Yuhan Liu, Ananda Theertha Suresh, Felix Yu, Sanjiv Kumar, and Michael Riley. Learning discrete distributions: User vs item-level privacy. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 20965–20976, 2020.

Bartłomiej Marek, Lorenzo Rossi, Vincent Hanke, Xun Wang, Michael Backes, Franziska Boenisch, and Adam Dziedzic. Benchmarking empirical privacy protection for adaptations of large language models. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Ryan McKenna, Yangsibo Huang, Amer Sinha, Borja Balle, Zachary Charles, Christopher A. Choquette-Choo, Badih Ghazi, Georgios Kaissis, Ravi Kumar, Ruibo Liu, Da Yu, and Chiyuan Zhang. Scaling laws for differentially private language models. In *Forty-second International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pages 43375–43398. PMLR, 2025.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 94–103. IEEE, 2007.

Anay Mehrotra, Grigoris Velegkas, Xifan Yu, and Felix Zhou. Language generation with infinite contamination. In *Proceedings of the 39th Conference on Learning Theory (COLT)*, volume 336 of *Proceedings of Machine Learning Research*, pages 5053–5112. PMLR, 2026a.

Anay Mehrotra, Grigoris Velegkas, Xifan Yu, and Felix Zhou. Differentially private language generation and identification in the limit (extended abstract). In *Proceedings of the Thirty-Ninth Conference on Learning Theory (COLT)*, volume 336 of *Proceedings of Machine Learning Research*, pages 5113–5114. PMLR, 2026b. Full version: arXiv:2604.08504.

Hristo Papazov and Nicolas Flammarion. Learning algorithms in the limit. In *The Thirty-Eighth Annual Conference on Learning Theory (COLT)*, pages 4486–4510. PMLR, 2025.

Charlotte Peale, Vinod Raman, and Omer Reingold. Representative language generation. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pages 48518–48541. PMLR, 2025.

Binghui Peng, Amin Saberi, and Grigoris Velegkas. Language identification in the limit with computational trace. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Natalia Ponomareva, Hussein Hazimeh, Alex Kurakin, Zheng Xu, Carson Denison, H. Brendan McMahan, Sergei Vassilvitskii, Steve Chien, and Abhradeep Guha Thakurta. How to DP-fy ML: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77:1113–1201, 2023. doi: 10.1613/jair.1.14649.

Natalia Ponomareva, Zheng Xu, H. Brendan McMahan, Peter Kairouz, Lucas Rosenblatt, Vincent Cohen-Addad, Cristóbal Guzmán, Ryan McKenna, Galen Andrew, Alex Bie, Da Yu, Alex Kurakin, Morteza Zadimoghaddam, Sergei Vassilvitskii, and Andreas Terzis. How to DP-fy your data: A practical guide to generating synthetic data with differential privacy. *Journal of Artificial Intelligence Research*, 86, 2026. doi: 10.1613/jair.1.21111.

Giorgio Racca, Michal Valko, and Amartya Sanyal. Language generation with replay: A learning-theoretic view of model collapse. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, volume 306 of *Proceedings of Machine Learning Research*, pages 102821–102843. PMLR, 2026.

Ananth Raman and Vinod Raman. Generation from noisy examples. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pages 51079–51093. PMLR, 2025.

Tom Sander, Pierre Stock, and Alexandre Sablayrolles. TAN without a burn: Scaling laws of DP-SGD. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 29937–29949. PMLR, 2023.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 813–822, 2011.

Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7(2), 2017. doi: 10.29012/jpc.v7i2.648.

Rushil Thareja, Preslav Nakov, Praneeth Vepakomma, and Nils Lukas. DP-fusion: Token-level differentially private inference for large language models. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018.

Martin J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2019.

Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997. doi: 10.1007/978-1-4612-1880-7_29.

Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A. Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, Sergey Yekhanin, and Huishuai Zhang. Differentially private fine-tuning of language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Xiang Yue, Huseyin A. Inan, Xuechen Li, Girish Kumar, Julia McAnallen, Hoda Shajari, Huan Sun, David Levitan, and Robert Sim. Synthetic text generation with differential privacy: A simple and practical recipe. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1321–1342. Association for Computational Linguistics, 2023.

Appendix

Organization of the Appendix. Section A lists the notation (Table 2) and collects concentration inequalities and privacy tools used throughout, including the private selection lemma behind every upper bound (Lemma A.5). Section B reviews the non-private algorithms. Section C contains the full proofs for the DP identification upper bounds. Section D contains the full proofs for the DP generation upper bounds. Section E checks the hard instances of the lower bounds proved in Section 5. Section F discusses the limitations and future directions, and Section G the broader impacts. Section H provides additional related work.

A Notation and Useful Lemmas

Table 2: Notation. The symbol $i^\star$ has the meaning of Assumption 3.1 in Section 3 and of Assumption 4.3 in Section 4.

Symbol	Meaning	Where
$\mathcal{U} = \{u_1, u_2, \ldots\}$, $\mathcal{C} = \{L_1, L_2, \ldots\}$	universe with a fixed enumeration; language collection	Section 2
$\mathcal{D}$, $S = (x_1, \ldots, x_n) \sim \mathcal{D}^n$	unknown distribution; sample	Section 2
$\mathrm{err}_{\mathcal{D}}(L)$, $\mathrm{err}_S(L)$	population and empirical risk of L	Section 2
IdErr, GenErr	identification and generation error	Section 2
$S \sim S'$, $\mathrm{d}_{\mathrm{Ham}}(S, S')$	neighboring samples; Hamming distance	Sections 2 and 5
Δq, Δ_2	sensitivity bound of a score; ℓ_2-sensitivity	Section 2
$\mathrm{EM}_\varepsilon(S, q, \mathcal{R})$, Γ	exponential mechanism; margin of a set of good outputs	Lemma 2.3
$\theta_\varepsilon = 1 - e^{-\varepsilon}$	survival probability of a differing coordinate	Lemma 5.1
$f(n)$	horizon: candidate languages $L_1, \ldots, L_{f(n)}$	Sections 3 and 4
c_{gap}	risk gap between $L_{i^\star}$ and the earlier languages	(2)
$M_S(i)$, $d_S(i)$, $q(S, i)$	empirical margin, deficit and score for identification	Section 3
$\mathcal{E}$; $\widetilde{M}_S(i)$, $\mathcal{F}$	concentration event; noisy margin and noise event	Sections 3.1 and 3.2
$i(\mathcal{C}, \mathcal{D})$, W	witness index; public witness bound $W \geq i(\mathcal{C}, \mathcal{D})$	Definition 4.1, Section 4.1
$N_S(u)$	number of occurrences of u in S	Section 4
$I_i(W)$, $a_i^W(S)$, $d_i^W(S)$, $q^W(S, i)$	prefix indices, prefix count, deficit and score for generation	Section 4
$g(n)$	count threshold	Section 4
$I_W^\star$, $p_W^\star$; p_0	target prefix and its mass floor; public lower bound on $p_W^\star$	Theorem 4.4, Corollary 4.6
$\mathcal{E}_W$, $\mathcal{E}_h$	coverage events	Section D
$h(n)$, $a_{i,t}(S)$, $d_{i,t}(S)$, q_{pair}	threshold range; pair count, deficit and score	Section 4.2
$I_h^\star$, $p_h^\star$	target prefix and mass floor at level $h(n)$	Theorem 4.5
s_0, s_1, s_2; (a_k), (b_k)	shared and private strings of the hard instances	Section 5
$\mathcal{D}, \mathcal{D}'$, $\mathcal{F}$	hard pair; event used in the two-point inequality	Theorems 5.3 and 5.5

A.1 Concentration Inequalities

We state some standard concentration inequalities and tail bounds; see, e.g., Boucheron et al. [2013], Vershynin [2018], Wainwright [2019] for comprehensive references.

Lemma A.1 (Hoeffding's Inequality)**.** *Let $X = \frac{1}{n}\sum_{i=1}^n X_i$ be the average of n independent random variables taking values in $[a, b]$, and let $\mu = E[X]$. For any $t \geq 0$, we have*

$$\Pr[|X - \mu| \geq t] \leq 2\exp\Big(-\frac{2nt^2}{(b-a)^2}\Big).$$

Lemma A.2 (Chernoff Bounds). *Let $X = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the average of n independent random variables taking values in $[0,1]$, and let $\mu = E[X]$. For any $t \geq 0$,*

$$P(X \leq (1-t)\mu) \leq \exp\left(-\frac{n\mu t^2}{2}\right).$$

For any $t \in [0,1]$,

$$P(X \geq (1+t)\mu) \leq \exp\left(-\frac{n\mu t^2}{3}\right).$$

Lemma A.3 (Gaussian Tail Bounds). *If $Z \sim \mathcal{N}(0,\sigma^2)$, then for every $t \geq 0$, we have*

$$\Pr[Z \geq t] \leq \exp\Big(-\frac{t^2}{2\sigma^2}\Big) \quad \text{and} \quad \Pr[|Z| \geq t] \leq 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

A.2 Tools from Differential Privacy

Lemma A.4 (Group Privacy Dwork and Roth [2014]). *Let $\mathcal{A} : \mathcal{U}^n \to \mathcal{R}$ be ε-differentially private. If $S, S' \in \mathcal{U}^n$ differ in at most k entries, then for every measurable $\mathcal{F} \subseteq \mathcal{R}$, we have*

$$\Pr[\mathcal{A}(S) \in \mathcal{F}] \leq \exp(\varepsilon k)\Pr[\mathcal{A}(S') \in \mathcal{F}],$$

where the probability is taken over the randomness of $\mathcal{A}$.

The proof of the private two-point inequality (Lemma 5.1) is complete in Section 5. All our upper bounds use the private mechanisms only through the next lemma. It isolates the two quantities that each proof computes: a margin Γ by which the correct candidates beat all others on a high-probability event, and the sensitivity of the score.

Lemma A.5 (Private Selection with a Margin). *Let $\mathcal{R}$ be a finite set, let $q\colon \mathcal{U}^n \times \mathcal{R} \to \mathbb{R}$ be a score with sensitivity Δ, fix a sample $S \in \mathcal{U}^n$, and let $\mathcal{G} \subseteq \mathcal{R}$ be a set of good candidates.*

(a) If $\Gamma := \max_{c\in\mathcal{G}} q(S,c) - \max_{c\in\mathcal{R}\setminus\mathcal{G}} q(S,c) > 0$, then $\Pr[\mathrm{EM}_\varepsilon(S,q,\mathcal{R}) \notin \mathcal{G}] \leq |\mathcal{R}\setminus\mathcal{G}|\, e^{-\varepsilon\Gamma/(2\Delta)}$.

(b) Let $T\colon \mathcal{U}^n \to \mathbb{R}^d$ have ℓ_2-sensitivity Δ_2, let $\varepsilon, \delta \in (0,1)$, and let $Z \sim \mathcal{N}(0,\sigma^2 I_d)$ with $\sigma = \Delta_2\sqrt{2\log(1.25/\delta)}/\varepsilon$. Then the release of $T(S)+Z$ is (ε,δ)-differentially private. If a rule $\phi\colon \mathbb{R}^d \to \mathcal{R}$ satisfies $\phi(T(S)+z) \in \mathcal{G}$ for every z with $\|z\|_\infty < \eta$, then $\Pr[\phi(T(S)+Z) \notin \mathcal{G}] \leq 2d\, e^{-\eta^2/(2\sigma^2)}$.

(c) In particular, let $T(S) = (q(S,c))_{c\in\mathcal{R}}$, so that $d = |\mathcal{R}|$ and $\Delta_2 \leq \Delta\sqrt{|\mathcal{R}|}$, and let ϕ return a maximizer of $T(S)+Z$. If $\Gamma > 0$ as in part (a), then $\Pr[\phi(T(S)+Z) \notin \mathcal{G}] \leq 2|\mathcal{R}|\, e^{-\Gamma^2/(8\sigma^2)}$.

Proof. (a) Let $c^\star \in \mathcal{G}$ have the largest score among the good candidates, so that every $c \notin \mathcal{G}$ has $q(S,c) \leq q(S,c^\star) - \Gamma$. The exponential mechanism outputs c with probability $\exp(\varepsilon q(S,c)/(2\Delta))$ divided by the sum of these weights over $\mathcal{R}$, and that sum is at least the weight of $c^\star$. Hence every bad candidate c is selected with probability at most $\exp(\varepsilon(q(S,c)-q(S,c^\star))/(2\Delta)) \leq e^{-\varepsilon\Gamma/(2\Delta)}$, and a union bound over the bad candidates gives part (a).

(b) The release is (ε,δ)-differentially private by Lemma 2.4. Each coordinate Z_j has the law $\mathcal{N}(0,\sigma^2)$, so $\Pr[|Z_j| \geq \eta] \leq 2e^{-\eta^2/(2\sigma^2)}$ by Lemma A.3, and a union bound over the d coordinates gives $\Pr[\|Z\|_\infty \geq \eta] \leq 2d\, e^{-\eta^2/(2\sigma^2)}$. On the complementary event, $\phi(T(S)+Z) \in \mathcal{G}$ by assumption.

(c) A neighboring sample changes each of the $|\mathcal{R}|$ coordinates of $T(S)$ by at most Δ, so $\Delta_2 \leq \Delta\sqrt{|\mathcal{R}|}$, and the release with σ computed from the bound $\Delta\sqrt{|\mathcal{R}|}$ is private as well. Let $c^\star$ be as in part (a), and suppose that $\|z\|_\infty < \Gamma/2$. Then every $c \notin \mathcal{G}$ satisfies

$$q(S,c^\star) + z_{c^\star} > q(S,c^\star) - \Gamma/2 \geq q(S,c) + \Gamma/2 > q(S,c) + z_c,$$

so no bad candidate maximizes the noisy scores. Part (b) with $\eta = \Gamma/2$ gives the bound. □

In words, one changed sample moves every score by at most Δ and so closes the margin by at most 2Δ; the exponential mechanism turns the number $\Gamma/(2\Delta)$ of changes that an adversary needs into the failure probability $e^{-\varepsilon\Gamma/(2\Delta)}$, while Gaussian noise turns it into $e^{-\Theta(\varepsilon^2\Gamma^2/(\Delta^2|\mathcal{R}|\log(1/\delta)))}$, which is quadratic in Γ/Δ.

Lemma A.6 (Elementary Inequalities). *(a) For every $\varepsilon \in (0,\infty]$, $(1-e^{-1})\min\{1,\varepsilon\} \le 1-e^{-\varepsilon} \le \min\{1,\varepsilon\}$.*

(b) For every $x \in [0,1/4]$, $1-x \ge e^{-4x/3}$.

(c) For every integer $n \ge 14$, $n2^{-n} \le e^{-n/2}$.

Proof. (a) The upper bound holds because $1-e^{-\varepsilon} \le 1$ and $e^{-\varepsilon} \ge 1-\varepsilon$. For the lower bound, if $\varepsilon \ge 1$, then $1-e^{-\varepsilon} \ge 1-e^{-1}$. If $\varepsilon < 1$, then concavity gives the bound: the map $\varepsilon \mapsto 1-e^{-\varepsilon}$ is concave and vanishes at 0, so on $[0,1]$ it lies above its chord, and $1-e^{-\varepsilon} \ge (1-e^{-1})\varepsilon$.

(b) Let $x \in [0,1/4]$, so that $1-x \ge 3/4$. Expand the logarithm and compare it with a geometric series:

$$\log(1-x) = -\sum_{k\ge1}\frac{x^k}{k} \ge -\sum_{k\ge1}x^k = -\frac{x}{1-x} \ge -\frac{4x}{3}.$$

(c) Taking logarithms, the claim is $\log n \le (\log 2 - \frac{1}{2})\,n$. It holds at $n = 14$, because $\log 14 < 2.64 < 0.193 \cdot 14$. Beyond $n = 14$, the right side grows by more than 0.19 per unit step, and the left side grows by less than $1/14 < 0.08$. □

B Non-Private Algorithms for Language Identification and Generation

For completeness, we restate the non-private algorithms of Høgsgaard and Pabbaraju [2026] that serve as the starting points for our private constructions.

Algorithm 4 performs agnostic identification via a margin-based selection rule: it selects the largest-indexed language within a growing horizon $[f(n)]$ that beats all predecessors in empirical risk by a margin of $2/f(n)$.

Algorithm 4 Non-Private Agnostic Identification [Høgsgaard and Pabbaraju, 2026]

Require: A dataset $S = (x_1,\dots,x_n) \in \mathcal{U}^n$, a function $f\colon \mathbb{N} \to \mathbb{N}$ with $f(n) \to \infty$, a language collection $\mathcal{C} = \{L_1, L_2, \dots\}$

Ensure: A language $L_{A(S)} \in \mathcal{C}$

1: **for** $i \in [f(n)]$ **do**
2: $\quad \mathrm{err}_S(L_i) \leftarrow \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \notin L_i\}$
3: **end for**
4: $A(S) \leftarrow$ largest index $i \in [f(n)]$ such that $\mathrm{err}_S(L_j) - \mathrm{err}_S(L_i) > \frac{2}{f(n)}$ for all $j < i$
5: **return** $L_{A(S)}$

Algorithm 5 performs agnostic generation via a pointer-based rule: for each language, it maintains a pointer to the smallest-indexed string not yet seen in the sample, and selects the language whose pointer has advanced the farthest. Note that the original presentation in Høgsgaard and Pabbaraju [2026] iterates over all $i \in \mathbb{N}$. We restrict to a finite horizon $[f(n)]$ with $f(n) \to \infty$. This is without loss of generality since $f(n) \ge i^\star$ for all sufficiently large n, and it is necessary for our privatization.

