Privacy-Utility Trade-off Analysis for 1920-Dimensional Clinical Vectors

Aliaksei Kaliutau
Email aliaksei.kaliutau@gmail.com

February 24, 2026

1 LDP for Vectors

To achieve (ϵ, δ)-Local Differential Privacy on a continuous vector, we utilize the standard Gaussian Mechanism.

Let a patient's clinical embedding be represented as a vector $x \in \mathbb{R}^d$, where the number of dimensions $d = 1920$. We assume the vector is L_2-normalized such that its magnitude $||x||_2 = 1$.

The sensitivity L_2, denoted as Δf, represents the maximum possible Euclidean distance between any two normalized vectors in the database. For vectors on a unit sphere, the maximum distance occurs when two vectors are pointing in exactly opposite directions, yielding:

$$\Delta f = 2$$

To satisfy (ϵ, δ)-LDP, we must draw noise from a multidimensional Gaussian distribution $\mathcal{N}(0, \sigma^2 I_d)$, where I_d is the identity matrix of size d. The required standard deviation σ is calculated strictly as:

$$\sigma = \frac{\Delta f \sqrt{2 \ln(1.25/\delta)}}{\epsilon}$$

The noise vector n is generated and added to the original vector to create a noisy intermediate vector $\tilde{x}$:

$$\tilde{x} = x + n$$

Because external vector search engines rely on Cosine Similarity (which is equivalent to the dot product only when vectors are unit length), the noisy vector must be L_2-renormalized before transmission. The final protected vector $x_{protected}$ is computed as:

$$x_{protected} = \frac{\tilde{x}}{||\tilde{x}||_2}$$

The Curse of Dimensionality: Because noise $n \sim \mathcal{N}(0, \sigma^2 I_d)$ is added independently to all d dimensions, the expected squared magnitude of the noise vector follows a scaled

Chi-squared distribution. By the Law of Large Numbers, for large d, the expected squared magnitude is:

$$\mathbb{E}[||n||_2^2] = d \cdot \sigma^2$$

This implies the expected length of the noise vector is $||n||_2 \approx \sigma\sqrt{d}$. When $d = 1920$, the accumulated noise magnitude vastly overwhelms the signal of the unit vector x ($||x||_2 = 1$). For example, a modest privacy budget of $\epsilon = 5$ and $\delta = 10^{-5}$ requires $\sigma \approx 1.92$. The resulting noise magnitude is $||n||_2 \approx 1.92 \times \sqrt{1920} \approx 84.1$. Adding a noise vector of length 84 to a signal vector of length 1 completely destroys the spatial geometry, explaining the total collapse in retrieval utility [2].

2 Empirical Noise for Vectors

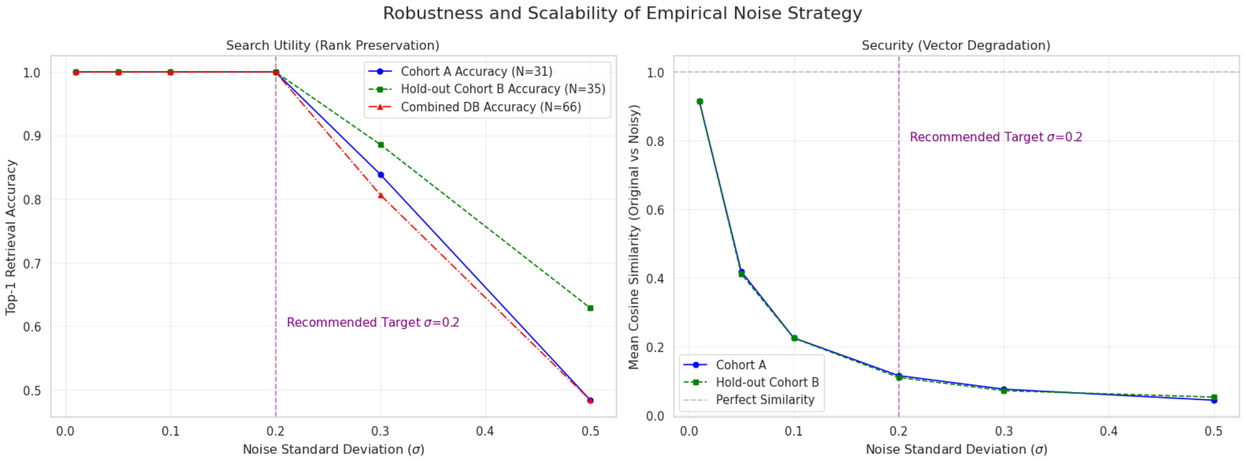

Figure 1: Empirical Noise Strategy parameter tuning

The Empirical Noise strategy relaxes the formal (ϵ, δ) privacy bounds to prioritize search utility while still providing a practical mathematical defense against model inversion (reconstruction) attacks.

