

Walkthrough

Privacy by Design

A Strategic Framework for Choosing Between Encryption and Noise

Group Members:

Max Forrest (30301436)

Sukhmanveer Singh (30291188)

Chirag Kumar (30273579)

Karanveer Singh (30282286)

Have you ever wondered how organizations can analyze your data while keeping it private? Or how companies like Apple and Google can gather insights from millions of users without seeing anyone's individual information? And most importantly, how do they decide which privacy technology to use?

This project takes you behind the scenes of how organizations can perform computations on sensitive data, how they protect individual privacy, and how to choose the right approach for different scenarios.

This handbook is designed to make it simple to learn about privacy-preserving technologies. You don't need to be a privacy expert to follow along, just a curious mind and willingness to explore how modern technology keeps your data safe while still being useful.

How to Access & Use the Material:

Our project comes with three parts: Walkthrough, Handbook, and a Python-based Jupyter Notebook. Start with the Walkthrough; it gives you a quick overview of what to expect and how everything connects. Then, move on to the Handbook for a deeper dive into the concepts and comparisons, which then leads you to our decision framework. Finally, the Jupyter Notebook walks you through our demonstrations. The Jupyter Notebook includes direct hyperlinks to the real-world datasets used in each demonstration, allowing readers to inspect the data sources and reproduce the results independently. These datasets are loaded within the notebook, and readers can modify parameters such as dataset size or privacy settings to observe how results change in real time. It shows exactly how Homomorphic Encryption (HE) and Differential Privacy (DP) work in practice with real datasets, and we compare them based on seven points of comparison. Together, these pieces make it easy to follow along, learn the theory, and see it all in action.

Overview of Content:

We begin with a clear overview of the privacy paradox facing modern organizations and why privacy-preserving technologies are essential in today's data-driven world. From there, you'll explore two fundamental approaches to protecting privacy: HE and DP, understanding how each works, when to use them, and their respective strengths and limitations.

The background covers the computational privacy gap, introduces Privacy-Preserving Technologies (PPT), and establishes why choosing the right approach matters for compliance with regulations like General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and Health Insurance Portability and Accountability Act (HIPAA).

You'll discover how Homomorphic Encryption allows computations directly on encrypted data without decryption. We explain the different types like Partially Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SHE), Fully Homomorphic Encryption (FHE), major schemes like Brakerski-Gentry-Vaikuntanathan (BGV), Brakerski-Fan-Vercauteren (BFV), and Cheon-Kim-Kim-Song (CKKS), and how they provide cryptographic security guarantees. You'll learn about supported operations, performance characteristics, and the trade-offs between security and computational cost.

You'll also learn about Differential Privacy, which adds carefully calibrated statistical noise to protect individual privacy while preserving data utility. We cover the formal mathematical definition, epsilon (ϵ) and delta (δ) parameters, noise injection mechanisms (Laplace and Gaussian), and how privacy budgets work. You'll understand the privacy-utility tradeoff and how DP scales efficiently with large datasets.

We present a comparison across seven critical dimensions: data granularity, processing efficiency, range of supported operations, data utility as size increases, scalability, hardware and computational cost, and representative use cases. This systematic comparison helps you understand when to choose HE versus DP based on your specific requirements.

To bridge theory and practice, we present hands-on demonstrations using Python that show both technologies in action. Using the TenSEAL library for Homomorphic Encryption and OpenDP for Differential Privacy, we process real healthcare and crime datasets. The demos measure encryption time, computation time, accuracy, memory overhead, and scalability across different dataset sizes. You'll see concrete performance metrics, visualizations comparing both approaches, and practical examples of computing sums and means while preserving privacy.

What Makes Our Approach Stand Out:

We don't only tell you about concepts, we demonstrate them. Our Jupyter Notebook acts as an interactive companion to the handbook and includes code implementation with realistic datasets, performance benchmarks, and detailed visualizations which play a central role in our notebook, with graphs, histograms and comparative plots, so that complex ideas become easier to grasp which otherwise might be difficult to convey through text alone. You'll see exactly how Homomorphic Encryption maintains perfect accuracy but requires significant computational resources, while Differential Privacy offers tunable privacy guarantees with minimal overhead. If you're unfamiliar with Jupyter Notebooks, we provide a comprehensive readme that walks you through the process of setting up and exploring the notebook.

If you're new to privacy-preserving technologies or cryptography, our handbook is your new best friend. It's written in plain language and explains every key term. It helps you understand the mathematical foundations behind encryption schemes, differential privacy guarantees, and the formal security properties that protect your data.

Why This Matters:

Understanding which privacy-preserving technology to use isn't just for data scientists, but for anyone working with sensitive data in healthcare, finance, research, or any field that values privacy. This handbook illustrates how organizations can make informed decisions about privacy technologies, how they balance privacy protection with data utility, and how to evaluate trade-offs between security, performance, and cost.

By the end of this, you will have a solid understanding of:

- How Homomorphic Encryption and Differential Privacy work, and how cryptographic and statistical privacy approaches differ.
- When to choose HE versus DP based on your use case requirements.
- The practical implications of privacy-utility trade-offs and how to evaluate privacy technologies using concrete comparison dimensions.

You'll be ready to dive into the full handbook, explore our interactive demos, and gain the knowledge needed to make strategic decisions about privacy-preserving technologies in real-world applications.