The key challenge in privatizing these algorithms is that their score functions have very different sensitivity properties. The identification margin test (Algorithm 4, line 4) is discontinuous: changing a single sample can flip which indices satisfy the margin constraint. The generation pointer (Algorithm 5, lines 5–7) has unbounded sensitivity: removing the sole occurrence of some u_k from the sample can reset the pointer from beyond the witness window back to k, an arbitrary jump. Our private algorithms resolve these issues by replacing both rules with smooth score functions amenable to the exponential and Gaussian mechanisms.

Algorithm 5 Non-Private Agnostic Generation [Høgsgaard and Pabbaraju, 2026]

Require: A dataset $S = (x_1, \ldots, x_n) \in \mathcal{U}^n$, a function $f\colon \mathbb{N} \to \mathbb{N}$ with $f(n) \to \infty$, a language collection $\mathcal{C} = \{L_1, L_2, \ldots\}$
Ensure: A string $\widehat{x} \in \mathcal{U}$

```
for i ∈ [f(n)] do
    r_i ← min{k ∈ ℕ : u_k ∈ L_i}                ▷ Initialize pointer to first string in L_i
end for
for i ∈ [f(n)] do
    while ∃ t ∈ [n] : x_t = u_{r_i} do
        r_i ← min{k ∈ ℕ : u_k ∈ L_i, k > r_i}   ▷ Advance pointer past seen strings
    end while
end for
o ← arg max_{i∈[f(n)]} r_i                      ▷ Select language with largest pointer
return any novel string from L_o
```

C Proofs for Section 3 (Upper Bounds for DP Identification)

Assumption C.1 (Agnostic Optimum is Attainable)**.** *There exists an index $i^\star \in \mathbb{N}$ such that $\mathrm{err}_{\mathcal{D}}(L_{i^\star}) = \inf_{L \in \mathcal{C}} \mathrm{err}_{\mathcal{D}}(L)$. Let $i^\star$ denote the smallest such index.*

When $i^\star > 1$, we define the least risk gap between $i^\star$ and the earlier indices as

$$c_{\mathrm{gap}} := \min_{j \in [i^\star - 1]} \left(\mathrm{err}_{\mathcal{D}}(L_j) - \mathrm{err}_{\mathcal{D}}(L_{i^\star})\right) > 0. \tag{2}$$

Let $f : \mathbb{N} \to \mathbb{N}$ be a function with $f(n) \to \infty$. There exists n_0 (depending on $\mathcal{C}, \mathcal{D}, f$) such that for all $n \geq n_0$, we have

(i) $f(n) \geq i^\star$ (the optimal language is within the horizon).

(ii) $f(n) > 3/c_{\mathrm{gap}}$ if $i^\star > 1$ (the margin threshold $2/f(n)$ is smaller than the gap).

(iii) $f(n) \geq 3$ (the index 1, whose margin is 1, passes the test $M > 2/f(n)$).

The threshold n_0 does not depend on ε or δ. All results below assume $n \geq n_0$ without further mention.

C.1 Proof of Theorem 3.2 (Pure DP Identification)

For $S \in \mathcal{U}^n, i \in [f(n)]$, we define the margin

$$M_S(i) := \begin{cases} 1, & \text{if } i = 1, \\ \min_{j \in [i-1]} \big(\mathrm{err}_S(L_j) - \mathrm{err}_S(L_i)\big), & \text{if } i \geq 2, \end{cases} \tag{3}$$

the deficit

$$d_S(i) := \Big(\frac{2}{f(n)} - M_S(i)\Big)_+, \tag{4}$$

and the score

$$q(S, i) := i - f(n)^2 d_S(i). \tag{5}$$

C.1.1 Privacy Analysis

The privacy proof follows directly from the exponential mechanism once we bound the sensitivity of the score function.

Lemma C.2 (Sensitivity of the Score Function)**.** *The score function q defined in (5) has sensitivity $2f(n)^2/n$. That is, for any neighboring $S, S' \in \mathcal{U}^n$ and any $i \in [f(n)]$, we have*

$$|q(S, i) - q(S', i)| \leq \frac{2f(n)^2}{n}.$$

Proof. Let $S, S' \in \mathcal{U}^n$ be two neighboring datasets. Since each $\mathrm{err}_S(L_j) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \notin L_j\}$ is an average of n binary indicators, changing one element in a dataset changes at most one indicator, so for $j \in [f(n)]$, we have

$$|\mathrm{err}_S(L_j) - \mathrm{err}_{S'}(L_j)| \le \frac{1}{n}. \tag{6}$$

For $i \ge 2$, define $\Delta_{j,i}(S) := \mathrm{err}_S(L_j) - \mathrm{err}_S(L_i)$. By the triangle inequality and (6), we have $|\Delta_{j,i}(S) - \Delta_{j,i}(S')| \le 2/n$. Since the pointwise minimum of functions preserves Lipschitz constants, it holds that

$$|M_S(i) - M_{S'}(i)| \le \frac{2}{n}. \tag{7}$$

For $i = 1$, we have $M_S(1) = M_{S'}(1) = 1$, so (7) holds trivially.

Note that the map $x \mapsto (2/f(n) - x)_+$ is 1-Lipschitz. Composing with (7), so $|d_S(i) - d_{S'}(i)| \le 2/n$. Therefore

$$|q(S,i) - q(S',i)| = f(n)^2 |d_S(i) - d_{S'}(i)| \le f(n)^2 \cdot \frac{2}{n} = \frac{2f(n)^2}{n}.$$

□

Lemma C.3 (Privacy). *Algorithm 1 is ε-differentially private.*

Proof. The only data-dependent output is $\widehat{i}$, produced by the exponential mechanism with score q over range $[f(n)]$. By Lemma C.2, q has sensitivity $\Delta = 2f(n)^2/n$. The exponential mechanism guarantee (Lemma 2.3) yields ε-DP. □

C.1.2 Utility Analysis

The utility analysis proceeds in three steps: (i) a uniform concentration event $\mathcal{E}$ under which all empirical errors in the horizon are close to their population values (Lemma C.4); (ii) on $\mathcal{E}$, the optimal index $i^\star$ is the unique maximizer of the score function with a gap of at least 1 (Lemma C.6); (iii) the exponential mechanism exploits this gap to select $i^\star$ with high probability (Lemma C.7).

Lemma C.4 (Uniform Concentration). *Define the event*

$$\mathcal{E} := \Big\{ |\mathrm{err}_S(L_i) - \mathrm{err}_{\mathcal{D}}(L_i)| \le \frac{1}{4f(n)}, \quad \forall i \in [f(n)] \Big\}.$$

Then $\Pr_{S\sim\mathcal{D}^n}[\mathcal{E}^c] \le 2f(n)\exp\big(-n/(8f(n)^2)\big)$.

Proof. For each $i \in [f(n)]$, Hoeffding's inequality (Lemma A.1) gives

$$\Pr\Big[|\mathrm{err}_S(L_i) - \mathrm{err}_{\mathcal{D}}(L_i)| > \frac{1}{4f(n)}\Big] \le 2\exp\Big(-2 \cdot \frac{n}{16f(n)^2}\Big) = 2\exp\Big(-\frac{n}{8f(n)^2}\Big).$$

A union bound over $[f(n)]$ yields

$$\Pr[\mathcal{E}^c] \le \sum_{i=1}^{f(n)} 2\exp\left(-\frac{n}{8f(n)^2}\right) = 2f(n)\exp\left(-\frac{n}{8f(n)^2}\right).$$

□

Remark C.5 (Choice of Precision $1/(4f(n))$). We choose the precision $1/(4f(n))$ so that on $\mathcal{E}$: for $j < i^\star$, the empirical gap $\mathrm{err}_S(L_j) - \mathrm{err}_S(L_{i^\star})$ remains above $2/f(n)$, ensuring $i^\star$ passes the margin test; for $i > i^\star$, the empirical gap $\mathrm{err}_S(L_{i^\star}) - \mathrm{err}_S(L_i)$ is at most $1/(2f(n)) < 2/f(n)$, ensuring such i fail the margin test with a deficit large enough to push their score below that of $i^\star$. In fact, any precision η with $2\eta < c_{\mathrm{gap}} - 2/f(n)$ and $\eta \le 1/(2f(n))$ would work; $1/(4f(n))$ satisfies both since $f(n) > 3/c_{\mathrm{gap}}$.

Lemma C.6 (Score separation on $\mathcal{E}$). *On the event $\mathcal{E}$, we have*

(a) $q(S, i^\star) = i^\star$.

(b) $q(S,i) \le i^\star - 1$ *for every* $i \in [f(n)] \setminus \{i^\star\}$.

Consequently, the set $\{i^\star\}$ *has margin at least* 1 *over the other indices.*

Proof. Assume $\mathcal{E}$ holds throughout.

Case 1: $q(S,i^\star) = i^\star$**.** We show $d_S(i^\star) = 0$, i.e., $M_S(i^\star) \ge 2/f(n)$. If $i^\star = 1$, then $M_S(1) = 1 \ge 2/f(n)$, so $d_S(1) = 0$ and $q(S,1) = 1 = i^\star$. If $i^\star \ge 2$, then for every $j < i^\star$, we have $\mathrm{err}_{\mathcal{D}}(L_j) - \mathrm{err}_{\mathcal{D}}(L_{i^\star}) \ge c_{\mathrm{gap}} \ge 3/f(n)$, where the last inequality uses $f(n) \ge 3/c_{\mathrm{gap}}$. On $\mathcal{E}$, each empirical risk deviates from the corresponding population risk by at most $1/(4f(n))$, so

$$\mathrm{err}_S(L_j) - \mathrm{err}_S(L_{i^\star}) \ge \big(\mathrm{err}_{\mathcal{D}}(L_j) - \mathrm{err}_{\mathcal{D}}(L_{i^\star})\big) - \frac{2}{4f(n)} \ge \frac{3}{f(n)} - \frac{1}{2f(n)} = \frac{5}{2f(n)} > \frac{2}{f(n)}.$$

Taking the minimum over $j < i^\star$ gives $M_S(i^\star) \ge 5/(2f(n)) > 2/f(n)$, hence $d_S(i^\star) = 0$ and $q(S,i^\star) = i^\star$.

Case 2: $q(S,i) \le i^\star - 1$ **for** $i < i^\star$**.** Since $d_S(i) \ge 0$, we have

$$q(S,i) = i - f(n)^2 d_S(i) \le i \le i^\star - 1$$

Case 3: $q(S,i) \le i^\star - 1$ **for** $i > i^\star$**.** By the optimality of $i^\star$, we have $\mathrm{err}_{\mathcal{D}}(L_i) \ge \mathrm{err}_{\mathcal{D}}(L_{i^\star})$. On $\mathcal{E}$,

$$\begin{aligned}\mathrm{err}_S(L_{i^\star}) - \mathrm{err}_S(L_i) &\le \Big(\mathrm{err}_{\mathcal{D}}(L_{i^\star}) + \frac{1}{4f(n)}\Big) - \Big(\mathrm{err}_{\mathcal{D}}(L_i) - \frac{1}{4f(n)}\Big) \\ &= \big(\mathrm{err}_{\mathcal{D}}(L_{i^\star}) - \mathrm{err}_{\mathcal{D}}(L_i)\big) + \frac{1}{2f(n)} \le \frac{1}{2f(n)}.\end{aligned}$$

Since $j = i^\star$ is among the candidates in the minimum defining $M_S(i)$, we obtain $M_S(i) \le 1/(2f(n)) < 2/f(n)$. Therefore

$$d_S(i) = \frac{2}{f(n)} - M_S(i) \ge \frac{2}{f(n)} - \frac{1}{2f(n)} = \frac{3}{2f(n)},$$

and, since $i \le f(n)$, the score satisfies

$$q(S,i) = i - f(n)^2 \cdot d_S(i) \le i - f(n)^2 \cdot \frac{3}{2f(n)} = i - \frac{3f(n)}{2} \le -\frac{f(n)}{2} \le i^\star - 1.$$

□

Lemma C.7 (Private Selection)**.** *On the event* $\mathcal{E}$*, Algorithm 1 outputs some* $\widehat{i} \ne i^\star$ *with probability at most* $f(n)\exp\big(-\varepsilon n/(4f(n)^2)\big)$.

Proof. Fix a sample S in $\mathcal{E}$. By Lemma C.6, the set $\{i^\star\}$ has margin at least 1 over the other indices of $[f(n)]$, and by Lemma C.2 the mechanism uses the sensitivity bound $\Delta = 2f(n)^2/n$. Lemma A.5(a) bounds the probability of selecting an index other than $i^\star$ by $(f(n)-1)\exp(-\varepsilon/(2\Delta)) \le f(n)\exp(-\varepsilon n/(4f(n)^2))$. □

C.1.3 Putting Things Together

Theorem C.8 (Guarantees of Algorithm 1, Restatement of Theorem 3.2)**.** *Let* $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$ *be a countable collection of languages and let* $\mathcal{D}$ *be any distribution over* $\mathcal{U}$*. Suppose Assumption C.1 holds. Let* $\varepsilon > 0$ *and let* $f : \mathbb{N} \to \mathbb{N}$ *satisfy* $f(n) \to \infty$*. Then for all* $n \ge n_0$*, Algorithm 1 has the following guarantees:*

- ***Privacy.*** *Algorithm 1 is* ε*-differentially private.*
- ***Utility.*** *The identification error satisfies*

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) \le 2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big) + f(n)\exp\Big(-\frac{\varepsilon n}{4f(n)^2}\Big). \tag{8}$$

Proof. **Privacy.** This directly follows from Lemma C.3.

Utility. We note that the excess error $\mathrm{err}_{\mathcal{D}}(L_{\widehat{i}}) - \inf_{L\in\mathcal{C}} \mathrm{err}_{\mathcal{D}}(L)$ lies in $[0,1]$ and equals 0 when $\widehat{i} = i^\star$. Therefore

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) = \mathop{\mathbb{E}}_{S,r}\left[\mathrm{err}_{\mathcal{D}}(L_{\widehat{i}}) - \inf_{L\in\mathcal{C}} \mathrm{err}_{\mathcal{D}}(L)\right] \le \Pr_{S,r}[\widehat{i} \neq i^\star].$$

We decompose using the event $\mathcal{E}$ from Lemma C.4:

$$\begin{aligned}\Pr[\widehat{i} \neq i^\star] &= \Pr[\widehat{i} \neq i^\star \mid \mathcal{E}]\Pr[\mathcal{E}] + \Pr[\widehat{i} \neq i^\star \mid \mathcal{E}^c]\Pr[\mathcal{E}^c] \\ &\le \Pr[\widehat{i} \neq i^\star \mid \mathcal{E}] + \Pr[\mathcal{E}^c].\end{aligned}$$

Applying Lemma C.7 and Lemma C.4:

$$\Pr[\widehat{i} \neq i^\star] \le f(n)\exp\Big(-\frac{\varepsilon n}{4f(n)^2}\Big) + 2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big).$$

□

Corollary C.9 (Explicit Rates). *Let $\varepsilon > 0$ be fixed. The bound (8) gives the following rates.*

(a) For $f(n) = \lceil\sqrt{c\log n}\,\rceil$ with any constant $c > 0$, $\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) \le \exp\big(-\Omega(\min\{1,\varepsilon\}\cdot n/\log n)\big)$.

(b) For $f(n) = \lceil n^\alpha\rceil$ with $\alpha \in (0, 1/2)$, $\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) \le \exp\big(-\Omega(\min\{1,\varepsilon\}\cdot n^{1-2\alpha})\big)$.

(c) For every $r\colon \mathbb{N} \to [0,\infty)$ with $r(n) = o(n)$, let $R(n) := 2r(n) + \log^2 n$ and $f(n) := \lceil\sqrt{n/(8R(n))}\,\rceil$. Then $\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,f}, \mathcal{D}, \mathcal{C}, n) \le \exp(-\min\{1,\varepsilon\}\cdot r(n))$ for every $n \ge n_0$ with $n \ge 335R(n)$ and $\frac{3}{4}\min\{1,\varepsilon\}\log^2 n \ge \log 3 + \frac{1}{2}\log n$; since $r(n) = o(n)$, this holds for all large n.

Proof. Let $m := \min\{1, 2\varepsilon\}$, so that $m \ge \min\{1,\varepsilon\}$. Both exponents in (8) are at least $m\,n/(8f(n)^2)$, so the bound is at most $3f(n)\exp(-m\,n/(8f(n)^2))$. In each case $f(n) \to \infty$, so $f(n)^2 \le \frac{4}{3}x^2$ for large n, where x is the quantity inside the ceiling.