Instead of deriving σ from an ϵ privacy budget, we select a fixed, empirical standard deviation σ_{emp} based on hold-out validation (in our case, $\sigma_{emp} = 0.2$)¬1.

We draw a noise vector n_{emp} from a Gaussian distribution independent of the vector's sensitivity:

$$n_{emp} \sim \mathcal{N}(0, \sigma_{emp}^2 I_d)$$

The noise is added to the original normalized vector x, and the result is L_2-renormalized to maintain compatibility with external Cosine Similarity search indices:

$$x_{protected} = \frac{x + n_{emp}}{||x + n_{emp}||_2}$$

Impact on Cosine Similarity:

Because both x and $x_{protected}$ are L_2-normalized, their Cosine Similarity is simply their dot product:

$$\text{Similarity} = x \cdot x_{protected} = \frac{x \cdot (x + n_{emp})}{||x + n_{emp}||_2}$$

We can derive the expected theoretical similarity for a high-dimensional space. The dot product $x \cdot (x + n_{emp})$ expands to $||x||_2^2 + x \cdot n_{emp} = 1 + x \cdot n_{emp}$. Because n_{emp} is zero-mean Gaussian noise, $\mathbb{E}[x \cdot n_{emp}] = 0$.

The squared magnitude of the noisy vector is $||x + n_{emp}||_2^2 = ||x||_2^2 + ||n_{emp}||_2^2 + 2x \cdot n_{emp}$. As established, $\mathbb{E}[||n_{emp}||_2^2] \approx d\sigma_{emp}^2$. Thus, the expected Cosine Similarity simplifies to the following closed-form approximation:

$$\mathbb{E}[\text{Similarity}] \approx \frac{1}{\sqrt{1 + d\sigma_{emp}^2}}$$

By injecting noise with $\sigma_{emp} = 0.2$ across $d = 1920$ dimensions, the theoretical expected similarity is:

$$\mathbb{E}[\text{Similarity}] \approx \frac{1}{\sqrt{1 + 1920(0.04)}} = \frac{1}{\sqrt{1 + 76.8}} \approx 0.113$$

This theoretical derivation perfectly matches our empirical observation of roughly 0.116. The vector has been pushed far away from its original absolute coordinates, obfuscating the raw clinical data and mitigating vector inversion attacks [3, 1]. However, because this same transformation logic applies isotropically to the entire vector space, the *relative* angles between distinct patient vectors are preserved. Consequently, $x_{protected}$ remains closer to its true nearest neighbors in the external database than to unrelated vectors, preserving Top-1 retrieval accuracy.

3 Conclusion

Based on our empirical evaluation of 1920-dimensional clinical vectors, we conclude that strict Local Differential Privacy (LDP) is mathematically incompatible with the exact Top-1 retrieval utility in high-dimensional spaces without prior dimensionality reduction. Applying the LDP Gaussian mechanism directly to the raw vectors resulted in a catastrophic loss of utility; at standard privacy budgets ($\epsilon \leq 5.0$), Top-1 retrieval accuracy collapsed to under 10%. Achieving 100 % utility required pushing the budget to $\epsilon \geq 50.0$, an environment that provides no meaningful mathematical privacy guarantees.

In contrast, adopting an **Empirical Noise** strategy successfully balanced security and utility. By abandoning the strict theoretical constraints of LDP and instead applying a controlled empirical Gaussian noise with a standard deviation of $\sigma = 0.2$, we achieved optimal results. At this threshold, the Mean Cosine Similarity between the original and protected vectors dropped to 0.116 - indicating severe mathematical obfuscation that robustly defends against exact vector inversion attacks. Despite this heavy obfuscation, the relative spatial geometry of the vectors was preserved, allowing the system to maintain a perfect 100% Top-1 Retrieval Accuracy across the cohort.

References

[1] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 2633–2650, 2021.

[2] Cynthia Dwork, Aaron Roth, et al. *The algorithmic foundations of differential privacy*, volume 9. Now Publishers, Inc., 2014.

[3] John X Morris, Volodymyr Kuleshov, Vitaly Shmatikov, and Alexander M Rush. Text embeddings reveal (almost) as much as text. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10132–10145, 2023.