(a) Here $n/(8f(n)^2) \ge 3n/(32c\log n)$, and the prefactor $3f(n) = O(\sqrt{\log n})$ is absorbed into the exponent.

(b) Here $n/(8f(n)^2) \ge 3n^{1-2\alpha}/32$, and the prefactor $3f(n) = O(n^\alpha)$ is absorbed because $n^{1-2\alpha}$ grows faster than $\log n$.

(c) Since $r(n) = o(n)$, also $R(n) = o(n)$, so $f(n) \to \infty$. When $n \ge 335R(n)$, the quantity $x = \sqrt{n/(8R(n))}$ is at least 6.47, so $f(n)^2 \le (x+1)^2 \le \frac{4}{3}x^2$ and $f(n) \le x + 1 \le \sqrt{n}$. Hence $n/(8f(n)^2) \ge \frac{3}{4}R(n) = \frac{3}{2}r(n) + \frac{3}{4}\log^2 n$. With $m \ge \min\{1,\varepsilon\}$, the bound is at most $3\sqrt{n}\exp(-\min\{1,\varepsilon\}\,r(n))\exp(-\frac{3}{4}\min\{1,\varepsilon\}\log^2 n)$, and the second condition makes $3\sqrt{n}\exp(-\frac{3}{4}\min\{1,\varepsilon\}\log^2 n) \le 1$. □

C.2 Proof of Theorem 3.3 (Approximate DP Identification)

For the noisy empirical errors $\widetilde{\mathrm{err}}_S(L_i) := \mathrm{err}_S(L_i) + Z_i$, we define the *noisy margin*

$$\widetilde{M}_S(i) := \begin{cases} 1, & \text{if } i = 1, \\ \min_{j\in[i-1]}\big(\widetilde{\mathrm{err}}_S(L_j) - \widetilde{\mathrm{err}}_S(L_i)\big), & \text{if } i \ge 2. \end{cases} \tag{9}$$

C.2.1 Privacy Analysis

Lemma C.10 (ℓ_2-Sensitivity of the Empirical Error Vector). *Define $v\colon \mathcal{U}^n \to \mathbb{R}^{f(n)}$ by*

$$v(S) := \big(\mathrm{err}_S(L_1), \ldots, \mathrm{err}_S(L_{f(n)})\big).$$

Then v has ℓ_2-sensitivity $\sqrt{f(n)}/n$.

Proof. For neighboring $S, S' \in \mathcal{U}^n$, by (6), we have $|\mathrm{err}_S(L_i) - \mathrm{err}_{S'}(L_i)| \le 1/n$ for each $i \in [f(n)]$. Therefore

$$\|v(S) - v(S')\|_2^2 = \sum_{i=1}^{f(n)} \big|\mathrm{err}_S(L_i) - \mathrm{err}_{S'}(L_i)\big|^2 \le \sum_{i=1}^{f(n)} \frac{1}{n^2} = \frac{f(n)}{n^2}.$$

Taking the square root gives $\|v(S) - v(S')\|_2 \le \sqrt{f(n)}/n$. □

Lemma C.11 (Privacy)**.** *For $\varepsilon, \delta \in (0,1)$, Algorithm 2 is (ε, δ)-differentially private.*

Proof. By Lemma C.10, the empirical error vector has ℓ_2-sensitivity $\Delta_2 = \sqrt{f(n)}/n$. The Gaussian mechanism (Lemma 2.4) with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{2\log(1.25/\delta)} = \frac{\sqrt{f(n)}}{\varepsilon n}\sqrt{2\log(1.25/\delta)}$ ensures that the noisy vector $\big(\widetilde{\mathrm{err}}_S(L_1), \ldots, \widetilde{\mathrm{err}}_S(L_{f(n)})\big)$ is (ε, δ)-differentially private. The subsequent margin selection (lines 3–4 of Algorithm 2) is a deterministic function of this noisy vector and is therefore (ε, δ)-DP by post-processing (Lemma 2.2). □

C.2.2 Utility Analysis

The utility analysis proceeds in three steps: (i) a uniform concentration event $\mathcal{E}$ under which all empirical errors in the horizon are close to their population values (Lemma C.4); (ii) a noise concentration event $\mathcal{F}$ under which the Gaussian perturbations are small (Lemma C.12); (iii) on $\mathcal{E} \cap \mathcal{F}$, the deterministic margin-selection rule outputs $i^\star$ (Lemma C.13).

Lemma C.12 (Noise Concentration)**.** *Define the event*

$$\mathcal{F} := \Big\{ |Z_i| \le \frac{1}{4f(n)}, \quad \forall i \in [f(n)] \Big\}.$$

Then

$$\Pr[\mathcal{F}^c] \le 2f(n)\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big).$$

Proof. Each $Z_i \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{2f(n)\log(1.25/\delta)}{\varepsilon^2 n^2}$. By the Gaussian tail bound (Lemma A.3), for each $i \in [f(n)]$,

$$\Pr\Big[|Z_i| > \frac{1}{4f(n)}\Big] \le 2\exp\Big(-\frac{1/(16f(n)^2)}{2\sigma^2}\Big) = 2\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big).$$

A union bound over $i \in [f(n)]$ yields

$$\Pr[\mathcal{F}^c] \le \sum_{i=1}^{f(n)} 2\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big) = 2f(n)\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big).$$

□

Lemma C.13 (Deterministic Correctness on $\mathcal{E} \cap \mathcal{F}$)**.** *On the event $\mathcal{E} \cap \mathcal{F}$, Algorithm 2 outputs $\widehat{i} = i^\star$.*

Proof. Assume $\mathcal{E} \cap \mathcal{F}$ holds throughout. Since $\widetilde{\mathrm{err}}_S(L_i) = \mathrm{err}_S(L_i) + Z_i$, the triangle inequality gives

$$\big|\widetilde{\mathrm{err}}_S(L_i) - \mathrm{err}_{\mathcal{D}}(L_i)\big| \le \big|\mathrm{err}_S(L_i) - \mathrm{err}_{\mathcal{D}}(L_i)\big| + |Z_i| \le \frac{1}{4f(n)} + \frac{1}{4f(n)} = \frac{1}{2f(n)} \tag{10}$$

for all $i \in [f(n)]$. We now verify that $i^\star$ is the largest index passing the noisy margin test.

Part 1: $i^\star$ passes the noisy margin test. If $i^\star = 1$, then $\widetilde{M}_S(1) = 1 > 2/f(n)$. If $i^\star \ge 2$, then for every $j < i^\star$, using (10) on both indices,

$$\begin{aligned}\widetilde{\mathrm{err}}_S(L_j) - \widetilde{\mathrm{err}}_S(L_{i^\star}) &\ge \big(\mathrm{err}_{\mathcal{D}}(L_j) - \tfrac{1}{2f(n)}\big) - \big(\mathrm{err}_{\mathcal{D}}(L_{i^\star}) + \tfrac{1}{2f(n)}\big) \\ &= \big(\mathrm{err}_{\mathcal{D}}(L_j) - \mathrm{err}_{\mathcal{D}}(L_{i^\star})\big) - \frac{1}{f(n)}\end{aligned}$$

$$\geq c_{\mathrm{gap}} - \frac{1}{f(n)}.$$

For $n \geq n_0$ we have $c_{\mathrm{gap}} > 3/f(n)$, so $c_{\mathrm{gap}} - 1/f(n) > 2/f(n)$. Taking the minimum over $j < i^\star$ gives $\widetilde{M}_S(i^\star) > 2/f(n)$.

Part 2: No $i > i^\star$ passes the noisy margin test. Fix $i > i^\star$. By the optimality of $i^\star$, $\mathrm{err}_{\mathcal{D}}(L_i) \geq \mathrm{err}_{\mathcal{D}}(L_{i^\star})$. Using (10),

$$\begin{aligned}
\widetilde{\mathrm{err}}_S(L_{i^\star}) - \widetilde{\mathrm{err}}_S(L_i) &\leq \left(\mathrm{err}_{\mathcal{D}}(L_{i^\star}) + \tfrac{1}{2f(n)}\right) - \left(\mathrm{err}_{\mathcal{D}}(L_i) - \tfrac{1}{2f(n)}\right) \\
&= \left(\mathrm{err}_{\mathcal{D}}(L_{i^\star}) - \mathrm{err}_{\mathcal{D}}(L_i)\right) + \frac{1}{f(n)} \\
&\leq \frac{1}{f(n)}.
\end{aligned}$$

Since $j = i^\star$ is among the candidates in the minimum defining $\widetilde{M}_S(i)$, we have $\widetilde{M}_S(i) \leq 1/f(n) < 2/f(n)$, so i fails the noisy margin test.

Since $i^\star$ passes and no $i > i^\star$ passes, the algorithm outputs $\widehat{i} = i^\star$. □

C.2.3 Putting Things Together

Theorem C.14 (Guarantees of Algorithm 2, Restatement of Theorem 3.3)**.** *Let $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$ be a countable collection of languages and let $\mathcal{D}$ be any distribution over $\mathcal{U}$. Suppose Assumption C.1 holds. Let $\varepsilon, \delta \in (0,1)$, and let $f : \mathbb{N} \to \mathbb{N}$ satisfy $f(n) \to \infty$. Then for all $n \geq n_0$, Algorithm 2 has the following guarantees:*

- ***Privacy.*** *Algorithm 2 is (ε, δ)-differentially private.*
- ***Utility.*** *The identification error satisfies*

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,\delta,f}, \mathcal{D}, \mathcal{C}, n) \leq 2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big) + 2f(n)\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big). \tag{11}$$

Proof. **Privacy.** This directly follows from Lemma C.11.

Utility. By Lemma C.13, $\widehat{i} = i^\star$ whenever $\mathcal{E} \cap \mathcal{F}$ holds. Therefore

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,\delta,f}, \mathcal{D}, \mathcal{C}, n) \leq \Pr_{S,r}[\widehat{i} \neq i^\star] \leq \Pr[\mathcal{E}^c] + \Pr[\mathcal{F}^c].$$

Applying Lemma C.4 and Lemma C.12:

$$\mathrm{IdErr}(\mathcal{A}^{\mathrm{id}}_{\varepsilon,\delta,f}, \mathcal{D}, \mathcal{C}, n) \leq 2f(n)\exp\Big(-\frac{n}{8f(n)^2}\Big) + 2f(n)\exp\Big(-\frac{\varepsilon^2 n^2}{64 f(n)^3 \log(1.25/\delta)}\Big).$$

□

Remark C.15 (Comparison with Pure DP and Interpretation)**.** The approximate DP analysis is structurally simpler: once both concentration events hold, correctness is deterministic (no exponential mechanism). The cost is that the noise event $\mathcal{F}$ depends on ε, δ, and $f(n)$ jointly. The bound (11) decomposes into a statistical term $2f(n)\exp(-n/(8f(n)^2))$, identical to the pure DP case, and a privacy term scaling as $\exp(-\varepsilon^2 n^2/f(n)^3)$ rather than $\exp(-\varepsilon n/f(n)^2)$, reflecting the different noise mechanism. The resulting rates are made explicit in Corollary C.16.

Corollary C.16 (Explicit Rates for Approximate DP)**.** *Let $\varepsilon, \delta \in (0,1)$ and write $\ell := \log(1.25/\delta)$. The bound (11) gives the following rates.*

(a) For $f(n) = \lceil n^\alpha \rceil$ with $\alpha \in (0, 1/2)$, $\mathrm{IdErr} \leq O(n^\alpha)\big(\exp(-\Omega(n^{1-2\alpha})) + \exp(-\Omega(\varepsilon^2 n^{2-3\alpha}/\ell))\big)$. The first term dominates when $\varepsilon^2 n^{1-\alpha}/\ell \to \infty$, for instance for constant ε and $\delta \geq n^{-O(1)}$.

(b) For $f(n) = \lceil \sqrt{c \log n} \rceil$ *with any constant* $c > 0$, $\mathrm{IdErr} \le O(\sqrt{\log n})\big(\exp(-\Omega(n/\log n)) + \exp(-\Omega(\varepsilon^2 n^2/(\ell \log^{3/2} n)))\big)$*, and the first term dominates for constant* ε *and* $\delta \ge n^{-O(1)}$.

(c) For every $r\colon \mathbb{N} \to [0, \infty)$ *with* $r(n) = o(n)$*, constant* ε *and* $\delta \ge n^{-O(1)}$*, the horizon of Corollary C.9(c) gives* $\mathrm{IdErr} \le \exp(-r(n))$ *for every* $n \ge n_0$ *with* $n \ge 335R(n)$*,* $\sqrt{nR(n)} \ge 3.3\,\ell/\varepsilon^2$ *and* $\frac{3}{4}\log^2 n \ge \log 4 + \frac{1}{2}\log n$*; these hold for all large* n.

(d) For $\varepsilon = n^{-\gamma}$ *with* $\gamma > 0$ *and* $f(n) = \lceil \sqrt{c \log n} \rceil$, $\mathrm{IdErr} \le O(\sqrt{\log n})\big(\exp(-\Omega(n/\log n)) + \exp(-\Omega(n^{2-2\gamma}/(\ell \log^{3/2} n)))\big)$*, and the privacy term dominates when* $\gamma \ge 1/2$.

Proof. Parts (a), (b) and (d) follow by substituting the horizon into (11), using $f(n)^2 \le \frac{4}{3}x^2$ for the quantity x inside the ceiling and large n, and bounding the prefactor $2f(n)$ by $O(n^\alpha)$ or $O(\sqrt{\log n})$; the threshold n_0 does not depend on ε, which part (d) needs. For part (c), the proof of Corollary C.9(c) gives, when $n \ge 335R(n)$, both $n/(8f(n)^2) \ge \frac{3}{4}R(n)$ with $R(n) = 2r(n) + \log^2 n$ and $f(n)^2 \le \frac{4}{3} \cdot \frac{n}{8R(n)} = \frac{n}{6R(n)}$. Hence the privacy exponent satisfies

$$\frac{\varepsilon^2 n^2}{64 f(n)^3 \ell} \ge \frac{\varepsilon^2 n^2 (6R(n))^{3/2}}{64\, n^{3/2}\, \ell} = \frac{6^{3/2}\varepsilon^2}{64} \cdot \frac{\sqrt{nR(n)}}{\ell} \cdot R(n),$$

which is at least $\frac{3}{4}R(n)$ once $\sqrt{nR(n)} \ge 3.3\,\ell/\varepsilon^2$ (as $48/6^{3/2} < 3.3$); this holds for large n because $\sqrt{nR(n)}/\ell \ge \sqrt{n}\log n/\ell \to \infty$ when $\ell = O(\log n)$. The bound is then at most $4f(n)\exp(-\frac{3}{4}R(n)) \le 4\sqrt{n}\exp(-\frac{3}{2}r(n) - \frac{3}{4}\log^2 n)$, which the last condition makes at most $\exp(-r(n))$. □

D Proofs for Section 4 (Upper Bounds for DP Generation)

We fix an enumeration of the universe $\mathcal{U} = \{u_i\}_{i\in\mathbb{N}}$ and assume that every language $L \in \mathcal{C}$ is infinite. For a dataset $S = (x_1, \dots, x_n) \in \mathcal{U}^n$ and $x \in \mathcal{U}$, we define the *multiplicity*

$$N_S(x) := \sum_{t=1}^{n} \mathbf{1}\{x_t = x\}.$$

Definition D.1 (Witness Index $i(\mathcal{C}, \mathcal{D})$)**.** Given a collection $\mathcal{C}$ of languages and a distribution $\mathcal{D}$ over $\mathcal{U} = \{u_i\}_{i\in\mathbb{N}}$, the *witness index* is

$$i(\mathcal{C}, \mathcal{D}) := \min\big\{ i \in \mathbb{N} : \forall\, L \in \mathcal{C} \text{ with } L \not\subseteq \mathrm{supp}(\mathcal{D}),\ \exists\, k \le i \text{ s.t. } u_k \in L \setminus \mathrm{supp}(\mathcal{D}) \big\},$$

with the convention $\min \emptyset = \infty$.

In words, $i(\mathcal{C}, \mathcal{D})$ is the smallest index such that every language not fully contained in $\mathrm{supp}(\mathcal{D})$ has a *witness* — a string in the language but outside the support — appearing within the first $i(\mathcal{C}, \mathcal{D})$ elements of $\mathcal{U}$. For any finite collection $\mathcal{C}$, we always have $i(\mathcal{C}, \mathcal{D}) < \infty$.

We work under the following two assumptions throughout this section.

Assumption D.2 (Finite witness index)**.** *The witness index is finite, i.e.,* $i(\mathcal{C}, \mathcal{D}) < \infty$.

Assumption D.3 (Support contains a reference language)**.** *There exists an index* $i^\star \in \mathbb{N}$ *such that* $L_{i^\star} \subseteq \mathrm{supp}(\mathcal{D})$*. Let* $i^\star$ *denote the smallest such index.*

D.1 Proof of Theorem 4.4 (Pure DP Generation with a Public Witness Bound)

We first consider the setting where a public upper bound on the witness index is available.

Assumption D.4 (Public Witness Bound)**.** *There is a public integer* $W \ge i(\mathcal{C}, \mathcal{D})$.

Here "public" means that W is auxiliary input given to the algorithm before observing the private sample $S \sim \mathcal{D}^n$; it is not part of the data protected by differential privacy.

D.1.1 Setup and Notations

For each language L_i and the witness window $[W]$, define

$$I_i(W) := \{k \in [W] : u_k \in L_i\} \tag{12}$$

to be the set of indices of strings in L_i within the witness window. Define the *minimum prefix count*

$$a_i^W(S) := \begin{cases} \min_{k \in I_i(W)} N_S(u_k), & \text{if } I_i(W) \neq \emptyset, \\ n, & \text{if } I_i(W) = \emptyset, \end{cases} \tag{13}$$

the *deficit*

$$d_i^W(S) := \big(g(n) - a_i^W(S)\big)_+, \tag{14}$$

and the *score*

$$q^W(S, i) := -d_i^W(S). \tag{15}$$

Here $g : \mathbb{N} \to \mathbb{N}$ is a count threshold satisfying $g(n) \to \infty$; its role is analogous to the margin $2/f(n)$ in identification. The score is at most 0 (achieved when the prefix is well-covered) and equals $-g(n)$ when a witness string is never observed.

We also define the following quantities associated with the reference language $L_{i^\star}$:

$$I_W^\star := I_{i^\star}(W) = \{k \in [W] : u_k \in L_{i^\star}\}, \qquad p_W^\star := \min_{k \in I_W^\star} \Pr_{x \sim \mathcal{D}}[x = u_k]. \tag{16}$$

We use the convention $p_W^\star := 1$ if $I_W^\star = \emptyset$; in particular, the condition $g(n) \leq np_W^\star/2$ forces $g(n) \leq n/2$. Since $L_{i^\star} \subseteq \mathrm{supp}(\mathcal{D})$, every string u_k with $k \in I_W^\star$ has positive probability under $\mathcal{D}$, so $p_W^\star \in (0, 1]$.

D.1.2 Privacy Analysis

Lemma D.5 (Sensitivity of q^W)**.** *The score function q^W defined in (15) has sensitivity 1.*

Proof. Fix $i \in [f(n)]$ and let $S, S' \in \mathcal{U}^n$ be neighboring datasets. For every $x \in \mathcal{U}$, the multiplicity satisfies $|N_S(x) - N_{S'}(x)| \leq 1$.

If $I_i(W) \neq \emptyset$, then $a_i^W(S) = \min_{k \in I_i(W)} N_S(u_k)$. Since each $N_S(u_k)$ changes by at most 1 and the minimum of functions preserves Lipschitz constants,

$$|a_i^W(S) - a_i^W(S')| \leq 1.$$

If $I_i(W) = \emptyset$, then $a_i^W(S) = a_i^W(S') = n$ and the bound holds trivially.

The map $x \mapsto (g(n) - x)_+$ is 1-Lipschitz, so $|d_i^W(S) - d_i^W(S')| \leq 1$. Therefore

$$|q^W(S, i) - q^W(S', i)| = |d_i^W(S) - d_i^W(S')| \leq 1.$$

□

Remark D.6 (Comparison with Identification)**.** The generation score has constant sensitivity 1 (integer multiplicity counts), whereas the identification score has sensitivity $\Theta(f(n)^2/n)$ (empirical error averages amplified by the $f(n)^2$ coefficient). Moreover, the score separation is structurally simpler: there is no case analysis for $i < i^\star$ vs. $i > i^\star$, since the witness test treats all bad languages uniformly, and the score gap is $g(n)$ rather than 1. Together, these give a qualitatively smaller price of privacy: the non-private rate $\exp(-\Theta(n))$ of Høgsgaard and Pabbaraju [2026] is matched when $\varepsilon = \Omega(1)$ and a window and mass floor are public, and the only cost is the factor of ε in the exponent when $\varepsilon < 1$, compared with identification where $r(n) = o(n)$ rather than $\Theta(n)$.

Lemma D.7 (Privacy)**.** *Algorithm 3 is ε-differentially private.*

Proof. The language selection on line 2 applies the exponential mechanism with score q^W over range $[f(n)]$. By Lemma D.5, q^W has sensitivity $\Delta = 1$, so this step is ε-DP. The subsequent output step (lines 3–4) is a deterministic function of the private output $\widehat{i}$ and public randomness $\widehat{j}$, and is therefore ε-DP by post-processing (Lemma 2.2). □

D.1.3 Utility Analysis

The utility analysis proceeds in three steps: (i) a coverage event $\mathcal{E}_W$ under which every relevant string of $L_{i^\star}$ in the witness window appears at least $g(n)$ times (Lemma D.8); (ii) on $\mathcal{E}_W$, the good language $L_{i^\star}$ has score 0 while every bad language has score $-g(n)$, creating a score gap of $g(n)$ (Lemma D.9); (iii) the exponential mechanism exploits this gap to select a good language, and the output step produces a novel string from its support (Lemma D.10).

Lemma D.8 (Coverage of the Good Prefix)**.** *Define the event*

$$\mathcal{E}_W := \big\{N_S(u_k) \geq g(n) \quad \forall\, k \in I^\star_W\big\}.$$

If $g(n) \leq np^\star_W/2$ for all large enough n, then

$$\Pr_{S\sim\mathcal{D}^n}[\mathcal{E}^c_W] \leq |I^\star_W| \exp\Big(-\frac{np^\star_W}{8}\Big).$$

Proof. Fix $k \in I^\star_W$ and let $p_k := \Pr_{x\sim\mathcal{D}}[x = u_k]$, so that $p_k \geq p^\star_W > 0$. The multiplicity $N_S(u_k) = \sum_{t=1}^n \mathbf{1}\{x_t = u_k\}$ is a sum of n independent Bernoulli(p_k) random variables with mean $\mu_k = np_k$. Since $g(n) \leq np^\star_W/2 \leq np_k/2 = \mu_k/2$, the event $\{N_S(u_k) < g(n)\}$ is contained in $\{N_S(u_k) \leq \mu_k/2\}$. By the Chernoff bound (Lemma A.2) with $t = 1/2$,

$$\Pr\big[N_S(u_k) \leq \mu_k/2\big] \leq \exp\Big(-\frac{\mu_k}{8}\Big) \leq \exp\Big(-\frac{np^\star_W}{8}\Big).$$

A union bound over $k \in I^\star_W$ yields

$$\Pr[\mathcal{E}^c_W] \leq \sum_{k\in I^\star_W} \exp\Big(-\frac{np^\star_W}{8}\Big) = |I^\star_W| \exp\Big(-\frac{np^\star_W}{8}\Big).$$

□

Lemma D.9 (Score separation on $\mathcal{E}_W$)**.** *On the event $\mathcal{E}_W$:*

(a) $q^W(S, i^\star) = 0$.

(b) $q^W(S, i) = -g(n)$ *for every* $i \in [f(n)]$ *such that* $L_i \not\subseteq \mathrm{supp}(\mathcal{D})$.

Consequently, on $\mathcal{E}_W$ the set of good languages has margin at least $g(n)$ over the bad ones.

Proof. **Part (a): The good language achieves score** 0**.** On $\mathcal{E}_W$, every $k \in I^\star_W$ satisfies $N_S(u_k) \geq g(n)$. If $I^\star_W \neq \emptyset$, then $a^W_{i^\star}(S) = \min_{k\in I^\star_W} N_S(u_k) \geq g(n)$, so $d^W_{i^\star}(S) = (g(n) - a^W_{i^\star}(S))_+ = 0$. If $I^\star_W = \emptyset$ (i.e., $L_{i^\star}$ has no string with index $\leq W$), then $a^W_{i^\star}(S) = n \geq g(n)$, so $d^W_{i^\star}(S) = 0$ as well. In either case, $q^W(S, i^\star) = -d^W_{i^\star}(S) = 0$.

Part (b): Every bad language achieves score $-g(n)$**.** Fix $i \in [f(n)]$ with $L_i \not\subseteq \mathrm{supp}(\mathcal{D})$. Since $W \geq i(\mathcal{C}, \mathcal{D})$ (Assumption D.4), by Definition D.1 there exists $k \leq W$ such that $u_k \in L_i \setminus \mathrm{supp}(\mathcal{D})$. In particular, $k \in I_i(W)$ (so $I_i(W) \neq \emptyset$) and $u_k \notin \mathrm{supp}(\mathcal{D})$, meaning u_k can never appear in any sample from $\mathcal{D}$. Therefore $N_S(u_k) = 0$, which gives $a^W_i(S) \leq N_S(u_k) = 0$. Hence $d^W_i(S) = (g(n) - 0)_+ = g(n)$ and $q^W(S, i) = -g(n)$. □

Lemma D.10 (Private Selection and Output)**.** *Define*

$$\beta_{\mathrm{sel}} := f(n) \exp\Big(-\frac{\varepsilon g(n)}{2}\Big), \qquad \beta_{\mathrm{out}} := \exp\Big(-\frac{n}{2}\Big).$$

On the event $\mathcal{E}_W$:

(a) The exponential mechanism selects a good language (i.e., $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$) with probability at least $1 - \beta_{\mathrm{sel}}$.

(b) Conditioned on $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$, the output $\widehat{x}$ satisfies $\widehat{x} \in \mathrm{supp}(\mathcal{D}) \setminus S$ with probability at least $1 - \beta_{\mathrm{out}}$.

Proof. **Part (a): Language selection.** Fix a sample S in $\mathcal{E}_W$, and let $\mathcal{G} := \{i \in [f(n)] : L_i \subseteq \mathrm{supp}(\mathcal{D})\}$, which contains $i^\star$ because $f(n) \geq i^\star$. By Lemma D.9, $\mathcal{G}$ has margin at least $q^W(S, i^\star) - (-g(n)) = g(n)$ over $[f(n)] \setminus \mathcal{G}$, and by Lemma D.5 the score has sensitivity $\Delta = 1$. Lemma A.5(a) bounds the probability of selecting a bad language by $f(n)\exp(-\varepsilon g(n)/2) = \beta_{\mathrm{sel}}$.

Part (b): Output novelty. Suppose $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$. The algorithm outputs the $\widehat{j}$-th smallest-indexed string from the set $T_{\widehat{i}} := L_{\widehat{i}} \cap \{u_{W+1}, u_{W+2}, \ldots\}$, where $\widehat{j} \sim \mathrm{Unif}([2^n])$. Since $L_{\widehat{i}}$ is infinite, $T_{\widehat{i}}$ is infinite and in particular $|T_{\widehat{i}}| \geq 2^n$. Moreover, since $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$, every string in $T_{\widehat{i}}$ belongs to $\mathrm{supp}(\mathcal{D})$. The sample S has size n, so at most n strings from $T_{\widehat{i}}$ can appear in S. Since $\widehat{j}$ is uniform over $[2^n]$ and independent of S (it is public randomness),

$$\Pr[\widehat{x} \in S \mid L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})] \leq \frac{n}{2^n} \leq \exp\Big(-\frac{n}{2}\Big)$$

for all $n \geq 14$, by Lemma A.6(c). □

D.1.4 Putting Things Together

Theorem D.11 (Guarantees of Algorithm 3, Restatement of Theorem 4.4)**.** *Let $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$ be a countable collection of infinite languages and let $\mathcal{D}$ be any distribution over $\mathcal{U}$. Suppose Assumptions D.2, D.3, and D.4 hold. Let $\varepsilon > 0$, and let $f, g : \mathbb{N} \to \mathbb{N}$ satisfy $f(n) \to \infty$, $g(n) \to \infty$, and $g(n) \leq np^\star_W/2$ for all large enough n. Suppose further that $f(n) \geq i^\star$. Then for all sufficiently large n, Algorithm 3 has the following guarantees:*

- ***Privacy.*** *Algorithm 3 is ε-differentially private.*
- ***Utility.*** *The generation error satisfies*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) \leq \underbrace{|I^\star_W| \exp\Big(-\frac{np^\star_W}{8}\Big)}_{\text{coverage}} + \underbrace{f(n)\exp\Big(-\frac{\varepsilon g(n)}{2}\Big)}_{\text{privacy}} + \underbrace{\exp\Big(-\frac{n}{2}\Big)}_{\text{collision}}. \tag{17}$$

Proof. **Privacy.** This directly follows from Lemma D.7.

Utility. The generation fails (i.e., $\widehat{x} \notin \mathrm{supp}(\mathcal{D}) \setminus S$) only if at least one of the following three bad events occurs:

(i) The coverage event $\mathcal{E}_W$ fails: the good prefix of $L_{i^\star}$ is not well-covered.

(ii) The exponential mechanism selects a bad language: $L_{\widehat{i}} \not\subseteq \mathrm{supp}(\mathcal{D})$.

(iii) The selected language is good but the output string is already in S: $\widehat{x} \in S$.

By a union bound,

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) \leq \Pr[\mathcal{E}_W^{\mathrm{c}}] + \Pr[L_{\widehat{i}} \not\subseteq \mathrm{supp}(\mathcal{D}) \mid \mathcal{E}_W] + \Pr[\widehat{x} \in S \mid L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})].$$

Applying Lemma D.8 (term (i)), Lemma D.10(a) (term (ii)), and Lemma D.10(b) (term (iii)):

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) \leq |I^\star_W| \exp\Big(-\frac{np^\star_W}{8}\Big) + f(n)\exp\Big(-\frac{\varepsilon g(n)}{2}\Big) + \exp\Big(-\frac{n}{2}\Big).$$

□

Remark D.12 (Interpretation of the bound)**.** The bound (17) consists of three terms: a **coverage term** $|I^\star_W|\exp(-np^\star_W/8)$, present even without privacy, controlling under-representation of $L_{i^\star}$ strings in the witness window; a **privacy term** $f(n)\exp(-\varepsilon g(n)/2)$, controlling the exponential mechanism's failure to select a good language (with direct ε-dependence thanks to constant sensitivity); and a negligible **collision term** $\exp(-n/2)$. The no-public-bound case (23) has the same structure, but the coverage term now depends on $h(n)$ through $I^\star_h$ and $p^\star_h$, and the privacy term acquires an extra $h(n)$ prefactor from the enlarged search space $[f(n)] \times [h(n)]$.

Corollary D.13 (Fixed public window, Restatement of Corollary 4.6)**.** *Under the conditions of Theorem D.11, let $\varepsilon > 0$ be fixed and let f satisfy $f(n) \geq i^\star$.*

(a) If the learner is also given a constant $p_0 > 0$ with $p_0 \le p^\star_W$, then $g(n) = \max\{1, \lfloor np_0/2 \rfloor\}$ and $\log f(n) = o(n)$ give

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) \le |I^\star_W| \exp\Big(-\frac{np^\star_W}{8}\Big) + f(n)\exp\Big(-\frac{\varepsilon np_0}{8}\Big) + \exp\Big(-\frac{n}{2}\Big)$$
$$= \exp\big(-\Omega(\min\{1,\varepsilon\}\cdot n)\big).$$

(b) For any $r(n) \to \infty$ with $r(n) = o(n)$, $g(n) = \lfloor r(n) \rfloor$ and $\log f(n) = o(r(n))$ give $\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,W}, \mathcal{D}, \mathcal{C}, n) = \exp(-\Omega(\min\{1,\varepsilon\}\cdot r(n)))$, without knowledge of $p^\star_W$.

Proof. (a) Since $p_0 \le p^\star_W$, we have $g(n) = \lfloor np_0/2 \rfloor \le np^\star_W/2$ for large n, so the condition of Theorem D.11 is satisfied. Substituting $g(n) \ge np_0/2 - 1 \ge np_0/4$ (for large n) into (17) gives the privacy term $f(n)\exp(-\varepsilon g(n)/2) \le f(n)\exp(-\varepsilon np_0/8)$, and the factor $f(n)$ is absorbed because $\log f(n) = o(n)$. The coverage term is $\exp(-\Omega(n))$ because $|I^\star_W| \le W$ and $p^\star_W > 0$ do not depend on n, and the collision term is $\exp(-n/2)$.

(b) Because W is fixed, $p^\star_W > 0$ does not depend on n, and $r(n) = o(n)$ gives $g(n) \le r(n) \le np^\star_W/2$ for all large n, so Theorem D.11 applies. Its coverage term is $\exp(-\Omega(n))$, its privacy term is $f(n)\exp(-\varepsilon\lfloor r(n)\rfloor/2) = \exp(-\Omega(\varepsilon\, r(n)))$ because $\log f(n) = o(r(n))$, and its collision term is $\exp(-n/2)$. Since $r(n) = o(n)$, all three terms are $\exp(-\Omega(\min\{1,\varepsilon\}\cdot r(n)))$. □

D.2 Proof of Theorem 4.5 (Pure DP Generation without a Public Witness Bound)

We now remove the assumption that a public upper bound on the witness index is available. The main challenge is that the witness window $i(\mathcal{C}, \mathcal{D})$ is unknown and depends on the private data through $\mathcal{D}$. Our approach is to jointly search over both the language index i and a candidate witness threshold t, using a score that rewards large thresholds (indicating progress past the witness window) while penalizing languages that fail the witness test at threshold t.

D.2.1 Setup and Notation

Let $f, g, h : \mathbb{N} \to \mathbb{N}$ be functions satisfying $f(n) \to \infty$, $g(n) \to \infty$, and $h(n) \to \infty$. The parameter $f(n)$ controls the *language horizon*, $g(n)$ is the *count threshold* (as in the public-bound setting), and $h(n)$ is the *threshold horizon* that upper bounds the witness index for large enough n.

For each language L_i and candidate threshold $t \in [h(n)]$, define

$$I_i(t) := \{k \in [t] : u_k \in L_i\}. \tag{18}$$

For each $S \in \mathcal{U}^n$ and each pair $(i,t) \in [f(n)] \times [h(n)]$, we define the *minimum prefix count*

$$a_{i,t}(S) := \begin{cases} \min_{k \in I_i(t)} N_S(u_k), & \text{if } I_i(t) \neq \emptyset, \\ n, & \text{if } I_i(t) = \emptyset, \end{cases} \tag{19}$$

the deficit

$$d_{i,t}(S) := \big(g(n) - a_{i,t}(S)\big)_+, \tag{20}$$

and the score

$$q_{\mathrm{pair}}(S, (i,t)) := t - \frac{h(n)}{g(n)}\, d_{i,t}(S). \tag{21}$$

The term t rewards larger thresholds, while the penalty $\frac{h(n)}{g(n)}\, d_{i,t}(S)$ suppresses pairs where the language fails the witness test at threshold t. The coefficient $h(n)/g(n)$ is chosen so that a bad language with full deficit $g(n)$ incurs a penalty of exactly $h(n)$, ensuring its score is at most $t - h(n) \le 0$.

We also define quantities associated with the reference language $L_{i^\star}$ at the threshold horizon:

$$I^\star_h := I_{i^\star}(h(n)) = \{k \in [h(n)] : u_k \in L_{i^\star}\}, \qquad p^\star_h := \min_{k \in I^\star_h} \Pr_{x \sim \mathcal{D}}[x = u_k]. \tag{22}$$

Algorithm 6 Pure DP Generation without a Public Witness Bound $\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,h}$

Require: Dataset $S = (x_1, \ldots, x_n) \in \mathcal{U}^n$, privacy parameter $\varepsilon > 0$, functions $f, g, h : \mathbb{N} \to \mathbb{N}$, language collection $\mathcal{C} = \{L_i\}_{i \in \mathbb{N}}$.
Ensure: A string $\widehat{x} \in \mathcal{U}$.
1: **for** $k \in [h(n)]$ **do**
2: $\quad N_S(u_k) \leftarrow \sum_{t=1}^{n} \mathbf{1}\{x_t = u_k\}$
3: **end for**
4: **for** $i \in [f(n)], t \in [h(n)]$ **do**
5: $\quad$ Compute $I_i(t)$, $a_{i,t}(S)$, $d_{i,t}(S)$, $q_{\mathrm{pair}}(S, (i,t))$ via (18)–(21)
6: **end for**
7: Sample $\left(\widehat{i}, \widehat{t}\right) \sim \mathrm{EM}_{\varepsilon}\big(S,\, q_{\mathrm{pair}},\, [f(n)] \times [h(n)]\big)$ with $\Delta q = h(n)/g(n)$ ▷ Privately select a (language, threshold) pair
8: Sample $\widehat{j} \sim \mathrm{Unif}([2^n])$ ▷ Public randomness
9: **return** the $\widehat{j}$-th smallest-indexed string in $L_{\widehat{i}} \cap \{u_{\widehat{t}+1}, u_{\widehat{t}+2}, \ldots\}$

As for $p^\star_W$, we set $p^\star_h := 1$ if $I^\star_h = \emptyset$. Since $L_{i^\star} \subseteq \mathrm{supp}(\mathcal{D})$, every string u_k with $k \in I^\star_h$ has positive probability under $\mathcal{D}$, so $p^\star_h \in (0, 1]$. Note that $p^\star_h$ depends on $h(n)$ and may decrease as $h(n)$ grows, since the minimum is taken over a larger set. This creates a tension: larger $h(n)$ provides a wider search range but requires a smaller $g(n)$ for coverage, which in turn weakens the score gap exploited by the exponential mechanism.

Compared to Algorithm 3, the key difference is that the exponential mechanism now searches over *pairs* (i, t) rather than language indices i alone. The selected threshold $\widehat{t}$ replaces the role of the public witness bound W: the output string is drawn from the selected language beyond index $\widehat{t}$. Since both $\widehat{i}$ and $\widehat{t}$ are part of the exponential mechanism's output, the subsequent output step (lines 8–9) is post-processing and incurs no additional privacy cost.

D.2.2 Privacy Analysis

Lemma D.14 (Sensitivity of the pair score)**.** *The score function q_{pair} defined in (21) has sensitivity $h(n)/g(n)$.*

Proof. Fix $(i, t) \in [f(n)] \times [h(n)]$ and let $S, S' \in \mathcal{U}^n$ be neighboring datasets. By the same argument as in Lemma D.5, the minimum prefix count satisfies $|a_{i,t}(S) - a_{i,t}(S')| \le 1$ (since each multiplicity changes by at most 1, and the minimum preserves Lipschitz constants; or $a_{i,t}(S) = a_{i,t}(S') = n$ when $I_i(t) = \emptyset$). Since $x \mapsto (g(n) - x)_+$ is 1-Lipschitz, $|d_{i,t}(S) - d_{i,t}(S')| \le 1$. Therefore

$$|q_{\mathrm{pair}}(S, (i,t)) - q_{\mathrm{pair}}(S', (i,t))| = \frac{h(n)}{g(n)} \, |d_{i,t}(S) - d_{i,t}(S')| \le \frac{h(n)}{g(n)}.$$

□

Remark D.15 (Sensitivity and Score gap without a Public Bound)**.** The pair score has sensitivity $h(n)/g(n)$, and on the coverage event its margin is at least $h(n) - i(\mathcal{C}, \mathcal{D}) + 1 > h(n)/2$ when $h(n) \ge 2\, i(\mathcal{C}, \mathcal{D})$ (Lemma D.18). So the ratio of margin to sensitivity is at least $g(n)/2$, as in the public-bound case up to a factor 2. The cost of not knowing the witness bound lies elsewhere: in the larger range $[f(n)] \times [h(n)]$, and in the coverage term, which depends on $p^\star_h$ and so on $h(n)$ (Remark D.21).

Lemma D.16 (Privacy)**.** *Algorithm 6 is ε-differentially private.*

Proof. The pair selection on line 7 applies the exponential mechanism with score q_{pair} over the finite range $[f(n)] \times [h(n)]$. By Lemma D.14, q_{pair} has sensitivity $\Delta = h(n)/g(n)$, so this step is ε-DP. The output step (lines 8–9) is a deterministic function of the private output $(\widehat{i}, \widehat{t})$ and public randomness $\widehat{j}$, and is therefore ε-DP by post-processing (Lemma 2.2). □

D.2.3 Utility Analysis

The utility analysis proceeds in three steps, paralleling the public-bound case: (i) a coverage event $\mathcal{E}_h$ under which every relevant string of $L_{i^\star}$ up to index $h(n)$ appears at least $g(n)$ times (Lemma D.17); (ii) on $\mathcal{E}_h$, the good pair $(i^\star, h(n))$ achieves the maximum score while every bad pair has a much lower score (Lemma D.18); (iii) the exponential mechanism exploits this gap to select a good language (Lemma D.19).

Lemma D.17 (Coverage of the Good Prefix)**.** *Define the event*

$$\mathcal{E}_h := \big\{N_S(u_k) \geq g(n) \quad \forall\, k \in I_h^\star\big\}.$$

If $g(n) \leq np_h^\star/2$ for all large enough n, then

$$\Pr_{S\sim\mathcal{D}^n}[\mathcal{E}_h^c] \leq |I_h^\star| \exp\Big(-\frac{np_h^\star}{8}\Big).$$

Proof. The proof is identical to that of Lemma D.8, with W replaced by $h(n)$ and $p_W^\star$ replaced by $p_h^\star$. □

Lemma D.18 (Score Separation on $\mathcal{E}_h$)**.** *Suppose $h(n) \geq i(\mathcal{C},\mathcal{D})$. On the event $\mathcal{E}_h$:*

(a) $q_{\mathrm{pair}}(S,(i^\star,h(n))) = h(n)$.

(b) For every $(i,t) \in [f(n)] \times [h(n)]$ with $L_i \not\subseteq \mathrm{supp}(\mathcal{D})$: $q_{\mathrm{pair}}(S,(i,t)) \leq i(\mathcal{C},\mathcal{D}) - 1$.

Consequently, the largest score is $h(n)$, and every bad pair trails it by at least $h(n) - i(\mathcal{C},\mathcal{D}) + 1$.

Proof. Assume $\mathcal{E}_h$ holds throughout.

Part (a): The good pair achieves score $h(n)$. On $\mathcal{E}_h$, every $k \in I_h^\star$ satisfies $N_S(u_k) \geq g(n)$. If $I_{i^\star}(h(n)) \neq \emptyset$, then $a_{i^\star,h(n)}(S) = \min_{k\in I_{i^\star}(h(n))} N_S(u_k) \geq g(n)$, so $d_{i^\star,h(n)}(S) = 0$. If $I_{i^\star}(h(n)) = \emptyset$, then $a_{i^\star,h(n)}(S) = n \geq g(n)$, so $d_{i^\star,h(n)}(S) = 0$ as well. In either case,

$$q_{\mathrm{pair}}(S,(i^\star,h(n))) = h(n) - \frac{h(n)}{g(n)} \cdot 0 = h(n).$$

Since $q_{\mathrm{pair}}(S,(i,t)) \leq t \leq h(n)$ for all pairs (as the deficit is nonnegative), the largest score is $h(n)$.

Part (b): Every bad pair has low score. Fix $i \in [f(n)]$ with $L_i \not\subseteq \mathrm{supp}(\mathcal{D})$ and any $t \in [h(n)]$. We consider two cases depending on whether t is large enough to contain a witness.

Case $t \geq i(\mathcal{C},\mathcal{D})$: By Definition D.1, there exists $k \leq i(\mathcal{C},\mathcal{D}) \leq t$ such that $u_k \in L_i \setminus \mathrm{supp}(\mathcal{D})$. In particular, $k \in I_i(t)$ and u_k never appears in any sample from $\mathcal{D}$, so $N_S(u_k) = 0$. Therefore $a_{i,t}(S) \leq N_S(u_k) = 0$, giving $d_{i,t}(S) = g(n)$ and

$$q_{\mathrm{pair}}(S,(i,t)) = t - \frac{h(n)}{g(n)} \cdot g(n) = t - h(n) \leq 0.$$

Case $t < i(\mathcal{C},\mathcal{D})$: Regardless of the deficit, the score satisfies

$$q_{\mathrm{pair}}(S,(i,t)) \leq t \leq i(\mathcal{C},\mathcal{D}) - 1.$$

Combining both cases, every bad pair has score at most $\max\{0,\, i(\mathcal{C},\mathcal{D}) - 1\} = i(\mathcal{C},\mathcal{D}) - 1$ (since $i(\mathcal{C},\mathcal{D}) \geq 1$). The score gap between the good pair and any bad pair is at least $h(n) - (i(\mathcal{C},\mathcal{D}) - 1) = h(n) - i(\mathcal{C},\mathcal{D}) + 1$. □

Lemma D.19 (Private Selection and Output)**.** *Suppose $h(n) \geq 2\, i(\mathcal{C},\mathcal{D})$. Define*

$$\beta_{\mathrm{sel}} := f(n)\, h(n) \exp\Big(-\frac{\varepsilon\, g(n)}{4}\Big), \qquad \beta_{\mathrm{out}} := \exp\Big(-\frac{n}{2}\Big).$$

On the event $\mathcal{E}_h$:

(a) The exponential mechanism selects a good language (i.e., $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$) with probability at least $1 - \beta_{\mathrm{sel}}$.

(b) Conditioned on $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$, the output $\widehat{x}$ satisfies $\widehat{x} \in \mathrm{supp}(\mathcal{D}) \setminus S$ with probability at least $1 - \beta_{\mathrm{out}}$.

Proof. **Part (a): Pair selection.** Fix a sample S in $\mathcal{E}_h$, and let $\mathcal{G}$ be the set of pairs $(i,t) \in [f(n)] \times [h(n)]$ with $L_i \subseteq \mathrm{supp}(\mathcal{D})$; it contains $(i^\star, h(n))$. By Lemma D.18, $q_{\mathrm{pair}}(S,(i^\star,h(n))) = h(n)$ and every pair outside $\mathcal{G}$ scores at most $i(\mathcal{C},\mathcal{D}) - 1 \le h(n)/2 - 1$, so $\mathcal{G}$ has margin greater than $h(n)/2$. The score has sensitivity $\Delta = h(n)/g(n)$ (Lemma D.14), so Lemma A.5(a) bounds the probability of selecting a pair outside $\mathcal{G}$ by $f(n)h(n)\exp(-\varepsilon(h(n)/2)\,g(n)/(2h(n))) = \beta_{\mathrm{sel}}$.

Part (b): Output novelty. Suppose $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$. The algorithm outputs the $\widehat{j}$-th smallest-indexed string from $T_{\widehat{i},\widehat{t}} := L_{\widehat{i}} \cap \{u_{\widehat{t}+1}, u_{\widehat{t}+2}, \ldots\}$, where $\widehat{j} \sim \mathrm{Unif}([2^n])$. Since $L_{\widehat{i}}$ is infinite, $|T_{\widehat{i},\widehat{t}}| = \infty \ge 2^n$. Since $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$, every string in $T_{\widehat{i},\widehat{t}}$ belongs to $\mathrm{supp}(\mathcal{D})$. At most n strings from $T_{\widehat{i},\widehat{t}}$ can appear in S, so

$$\Pr[\widehat{x} \in S \mid L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})] \le \frac{n}{2^n} \le \exp\Big(-\frac{n}{2}\Big)$$

for $n \ge 14$ (Lemma A.6(c)). □

D.2.4 Putting Things Together

Theorem D.20 (Guarantees of Algorithm 6, Restatement of Theorem 4.5)**.** *Let $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$ be a countable collection of infinite languages and let $\mathcal{D}$ be any distribution over $\mathcal{U}$. Suppose Assumptions D.2 and D.3 hold. Let $\varepsilon > 0$, and let $f, g, h : \mathbb{N} \to \mathbb{N}$ satisfy:*

(i) $f(n) \ge i^\star$ and $h(n) \ge 2\,i(\mathcal{C},\mathcal{D})$ for all large enough n;

(ii) $g(n) \to \infty$ and $g(n) \le np_h^\star/2$ for all large enough n.

Then for all sufficiently large n, Algorithm 6 has the following guarantees:

- ***Privacy.*** *Algorithm 6 is ε-differentially private.*
- ***Utility.*** *The generation error satisfies*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,h}, \mathcal{D}, \mathcal{C}, n) \le |I_h^\star| \exp\Big(-\frac{np_h^\star}{8}\Big) + f(n)\,h(n)\exp\Big(-\frac{\varepsilon\,g(n)}{4}\Big) + \exp\Big(-\frac{n}{2}\Big). \tag{23}$$

Proof. **Privacy.** This directly follows from Lemma D.16.

Utility. By a union bound over the three failure events:

$$\mathrm{GenErr} \le \Pr[\mathcal{E}_h^c] + \Pr[L_{\widehat{i}} \not\subseteq \mathrm{supp}(\mathcal{D}) \mid \mathcal{E}_h] + \Pr[\widehat{x} \in S \mid L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})].$$

Applying Lemma D.17 (coverage failure), Lemma D.19(a) (selection failure), and Lemma D.19(b) (collision):

$$\mathrm{GenErr} \le |I_h^\star| \exp\Big(-\frac{np_h^\star}{8}\Big) + f(n)\,h(n)\exp\Big(-\frac{\varepsilon\,g(n)}{4}\Big) + \exp\Big(-\frac{n}{2}\Big).$$

□

Remark D.21 (No constant mass floor without a public window)**.** The coverage condition $g(n) \le np_h^\star/2$ of Theorem D.20 cannot be met with $g(n) = \Theta(n)$ in general. Since $h(n) \to \infty$ and $L_{i^\star}$ is infinite, $I_h^\star$ eventually contains strings of arbitrarily small probability, so $p_h^\star \to 0$ and no constant $p_0 \le p_h^\star$ can persist. Our analysis therefore gives fully exponential rates with a mass floor only for a fixed public window (Corollary D.13).

Corollary D.22 (No public information, Restatement of Corollary 4.7)**.** *Under the conditions of Theorem D.20 with $\varepsilon > 0$ fixed, let $r(n) \to \infty$ with $r(n) = o(n)$, set $g(n) = \lfloor r(n) \rfloor$, and choose f, h with $\log(f(n)\,h(n)) = o(r(n))$. If $r(n) \le np_h^\star/2$ for all large enough n, then*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,f,g,h}, \mathcal{D}, \mathcal{C}, n) = \exp\big(-\Omega(\min\{1,\varepsilon\}\cdot r(n))\big).$$

Proof. The hypothesis gives $g(n) \le r(n) \le np_h^\star/2$, so Theorem D.20 applies. The coverage term satisfies $|I_h^\star| \exp(-np_h^\star/8) \le h(n)\exp(-r(n)/4)$ since $np_h^\star \ge 2r(n)$ and $|I_h^\star| \le h(n)$. The privacy term is $f(n)\,h(n)\exp(-\varepsilon\lfloor r(n)\rfloor/4)$. Since $\log(f(n)\,h(n)) = o(r(n))$, both prefactors are absorbed, giving $\exp(-\Omega(r(n)))$ and $\exp(-\Omega(\varepsilon\, r(n)))$; the collision term is $\exp(-n/2)$. Hence the error is $\exp(-\Omega(\min\{1,\varepsilon\}\cdot r(n)))$. □

D.3 Proof of Theorem 4.8(a) (Approximate DP Generation with a Public Witness Bound)

Algorithm 7 Approximate DP Generation with a Public Witness Bound $\mathcal{A}^{\mathrm{gen}}_{\varepsilon,\delta,f,g,W}$

Require: Dataset $S = (x_1,\dots,x_n) \in \mathcal{U}^n$, privacy parameters $\varepsilon, \delta \in (0,1)$, a public witness bound W, functions $f, g : \mathbb{N} \to \mathbb{N}$, language collection $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$.
Ensure: A string $\widehat{x} \in \mathcal{U}$.

1: Compute $q^W(S,i)$ for every $i \in [f(n)]$ via (12)–(15)
2: $\sigma \leftarrow \frac{\sqrt{f(n)}}{\varepsilon}\sqrt{2\log(1.25/\delta)}$
3: Sample $Z_i \sim \mathcal{N}(0,\sigma^2)$ independently for $i \in [f(n)]$
4: $\widehat{i} \leftarrow \arg\max_{i\in[f(n)]}\big(q^W(S,i) + Z_i\big)$
5: Sample $\widehat{j} \sim \mathrm{Unif}([2^n])$ ▷ Public randomness
6: **return** the $\widehat{j}$-th smallest-indexed string in $L_{\widehat{i}} \cap \{u_{W+1}, u_{W+2}, \dots\}$

Lemma D.23 (Privacy)**.** *Algorithm 7 is (ε,δ)-differentially private.*

Proof. By Lemma D.5, each coordinate of $v(S) := (q^W(S,i))_{i\in[f(n)]}$ changes by at most 1 between neighboring datasets, so v has ℓ_2-sensitivity at most $\sqrt{f(n)}$. The Gaussian mechanism (Lemma 2.4) with the stated σ makes the noisy vector (ε,δ)-DP, and the remaining steps are post-processing (Lemma 2.2). □

Theorem D.24 (Guarantees of Algorithm 7, Restatement of Theorem 4.8(a))**.** *Under the conditions of Theorem D.11, let $\varepsilon,\delta \in (0,1)$. Then Algorithm 7 is (ε,δ)-differentially private and, for all sufficiently large n,*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,\delta,f,g,W}, \mathcal{D}, \mathcal{C}, n) \le |I_W^\star| \exp\Big(-\frac{np_W^\star}{8}\Big) + 2f(n)\exp\Big(-\frac{\varepsilon^2 g(n)^2}{16 f(n)\log(1.25/\delta)}\Big) + \exp\Big(-\frac{n}{2}\Big). \tag{24}$$

Proof. Privacy is Lemma D.23. Let $\mathcal{G}_W := \{|Z_i| < g(n)/2 \;\forall\, i \in [f(n)]\}$. By Lemma A.3 and a union bound,

$$\Pr[\mathcal{G}_W^c] \le 2f(n)\exp\Big(-\frac{g(n)^2}{8\sigma^2}\Big) = 2f(n)\exp\Big(-\frac{\varepsilon^2 g(n)^2}{16 f(n)\log(1.25/\delta)}\Big).$$

On $\mathcal{E}_W \cap \mathcal{G}_W$, Lemma D.9 gives $q^W(S,i^\star) + Z_{i^\star} > -g(n)/2$, while every i with $L_i \not\subseteq \mathrm{supp}(\mathcal{D})$ has $q^W(S,i) + Z_i < -g(n) + g(n)/2 = -g(n)/2$. Hence the argmax is a language with $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$. The output step is the same as in Algorithm 3, so Lemma D.10(b) bounds the collision probability by $\exp(-n/2)$. A union bound over $\mathcal{E}_W^c$ (Lemma D.8), $\mathcal{G}_W^c$, and the collision event gives (24). □

Corollary D.25 (Privacy is Essentially Free with a Public Window)**.** *Under the conditions of Theorem D.24, let $\varepsilon \in (0,1)$ be constant and $g(n) = \lfloor n/\log n \rfloor$. If $f(n)\log(1.25/\delta) = o(n/\log^2 n)$, then $\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,\delta,f,g,W}, \mathcal{D}, \mathcal{C}, n) = \exp(-\Omega(n))$. For example, this holds for $\delta \ge n^{-O(1)}$ and $f(n) = o(n/\log^3 n)$.*

Proof. Since W is fixed, $p_W^\star > 0$ does not depend on n, so $g(n) \le n/\log n \le np_W^\star/2$ once $\log n \ge 2/p_W^\star$, and the coverage term is $\exp(-\Omega(n))$; no knowledge of $p_W^\star$ is needed. Using $g(n) \ge n/(2\log n)$ for $n \ge 3$, the exponent of the noise term is at least

$$\frac{\varepsilon^2 n^2}{64\log^2 n \cdot f(n)\log(1.25/\delta)} = n \cdot \frac{\varepsilon^2}{64} \cdot \frac{n}{\log^2 n \cdot f(n)\log(1.25/\delta)} = \omega(n),$$

and the prefactor $2f(n) \le 2n$ is absorbed. The collision term is $\exp(-n/2)$, so the error is $\exp(-\Omega(n))$. □

D.4 Proof of Theorem 4.8(b) (Approximate DP Generation without a Public Witness Bound)

Algorithm 8 Approximate DP Generation without a Public Witness Bound $\mathcal{A}^{\text{gen}}_{\varepsilon,\delta,f,g,h}$

Require: Dataset $S = (x_1, \dots, x_n) \in \mathcal{U}^n$, privacy parameters $\varepsilon, \delta \in (0,1)$, functions $f, g, h : \mathbb{N} \to \mathbb{N}$, language collection $\mathcal{C} = \{L_i\}_{i \in \mathbb{N}}$.
Ensure: A string $\widehat{x} \in \mathcal{U}$.

for $k \in [h(n)]$ **do**
  $N_S(u_k) \leftarrow \sum_{t=1}^{n} \mathbf{1}\{x_t = u_k\}$
end for
for $i \in [f(n)], t \in [h(n)]$ **do**
  Compute $I_i(t)$, $a_{i,t}(S)$, $d_{i,t}(S)$, $q_{\text{pair}}(S,(i,t))$ via (18)–(21)
end for
$\sigma \leftarrow \frac{h(n)}{g(n)} \cdot \frac{\sqrt{f(n)\cdot h(n)}}{\varepsilon} \sqrt{2\log(1.25/\delta)}$
for $(i,t) \in [f(n)] \times [h(n)]$ **do**
  Sample $Z_{i,t} \sim \mathcal{N}(0, \sigma^2)$
  $\widetilde{q}_{\text{pair}}(S,(i,t)) \leftarrow q_{\text{pair}}(S,(i,t)) + Z_{i,t}$
end for
$(\widehat{i}, \widehat{t}) \leftarrow \arg\max_{(i,t)\in[f(n)]\times[h(n)]} \widetilde{q}_{\text{pair}}(S,(i,t))$
Sample $\widehat{j} \sim \text{Unif}([2^n])$ ▷ Public randomness
return the $\widehat{j}$-th smallest-indexed string in $L_{\widehat{i}} \cap \{u_{\widehat{t}+1}, u_{\widehat{t}+2}, \dots\}$

D.4.1 Privacy Analysis.

Lemma D.26 (ℓ_2-Sensitivity of q_{pair})**.** *Define* $v \colon \mathcal{U}^n \to \mathbb{R}^{f(n)\cdot h(n)}$ *by*

$$v(S) := (q_{\text{pair}}(S,(i,t)))_{(i,t)\in[f(n)]\times[h(n)]}.$$

Then v *has* ℓ_2*-sensitivity* $\frac{h(n)}{g(n)}\sqrt{f(n)\cdot h(n)}$.

Proof. By Lemma D.14, each coordinate satisfies $|q_{\text{pair}}(S,(i,t)) - q_{\text{pair}}(S',(i,t))| \le h(n)/g(n)$ for neighboring S, S'. Therefore

$$\|v(S) - v(S')\|_2^2 = \sum_{(i,t)} |q_{\text{pair}}(S,(i,t)) - q_{\text{pair}}(S',(i,t))|^2 \le f(n)\cdot h(n)\cdot \frac{h(n)^2}{g(n)^2}.$$

Taking the square root gives $\|v(S) - v(S')\|_2 \le \frac{h(n)}{g(n)}\sqrt{f(n)\cdot h(n)}$. □

Lemma D.27 (Privacy)**.** *For* $\varepsilon, \delta \in (0,1)$*, Algorithm 8 is* (ε,δ)*-differentially private.*

Proof. By Lemma D.26, the pair score vector has ℓ_2-sensitivity $\Delta_2 = \frac{h(n)}{g(n)}\sqrt{f(n)\cdot h(n)}$. The Gaussian mechanism with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{2\log(1.25/\delta)}$ ensures the noisy score vector is (ε,δ)-DP. The subsequent steps are post-processing and therefore (ε,δ)-DP by Lemma 2.2. □

D.4.2 Utility Analysis.

The utility analysis proceeds as in the pure DP case: (i) coverage (Lemma D.17, unchanged); (ii) noise concentration (Lemma D.28); (iii) deterministic correctness on the intersection of both events (Lemma D.29).

Lemma D.28 (Noise Concentration)**.** *Suppose* $h(n) \ge 2\, i(\mathcal{C},\mathcal{D})$*. Define the event*

$$\mathcal{G}_h := \left\{ |Z_{i,t}| < \frac{h(n)}{4} \quad \forall (i,t) \in [f(n)] \times [h(n)] \right\}.$$

Then

$$\Pr[\mathcal{G}_h^c] \le 2f(n)\,h(n)\exp\Big(-\frac{\varepsilon^2\,g(n)^2}{64\,f(n)\,h(n)\log(1.25/\delta)}\Big).$$

Proof. Each $Z_{i,t} \sim \mathcal{N}(0,\sigma^2)$ with

$$\sigma^2 = \frac{2h(n)^2\,f(n)\,h(n)\log(1.25/\delta)}{\varepsilon^2 g(n)^2} = \frac{2h(n)^3 f(n)\log(1.25/\delta)}{\varepsilon^2 g(n)^2}.$$

By the Gaussian tail bound (Lemma A.3),

$$\Pr\Big[|Z_{i,t}| \ge \frac{h(n)}{4}\Big] \le 2\exp\Big(-\frac{h(n)^2/16}{2\sigma^2}\Big) = 2\exp\Big(-\frac{\varepsilon^2 g(n)^2}{64\,f(n)\,h(n)\log(1.25/\delta)}\Big).$$

A union bound over $f(n)\cdot h(n)$ pairs gives the result. □

Lemma D.29 (Deterministic Correctness on $\mathcal{E}_h \cap \mathcal{G}_h$)**.** *Suppose $h(n) \ge 2\,i(\mathcal{C},\mathcal{D})$. On the event $\mathcal{E}_h \cap \mathcal{G}_h$, Algorithm 8 selects a good language, i.e., $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$.*

Proof. Assume $\mathcal{E}_h \cap \mathcal{G}_h$ holds. By Lemma D.18, $q_{\mathrm{pair}}(S,(i^\star,h(n))) = h(n)$ and every bad pair (i,t) satisfies $q_{\mathrm{pair}}(S,(i,t)) \le i(\mathcal{C},\mathcal{D}) - 1 \le h(n)/2 - 1$.

For the good pair $(i^\star,h(n))$:

$$\widetilde{q}_{\mathrm{pair}}(S,(i^\star,h(n))) = h(n) + Z_{i^\star,h(n)} > h(n) - \frac{h(n)}{4} = \frac{3h(n)}{4}.$$

For any bad pair (i,t):

$$\widetilde{q}_{\mathrm{pair}}(S,(i,t)) < \frac{h(n)}{2} - 1 + \frac{h(n)}{4} = \frac{3h(n)}{4} - 1.$$

Hence the good pair beats every bad pair, and the argmax selects a good pair. (This is Lemma A.5(c) with margin greater than $h(n)/2$.) □

D.4.3 Putting Things Together.

Theorem D.30 (Guarantees of Algorithm 8, Restatement of Theorem 4.8(b))**.** *Let $\mathcal{C} = \{L_i\}_{i\in\mathbb{N}}$ be a countable collection of infinite languages and let $\mathcal{D}$ be any distribution over $\mathcal{U}$. Suppose Assumptions D.2 and D.3 hold. Let $\varepsilon,\delta \in (0,1)$, and let $f,g,h : \mathbb{N}\to\mathbb{N}$ satisfy:*

(i) $f(n) \ge i^\star$ and $h(n) \ge 2\,i(\mathcal{C},\mathcal{D})$ for all large enough n;

(ii) $g(n)\to\infty$ and $g(n) \le np_h^\star/2$ for all large enough n.

Then for all sufficiently large n, Algorithm 8 has the following guarantees:

- ***Privacy.*** *Algorithm 8 is (ε,δ)-differentially private.*
- ***Utility.*** *The generation error satisfies*

$$\mathrm{GenErr}(\mathcal{A}^{\mathrm{gen}}_{\varepsilon,\delta,f,g,h},\mathcal{D},\mathcal{C},n) \le |I_h^\star|\exp\Big(-\frac{np_h^\star}{8}\Big) + 2f(n)h(n)\exp\Big(-\frac{\varepsilon^2 g(n)^2}{64f(n)h(n)\log(1.25/\delta)}\Big) + \exp\Big(-\frac{n}{2}\Big). \tag{25}$$

Proof. **Privacy.** This follows from Lemma D.27.

Utility. By Lemma D.29, $L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})$ whenever $\mathcal{E}_h \cap \mathcal{G}_h$ holds. By a union bound:

$$\mathrm{GenErr} \le \Pr[\mathcal{E}_h^c] + \Pr[\mathcal{G}_h^c] + \Pr[\widehat{x}\in S \mid L_{\widehat{i}} \subseteq \mathrm{supp}(\mathcal{D})].$$

Applying Lemma D.17, Lemma D.28, and the collision bound $n/2^n \le \exp(-n/2)$:

$$\mathrm{GenErr} \le |I_h^\star|\exp\Big(-\frac{np_h^\star}{8}\Big) + 2f(n)\,h(n)\exp\Big(-\frac{\varepsilon^2 g(n)^2}{64\,f(n)\,h(n)\log(1.25/\delta)}\Big) + \exp\Big(-\frac{n}{2}\Big).$$

□

Remark D.31 (Scope without a public bound)**.** Theorem D.30 is conditional on $g(n) \le np_h^\star/2$ and keeps the coverage term $|I_h^\star| \exp(-np_h^\star/8)$, which is not $\exp(-\Omega(n))$ in general because $p_h^\star \to 0$ as $h(n) \to \infty$ (Remark D.21). The fully exponential approximate DP rate is therefore stated only for a fixed public window (Corollary D.25). There, the score has sensitivity 1 per coordinate, so the noise $\sigma = O(\sqrt{f(n)\log(1.25/\delta)}/\varepsilon)$ is negligible relative to the score gap $g(n) = \lfloor n/\log n \rfloor$.

E Proofs for Section 5 (Lower Bounds)

The proofs of Theorems 5.3 and 5.5 in Section 5 are complete once the properties of the two hard instances are checked. This appendix checks them and compares the bounds with the non-private ones. Both proofs apply the private two-point inequality (Lemma 5.1) to a coordinatewise coupling in which the two samples share the string s_0 with probability $3/4$.

E.1 The Identification Instance (Theorem 5.3)

E.1.1 Hard Instance Construction

Definition E.1 (Private Element)**.** Let $\mathcal{C}$ be a collection of languages over $\mathcal{U}$. For a language $L \in \mathcal{C}$, an element $x \in \mathcal{U}$ is *private* to L with respect to $\mathcal{C}$ if

$$x \in L \setminus \Big(\bigcup_{\widetilde{L} \in \mathcal{C} \setminus \{L\}} \widetilde{L} \Big).$$

Definition E.2 (Intersecting Private-Pair (IPP) Condition)**.** A collection $\mathcal{C}$ of languages satisfies the *Intersecting Private-Pair (IPP) Condition* if there exist two distinct languages $L \neq L'$ in $\mathcal{C}$ such that both have at least one private element with respect to $\mathcal{C}$ and $L \cap L' \neq \emptyset$.

Definition E.3 (Hard Instance)**.** Assume $\mathcal{C}$ satisfies IPP. Let $L, L' \in \mathcal{C}$ be two languages witnessing IPP. Fix elements

$$s_0 \in L \cap L', \qquad s_1 \in L \setminus \Big(\bigcup_{\widetilde{L} \in \mathcal{C} \setminus \{L\}} \widetilde{L} \Big), \qquad s_2 \in L' \setminus \Big(\bigcup_{\widetilde{L} \in \mathcal{C} \setminus \{L'\}} \widetilde{L} \Big).$$

Note that s_0, s_1, s_2 are necessarily distinct (e.g., if $s_1 = s_0$ then $s_1 \in L \cap L' \subseteq L'$, contradicting that s_1 is private to L; similarly $s_2 \neq s_0$, and $s_1 \neq s_2$ since $s_1 \notin L'$ while $s_2 \in L'$). Define two distributions $\mathcal{D}, \mathcal{D}'$ over $\mathcal{U}$:

$$\mathcal{D}: \ \Pr_{x \sim \mathcal{D}}[x = s_0] = \tfrac{3}{4}, \quad \Pr_{x \sim \mathcal{D}}[x = s_1] = \tfrac{1}{4}, \quad \Pr_{x \sim \mathcal{D}}[x = s] = 0 \ \ \forall\, s \notin \{s_0, s_1\},$$
$$\mathcal{D}': \ \Pr_{x \sim \mathcal{D}'}[x = s_0] = \tfrac{3}{4}, \quad \Pr_{x \sim \mathcal{D}'}[x = s_2] = \tfrac{1}{4}, \quad \Pr_{x \sim \mathcal{D}'}[x = s] = 0 \ \ \forall\, s \notin \{s_0, s_2\}.$$

E.1.2 Properties of the Hard Instance

Lemma E.4 (Optimality and Gap)**.** *Consider the hard instance in Definition E.3. Then:*

(a) $\mathrm{err}_{\mathcal{D}}(L) = \inf_{\widetilde{L} \in \mathcal{C}} \mathrm{err}_{\mathcal{D}}(\widetilde{L}) = 0$ *and* $\inf_{\widetilde{L} \in \mathcal{C} \setminus \{L\}} \mathrm{err}_{\mathcal{D}}(\widetilde{L}) \ge 1/4$.

(b) $\mathrm{err}_{\mathcal{D}'}(L') = \inf_{\widetilde{L} \in \mathcal{C}} \mathrm{err}_{\mathcal{D}'}(\widetilde{L}) = 0$ *and* $\inf_{\widetilde{L} \in \mathcal{C} \setminus \{L'\}} \mathrm{err}_{\mathcal{D}'}(\widetilde{L}) \ge 1/4$.

Proof. We prove part (a), and part (b) follows by the symmetric argument, swapping $(L, s_1, \mathcal{D})$ with $(L', s_2, \mathcal{D}')$.

Since $s_0 \in L$ and $s_1 \in L$ by construction, the support of $\mathcal{D}$ satisfies $\mathrm{supp}(\mathcal{D}) = \{s_0, s_1\} \subseteq L$. It follows that every draw from $\mathcal{D}$ lands in L, so

$$\mathrm{err}_{\mathcal{D}}(L) = \Pr_{x \sim \mathcal{D}}[x \notin L] = 0.$$

Since $\mathrm{err}_{\mathcal{D}}(\widetilde{L}) \ge 0$ for every $\widetilde{L} \in \mathcal{C}$ and the value 0 is achieved by L, we conclude $\inf_{\widetilde{L} \in \mathcal{C}} \mathrm{err}_{\mathcal{D}}(\widetilde{L}) = 0$.

Now fix any $\widetilde{L} \in \mathcal{C} \setminus \{L\}$. By Definition E.1, s_1 is private to L with respect to $\mathcal{C}$, which means $s_1 \notin \widetilde{L}$ for any $\widetilde{L} \neq L$. Since $s_1 \notin \widetilde{L}$, any draw from $\mathcal{D}$ that equals s_1 is not covered by $\widetilde{L}$:

$$\mathrm{err}_{\mathcal{D}}(\widetilde{L}) = \Pr_{x\sim\mathcal{D}}[x \notin \widetilde{L}] \geq \Pr_{x\sim\mathcal{D}}[x = s_1] = \frac{1}{4}.$$

Since this holds for every $\widetilde{L} \in \mathcal{C} \setminus \{L\}$, the infimum over this set is at least $1/4$. □

Lemma E.5 (From Misidentification to Identification error)**.** *Consider the hard instance in Definition E.3. For any identification algorithm $\mathcal{A}$ (using randomness r),*

$$\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) \geq \frac{1}{4} \cdot \Pr_{S\sim\mathcal{D}^n, r}[\mathcal{A}(S) \neq L], \qquad \mathrm{IdErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n) \geq \frac{1}{4} \cdot \Pr_{S\sim(\mathcal{D}')^n, r}[\mathcal{A}(S) \neq L'].$$

Proof. We prove the first inequality; the second follows by the symmetric argument.

By Lemma E.4(a), the agnostic optimum under $\mathcal{D}$ equals 0 and is uniquely attained by L (in the sense that every other language incurs error at least $1/4$). Therefore the identification error simplifies to

$$\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) = \mathop{\mathbb{E}}_{S\sim\mathcal{D}^n, r}\big[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S))\big] - 0 = \mathop{\mathbb{E}}_{S\sim\mathcal{D}^n, r}\big[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S))\big].$$

We now bound the integrand pointwise by case analysis on the output of $\mathcal{A}$. If $\mathcal{A}(S) = L$, then $\mathrm{supp}(\mathcal{D}) \subseteq L$ implies $\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S)) = \mathrm{err}_{\mathcal{D}}(L) = 0$. If $\mathcal{A}(S) = \widetilde{L} \neq L$, then $\widetilde{L} \in \mathcal{C} \setminus \{L\}$, so $s_1 \notin \widetilde{L}$ by privacy of s_1. Consequently,

$$\mathrm{err}_{\mathcal{D}}(\widetilde{L}) = \Pr_{x\sim\mathcal{D}}[x \notin \widetilde{L}] \geq \Pr_{x\sim\mathcal{D}}[x = s_1] = \frac{1}{4}.$$

Combining both cases, we obtain the pointwise bound

$$\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S)) \geq \frac{1}{4} \cdot \mathbf{1}\{\mathcal{A}(S) \neq L\}$$

for every realization of S and r. Taking expectations over (S, r) on both sides yields

$$\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) = \mathop{\mathbb{E}}_{S,r}\big[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S))\big] \geq \frac{1}{4} \cdot \mathop{\mathbb{E}}_{S,r}\big[\mathbf{1}\{\mathcal{A}(S) \neq L\}\big] = \frac{1}{4} \cdot \Pr_{S\sim\mathcal{D}^n, r}[\mathcal{A}(S) \neq L].$$

□

Lemma E.6 (The Identification Coupling)**.** *For each coordinate $t \in [n]$, sample (X_t, Y_t) independently with $\Pr[(X_t, Y_t) = (s_0, s_0)] = 3/4$ and $\Pr[(X_t, Y_t) = (s_1, s_2)] = 1/4$. Then $S = (X_1, \ldots, X_n) \sim \mathcal{D}^n$, $S' = (Y_1, \ldots, Y_n) \sim (\mathcal{D}')^n$, and $\mathrm{d_{Ham}}(S, S')$ has the binomial law with parameters n and $1/4$. Consequently, for every ε-DP identification algorithm $\mathcal{A}$ and every $n \in \mathbb{N}$,*

$$\Pr_{S,r}[\mathcal{A}(S) \neq L] + \Pr_{S',r}[\mathcal{A}(S') = L] \geq \Big(1 - \frac{\theta_\varepsilon}{4}\Big)^n.$$

Proof. The first coordinate X_t equals s_0 with probability $3/4$ and s_1 with probability $1/4$, which is the law $\mathcal{D}$; in the same way, Y_t has the law $\mathcal{D}'$. The two coordinates differ exactly in the second branch, because $s_1 \neq s_2$ (Definition E.3), and the branches are independent across t. The inequality is Lemma 5.1 with $\mathcal{F} = \{L\}$ and $p = 1/4$. □

Together with Lemma E.5, this proves Theorem 5.3: under $\mathcal{D}'$, the output L differs from L', so $\Pr[\mathcal{A}(S') = L] \leq \Pr[\mathcal{A}(S') \neq L'] \leq 4\ \mathrm{IdErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)$.

Remark E.7 (Comparison with the Non-Private Bound)**.** At $\varepsilon = \infty$, Theorem 5.3 gives $\max\{\mathrm{IdErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n), \mathrm{IdErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)\} \geq \frac{1}{8}(3/4)^n$ for every algorithm and every n. Høgsgaard and Pabbaraju [2026, Theorem 2.2] prove $\mathrm{IdErr} \geq e^{-5n}$ for infinitely many n, assuming only that two languages have private strings; our instance also uses a common string $s_0 \in L \cap L'$, which makes the two samples equal with probability $(3/4)^n$.

Remark E.8 (Scales of the Two Bounds)**.** The lower bound of Theorem 5.3 and the upper bound of Corollary C.9 both carry the factor $\min\{1, \varepsilon\}$, the former at the linear scale n and the latter at the sublinear scale $r(n)$. The remaining gap between $r(n) = o(n)$ and n is inherited from the non-private setting [Høgsgaard and Pabbaraju, 2026].

E.2 The Generation Instance (Theorem 5.5)

The construction differs from the identification instance in two ways. First, the hard distributions have *infinite* supports, since generation requires novel strings from an infinite support. Second, the argument is one-sided: under $\mathcal{D}$, the algorithm should output strings from $L \setminus L'$, while under $\mathcal{D}'$ such outputs are failures.

E.2.1 Hard Instance Construction

Definition E.9 (Intersecting Infinite-Difference Pair (IIDP))**.** A collection $\mathcal{C}$ of languages over a countable universe $\mathcal{U}$ satisfies the *Intersecting Infinite-Difference Pair (IIDP)* condition if there exist two distinct languages $L, L' \in \mathcal{C}$, an element $s_0 \in L \cap L'$, and two infinite sequences of distinct elements $(a_k)_{k\geq 1} \subseteq L \setminus L'$ and $(b_k)_{k\geq 1} \subseteq L' \setminus L$.

Remark E.10 (On the Structural Conditions)**.** The IIDP condition (Definition E.9) is a condition on one pair of languages: it asks for infinitely many elements of $L \setminus L'$ and of $L' \setminus L$, but not that these elements be private with respect to $\mathcal{C}$. So IIDP and IPP (Definition E.2) are incomparable. The infinite differences give the hard distributions the infinite supports that generation needs.

Definition E.11 (Hard Instance for Generation)**.** Assume $\mathcal{C}$ satisfies IIDP and witnessed by $(L, L', s_0, (a_k)_{k\geq 1}, (b_k)_{k\geq 1})$. Define distributions $\mathcal{D}$ and $\mathcal{D}'$ over $\mathcal{U}$ by

$$\mathcal{D}: \Pr_{x\sim\mathcal{D}}[x = s_0] = \tfrac{3}{4}, \quad \Pr_{x\sim\mathcal{D}}[x = a_k] = \tfrac{1}{4}\cdot 2^{-k} \quad (k \geq 1),$$
$$\mathcal{D}': \Pr_{x\sim\mathcal{D}'}[x = s_0] = \tfrac{3}{4}, \quad \Pr_{x\sim\mathcal{D}'}[x = b_k] = \tfrac{1}{4}\cdot 2^{-k} \quad (k \geq 1).$$

All other strings have zero probability under both distributions.

E.2.2 Properties of the Hard Instance

Lemma E.12 (Support structure)**.** *Consider the hard instance in Definition E.11. Then:*

(a) $\mathrm{supp}(\mathcal{D}) = \{s_0\} \cup \{a_k : k \geq 1\} \subseteq L$ *and* $\mathrm{supp}(\mathcal{D}') = \{s_0\} \cup \{b_k : k \geq 1\} \subseteq L'$.

(b) Both supports are infinite.

(c) The sets $\mathcal{F} := \{a_k : k \geq 1\}$ *and* $\mathcal{F}' := \{b_k : k \geq 1\}$ *satisfy* $\mathcal{F} \cap \mathrm{supp}(\mathcal{D}') = \emptyset$ *and* $\mathcal{F}' \cap \mathrm{supp}(\mathcal{D}) = \emptyset$.

Proof. **Part (a):** By construction, $s_0 \in L \cap L'$ and $a_k \in L \setminus L'$ for all $k \geq 1$, so $\mathrm{supp}(\mathcal{D}) = \{s_0\} \cup \{a_k : k \geq 1\} \subseteq L$. The argument for $\mathcal{D}'$ is symmetric.

Part (b): The sequences $(a_k)_{k\geq 1}$ and $(b_k)_{k\geq 1}$ are infinite by the IIDP condition, and each element receives positive probability.

Part (c): Since $a_k \in L \setminus L'$ for all k, we have $a_k \notin L'$, hence $a_k \notin \{s_0\} \cup \{b_j : j \geq 1\} = \mathrm{supp}(\mathcal{D}')$. Thus $\mathcal{F} \cap \mathrm{supp}(\mathcal{D}') = \emptyset$. The argument for $\mathcal{F}'$ is symmetric. □

Lemma E.13 (Disjointness of private elements and opposing support)**.** *Consider the hard instance in Definition E.11. Let* $\mathcal{F} = \{a_k : k \geq 1\}$*. For any generation algorithm* $\mathcal{A}$*,*

$$\{\mathcal{A}(S') \in \mathcal{F}\} \subseteq \{\mathcal{A}(S') \notin \mathrm{supp}(\mathcal{D}') \setminus S'\},$$

where $S' \sim (\mathcal{D}')^n$*. In other words, if* $\mathcal{A}(S')$ *outputs an element of* $\mathcal{F}$*, it necessarily fails under* $\mathcal{D}'$*. Consequently,*

$$\Pr[\mathcal{A}(S') \in \mathcal{F}] \leq \mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n).$$

Proof. By Lemma E.12(c), $\mathcal{F} \cap \mathrm{supp}(\mathcal{D}') = \emptyset$. If $\mathcal{A}(S') \in \mathcal{F}$, then $\mathcal{A}(S') \notin \mathrm{supp}(\mathcal{D}')$, which immediately implies $\mathcal{A}(S') \notin \mathrm{supp}(\mathcal{D}') \setminus S'$. Taking probabilities gives the claimed inequality. □

Lemma E.14 (Success under $\mathcal{D}$ requires outputting from $\mathcal{F}$)**.** *Consider the hard instance in Definition E.11. Let* $\mathcal{F} = \{a_k : k \geq 1\}$*. For any generation algorithm* $\mathcal{A}$ *and* $S \sim \mathcal{D}^n$*,*

$$\Pr[\mathcal{A}(S) \in \mathcal{F}] \geq 1 - \mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) - 4^{-n}.$$

Proof. By Lemma E.12(a), $\mathrm{supp}(\mathcal{D}) = \{s_0\} \cup \mathcal{F}$. Hence the success event decomposes as

$$\{\mathcal{A}(S) \in \mathrm{supp}(\mathcal{D}) \setminus S\} \subseteq \{\mathcal{A}(S) \in \mathcal{F}\} \cup \{\mathcal{A}(S) = s_0,\ s_0 \notin S\} \subseteq \{\mathcal{A}(S) \in \mathcal{F}\} \cup \{s_0 \notin S\}.$$

Indeed, if $\mathcal{A}(S) = s_0$, this output is successful only if $s_0 \notin S$. Taking probabilities and rearranging:

$$1 - \mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) = \Pr[\mathcal{A}(S) \in \mathrm{supp}(\mathcal{D}) \setminus S] \leq \Pr[\mathcal{A}(S) \in \mathcal{F}] + \Pr[s_0 \notin S].$$

Since $\Pr_{x\sim\mathcal{D}}[x = s_0] = 3/4$, the probability that s_0 never appears in n i.i.d. draws is $\Pr[s_0 \notin S] = (1 - 3/4)^n = 4^{-n}$. Rearranging gives the claim. □

Lemma E.15 (The Generation Coupling)**.** *For each coordinate $t \in [n]$, independently, set $(X_t, Y_t) = (s_0, s_0)$ with probability $3/4$; otherwise draw $K_t \in \mathbb{N}$ with $\Pr[K_t = k] = 2^{-k}$ and set $(X_t, Y_t) = (a_{K_t}, b_{K_t})$. Then $S = (X_1, \ldots, X_n) \sim \mathcal{D}^n$, $S' = (Y_1, \ldots, Y_n) \sim (\mathcal{D}')^n$, and $\mathrm{d}_{\mathrm{Ham}}(S, S')$ has the binomial law with parameters n and $1/4$. Consequently, for every ε-DP generation algorithm $\mathcal{A}$ and every $n \in \mathbb{N}$,*

$$\mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) + \mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n) \geq \left(1 - \frac{\theta_\varepsilon}{4}\right)^n - 4^{-n} \geq \frac{2}{3}\left(1 - \frac{\theta_\varepsilon}{4}\right)^n.$$

Proof. The first coordinate equals s_0 with probability $3/4$ and a_k with probability $2^{-k}/4$, which is the law $\mathcal{D}$; in the same way, the second has the law $\mathcal{D}'$. In the second branch, $a_{K_t} \notin L'$ while $b_{K_t} \in L'$, so the two coordinates differ there, and they agree in the first branch. By Lemma 5.1 with $p = 1/4$, $\Pr[\mathcal{A}(S) \notin \mathcal{F}] + \Pr[\mathcal{A}(S') \in \mathcal{F}] \geq (1 - \theta_\varepsilon/4)^n$. Lemma E.14 bounds the first probability by $\mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n) + 4^{-n}$, and Lemma E.13 bounds the second by $\mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)$. The last inequality holds because $(1 - \theta_\varepsilon/4)^n \geq (3/4)^n \geq 3 \cdot 4^{-n}$ for $n \geq 1$. □

Remark E.16 (Asymmetry with the Identification Lower Bound)**.** In the identification argument, the event $\mathcal{F} = \{L\}$ is a correct output under $\mathcal{D}$ and an error under $\mathcal{D}'$, and the two misidentification probabilities enter symmetrically. In the generation argument, $\Pr[\mathcal{A}(S) \in \mathcal{F}]$ is large because of the success requirement under $\mathcal{D}$ (Lemma E.14), not because of privacy, and every output in $\mathcal{F}$ fails under $\mathcal{D}'$ (Lemma E.13). This asymmetry reflects the fact that the generation objective depends on $\mathrm{supp}(\mathcal{D})$. At $\varepsilon = \infty$, Lemma E.15 gives $\max\{\mathrm{GenErr}(\mathcal{A}, \mathcal{D}, \mathcal{C}, n), \mathrm{GenErr}(\mathcal{A}, \mathcal{D}', \mathcal{C}, n)\} \geq \frac{1}{3}(3/4)^n$ for every n; Høgsgaard and Pabbaraju [2026, Theorem 3.5] prove $\frac{1}{4}e^{-2n}$ for infinitely many n under the different condition $|L \cap L'| < \infty$ and $\mathcal{U} \setminus (L \cup L') \neq \varnothing$, which our proof does not need.

Remark E.17 (The Hard Instance Has Public Side Information)**.** Let $\mathcal{C} = \{L, L'\}$ satisfy IIDP with $L \cap L' = \{s_0\}$, and let (a_k) and (b_k) enumerate $L \setminus L'$ and $L' \setminus L$ in the order of $\mathcal{U}$. Then $\mathrm{supp}(\mathcal{D}) = L$ and $\mathrm{supp}(\mathcal{D}') = L'$, so each of $\mathcal{D}$ and $\mathcal{D}'$ has a language of $\mathcal{C}$ equal to its support (Assumption 4.3), and the other language is ruled out by its first element outside that support (Assumption 4.2). Hence $W := \max\{i(\mathcal{C}, \mathcal{D}), i(\mathcal{C}, \mathcal{D}')\}$ is a valid public window for both, and it depends only on $\mathcal{C}$ and the enumeration. The strings of L among $u_1, \ldots, u_W$ are s_0 and the elements a_k with index at most W, so $p^\star_W \geq 2^{-W}/4$ under $\mathcal{D}$, and likewise under $\mathcal{D}'$; thus $p_0 := 2^{-W}/4$ is a valid public mass floor for both.

In the instance of Remark 5.6, with $u_k = k$: under $\mathcal{D}$, the language L' has the witness $u_4 = 4$, and the strings of L among $u_1, \ldots, u_4$ are 1 and 3, with masses $3/4$ and $1/8$; under $\mathcal{D}'$, the language L has the witness $u_3 = 3$, and the strings of L' among $u_1, \ldots, u_4$ are 1 and 4, with masses $3/4$ and $1/8$. So both hard distributions admit the public values $W = 4$ and $p_0 = 1/8$. Theorem 4.4, whose proof uses $f(n) \to \infty$ only through $f(n) \geq i^\star$, then applies with $f \equiv 2$ and $g(n) = \lfloor n/16 \rfloor$ and gives $\mathrm{GenErr} \leq 2e^{-n/64} + 2e^{-\varepsilon\lfloor n/16 \rfloor/2} + e^{-n/2}$ under both distributions. Together with Theorem 5.5, the optimal pure-DP rate on this instance is $\exp(-\Theta(\min\{1, \varepsilon\} \cdot n))$ for every $\varepsilon \in (0, \infty)$.

F Limitations and Future Work

Our results are information-theoretic: we characterize nearly optimal error rates but do not address computational efficiency, and our algorithms require explicit enumeration of languages and universe elements. Bridging these guarantees with the empirical DP-SGD pipeline for LLM training [Abadi et al., 2016, Li et al., 2022, Ponomareva et al., 2023, 2026] remains an important challenge. Recent work on computational barriers for (non-private) language generation [Arenas et al., 2026] suggests that efficient algorithms may need more structural assumptions.

For identification, our pure-DP upper bound and our lower bound carry the factor $\min\{1, \varepsilon\}$ at different scales, the former at the sublinear scale $r(n)$ and the latter at the linear scale n; the gap between them is inherited from the non-private setting [Høgsgaard and Pabbaraju, 2026]. For generation, the fully exponential rates use a public witness window (and, under pure DP, a public mass floor); without public side information, our guarantees are conditional on a coverage requirement.

Future directions include studying DP language learning in the online setting of Gold [1967] and Kleinberg and Mullainathan [2024], where adversarial samples and infinite-horizon composition require fundamentally different privacy analyses (the concurrent work of Mehrotra et al. [2026b] studies pure DP in this online setting for realizable sequences under continual release), and exploring user-level privacy [Liu et al., 2020, Levy et al., 2021, Ghazi et al., 2023], where the protected unit is an entire sequence of strings from a single user. Finally, extending our results to richer generation objectives that incorporate safety constraints [Anastasopoulos et al., 2026] or representativeness requirements [Peale et al., 2025] while maintaining differential privacy is a natural next step.

G Broader Impacts

This work is foundational research on differentially private language identification and generation. The contribution is information-theoretic: we give upper and lower bounds on the error rates achievable under pure and approximate differential privacy, which match for generation with a public window and mass floor. We discuss potential societal impacts below.

Positive impacts. Large language models are increasingly trained on user-generated text, which often contains sensitive personal information. Our results clarify the fundamental cost of training such models with formal privacy guarantees: under approximate (ε, δ)-DP with constant ε, no asymptotic statistical price is paid for either identification or generation (for generation, given a public witness window); under pure ε-DP, the cost of our algorithms is a $\min\{1, \varepsilon\}$ factor in the convergence exponent (for generation, with a public window and mass floor), and our lower bounds show that this factor cannot be avoided at the linear scale. By providing rigorous bounds and a design principle—that reformulating an objective to reduce its sensitivity can collapse the privacy–utility gap—this work can guide the development of privacy-preserving language learning systems and inform decisions about which privacy notion (pure vs. approximate) is appropriate for a given application.

Potential negative impacts. Because our contribution is purely theoretical, it does not directly enable any malicious use. We do not release any models, datasets, or code that could be misused. One indirect concern is that strong privacy guarantees can in principle be invoked to legitimize the deployment of models trained on sensitive data without further scrutiny; we note that DP guarantees protect individual contributions to the training data but do not, by themselves, address downstream issues such as fairness, representativeness, or content safety, which require complementary mechanisms.

Mitigations and scope. Our upper bounds are asymptotic, and all our bounds are information-theoretic; practitioners deploying DP language learning systems should still validate end-to-end privacy and utility empirically and combine DP with appropriate safeguards (auditing, content filtering, user-level privacy where appropriate). We see our work as one step toward a complete theoretical foundation rather than a deployment-ready system.

H Additional Related Work

Language Identification. Lange et al. [2008] provided a comprehensive survey of the identification-in-the-limit paradigm. Charikar et al. [2026] studied list identification, showing that allowing k guesses per step strictly expands identifiability. Peng et al. [2026] showed that augmenting Gold's paradigm with computational traces enables identification of all recursively enumerable languages. Papazov and Flammarion [2025] extend Gold's inductive inference framework with computational observations and restricted input sources to study learnability of computable functions in the limit.

Language Generation. Following Kleinberg and Mullainathan [2024], Li et al. [2025] placed generation within a broader learning-theoretic hierarchy. The tension between output diversity and hallucination was examined by Kalavasis et al. [2025, 2026], Charikar and Pabbaraju [2025], Peale et al. [2025], and Kleinberg and Wei [2025, 2026a,b], with Pareto-optimal tradeoffs studied by Charikar and Pabbaraju [2026]. Robustness under noise models was studied by Raman and Raman [2025], Bai et al. [2026], Mehrotra et al. [2026a], and Li and Zhang [2026]. On the structural side, Hanneke et al. [2025] and Bai et al. [2026] showed that generatability is not closed under finite unions for uncountable collections, and Karbasi et al. [2025] investigated automated hallucination detection. Computational and sample-complexity barriers were analyzed by Arenas et al. [2026]. More recent extensions include generation in continuous metric spaces [Li et al., 2026] and safe generation [Anastasopoulos et al., 2026]. Racca et al. [2026] study model collapse from a learning-theoretic perspective, showing replay creates provable separations for weaker notions of language generation.

Private Learning and Language Models. There is a rich literature on differentially private PAC learning [Kasiviswanathan et al., 2011, Bun et al., 2015, Bun, 2020] and private statistical estimation [Chaudhuri and Hsu, 2012, Aden-Ali et al., 2021, Bie et al., 2022]. On the applied side, Abadi et al. [2016] introduced DP-SGD, which has been extensively applied to large-scale language model training and fine-tuning [Anil et al., 2022, Yu et al., 2022, Li et al., 2022]. Ponomareva et al. [2023] provided a comprehensive practical guide to training machine learning models with differential privacy. Sander et al. [2023] studied scaling laws for private training, Yue et al. [2023] developed practical recipes for private synthetic text generation, Ghazi et al. [2023] studied user-level privacy with few examples per user, and Ponomareva et al. [2026] provided a practical guide to DP synthetic data generation. Gao et al. [2025] proposed a data-adaptive framework that synthesizes differentially private few-shot demonstrations for in-context learning with LLMs, improving the privacy–utility tradeoff without requiring public data. Liu and Bu [2025] proposed a hyperparameter-free DP training framework that eliminates the need for manually tuning the clipping threshold in DP-SGD, reducing the computational overhead of private fine-tuning for large language models. Marek et al. [2026] benchmark the practical privacy protection of DP-adapted LLMs via membership inference and canary extraction attacks. Thareja et al. [2026] propose DP-Fusion, a token-level differentially private inference mechanism for LLMs that bounds the influence of sensitive context tokens on the model's output. These works focus on empirical sides, whereas we establish information-theoretic guarantees. To our knowledge, ours is the first work to study differential privacy in the agnostic statistical framework of language identification and generation.

Comparison with other private learning tasks. Table 3 compares the privacy-induced change in several private learning tasks. Here α is selection or release accuracy, η is PAC error, and β is failure probability. The first five rows report sample or database complexity (the threshold and counting-query rows show only the domain-dependent term at fixed accuracy, confidence, and privacy), the sixth reports a uniform generation time, and the last two report task error as a function of n: expected excess risk for identification and failure probability for generation. We therefore compare the privacy-induced change within each task, not raw magnitudes across tasks. Approximate DP removes the pure-DP penalty in some tasks and only reduces it in others. The concurrent work of Mehrotra et al. [2026b] is the closest topical comparison: it studies pure DP under the stronger requirement of continual release, but for realizable generation and identification in the limit, so its rates are complementary to ours rather than directly comparable.

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 state the main claims and these are formally established by the theorems in Sections 3 to 5, with full proofs in the Appendix.

   Guidelines:

Table 3: Privacy-induced change across private learning tasks. Non-private baselines in the last two rows are due to Høgsgaard and Pabbaraju [2026].

Setting (quantity)	Non-private	Pure ε-DP	Approx. (ε, δ)-DP	Remark
Finite hypothesis selection, M public candidates (sample complexity) [Bun et al., 2019]	$O(\log(M/\beta)/\alpha^2)$	$O(\log(M/\beta)/\alpha^2 + \log(M/\beta)/(\alpha\varepsilon))$	the pure-DP bound already applies	returns a candidate within $(3+\zeta)\alpha$ if one is within α; score sensitivity $1/n$, gap $\Theta(\alpha)$
PAC learning Line_p, $\mathrm{VC} = \mathrm{Ldim} = 2$ (sample complexity) [Feldman and Xiao, 2015]	$O(\eta^{-1}\log(1/\beta))$, independent of p	$\Theta(\log p/(\varepsilon\eta))$ in its p-dependence	$O(\log(1/\delta)\log(1/\beta)/(\varepsilon\eta))$, independent of p	approximate DP removes the domain penalty; the learner is improper
PAC thresholds on finite ordered X (domain term) [Alon et al., 2019, Cohen et al., 2023, Feldman and Xiao, 2015]	no $\lvert X\rvert$-dependence	$\Theta(\log \lvert X\rvert)$	$\widetilde{\Theta}(\log^* \lvert X\rvert)$	approximate DP sharply reduces, but does not remove, the domain cost
Counting-query release, worst-case large Q (domain term) [Bun et al., 2014]	$\Theta(\log \lvert Q\rvert)$	$\Theta(d \log \lvert Q\rvert)$	$\widetilde{\Theta}(\sqrt{d}\log \lvert Q\rvert)$	the d-cost becomes $\sqrt{d}$; privacy is not free
d one-way marginals, average ℓ_1 error α (database size) [Steinke and Ullman, 2017]	exact empirical means	$\Theta(d/(\alpha\varepsilon))$	$\Theta(\sqrt{d\log(1/\delta)}/(\alpha\varepsilon))$ for $2^{-\Omega(n)} \le \delta$ and $\delta \le n^{-1-\Omega(1)}$	tight pure-to-approximate separation: d becomes $\sqrt{d}$
Generation in the limit, uniform time [Mehrotra et al., 2026b]	$d_{\mathrm{cl}} + 1$	$d_{\mathrm{cl}} + \widetilde{O}((k/\varepsilon)\log(1/\beta))$	not studied	class size k, closure dimension d_{cl}; adversarial enumeration, continual release, realizable
Identification, this work (expected excess risk)	$\exp(-r(n))$, any $r(n) = o(n)$	$\exp(-\min\{1,\varepsilon\}\, r(n))$	$\exp(-r(n))$ in the stated range	score sensitivity $\le 2f(n)^2/n$, gap ≥ 1; the growing horizon causes the sublinear exponent
Generation, this work, public window (error probability)	$\exp(-\Omega(n))$	$\exp(-\min\{1,\varepsilon\}\,\Omega(n))$ with a public mass floor	$\exp(-\Omega(n))$ in the stated range	prefix-count sensitivity ≤ 1, gap $g(n)$, linear in n with a public mass floor

- The answer [N/A] 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 [N/A] 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: Limitations are discussed in Section F of the Appendix.

   Guidelines:

   - The answer [N/A] 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: Every theorem explicitly states its assumptions (e.g., Assumption 3.1 for identification and Assumptions 4.2 and 4.3 for generation). The main text contains the theorem statements, proof ideas for the upper bounds (Sections 3 and 4), and the proofs of the lower bounds (Section 5); the complete proofs of the upper bounds are in Sections C and D, and Section E checks the hard instances of the lower bounds. The supporting lemmas (concentration, group privacy, private selection with a margin, the private two-point inequality) are stated with proofs or citations (Section A and Section 5).

   Guidelines:

   - The answer [N/A] 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: [N/A]

   Justification: This paper is purely theoretical and does not contain experiments.

   Guidelines:

   - The answer [N/A] 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: [N/A]

   Justification: This paper is purely theoretical and does not contain experiments requiring code or data.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.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://neurips.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) necessary to understand the results?

   Answer: [N/A]

   Justification: This paper is purely theoretical and does not contain experiments.

   Guidelines:

   - The answer [N/A] 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: [N/A]

Justification: This paper is purely theoretical and does not contain experiments.

Guidelines:

- The answer [N/A] 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: [N/A]

   Justification: This paper is purely theoretical and does not contain experiments.

   Guidelines:

   - The answer [N/A] 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: We have reviewed the NeurIPS Code of Ethics and the work conforms with it. The paper is a theoretical study of differential privacy and does not involve human subjects, sensitive data, or release of artifacts.

   Guidelines:

   - The answer [N/A] 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: Broader impacts are discussed in a dedicated section in the Appendix (Section G).

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] 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., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper is theoretical and does not release any data or models.

    Guidelines:

    - The answer [N/A] 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: [N/A]

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

    Guidelines:

- The answer [N/A] 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: [N/A]

    Justification: The paper does not release any new assets (datasets, code, or models).

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: The paper is purely theoretical and does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: The paper is purely theoretical and does not involve human subjects.

    Guidelines:

    - The answer [N/A] 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 rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: An AI assistant (Claude, Anthropic) was used to check the proofs and to draft parts of the camera-ready revision: the scope corrections of Section 4, Lemmas 5.1 and A.5 with the results derived from them (Theorems 5.3 and 5.5, Remark 5.6, and the constants of Theorems 3.2, 4.4, 4.5 and 4.8), and parts (c) of Corollaries C.9 and C.16. The acknowledgments state this, and the authors verified all statements and proofs.

    Guidelines:

    - The answer [N/A] 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 in the NeurIPS handbook for what should or should not be described.