**All you need to know about Nabla's privacy and security features**

Clément Baudelaire

ML Product Manager

*Last updated: April 23, 2024.*

Security and compliance are the backbone of healthcare. At Nabla, we place **privacy at the top of our agenda** because it is fundamentally tied to our customers' experience of our products. We are committed by design to securing customer application data, eliminating system vulnerability, and ensuring continuity of access.

In practical terms, this means we use a variety of industry-standard technologies, services, and processes to **secure data** against unauthorized access, disclosure, use, and loss.

[Nabla](https://copilot.nabla.com/) is an **AI ambient assistant** designed to **reduce the clinical documentation burden**. We do this by automatically generating the clinical note of any encounter.

Because we believe transparency is everything, this article details **how data is captured, stored, and processed when a physician uses Nabla**. We built this product ensuring the highest level of security and compliance, without compromising the final output of the medical note. An approximate note means physicians would have to edit it and divert from care again - and that would be defeating the purpose, which is maximum quality with maximum privacy.

**Overview of the Nabla data flow**

**The short version is this:** Nabla turns a raw medical conversation that occurred during a consultation into a structured medical note that can be exported directly to the patient's EMR.

**The longer version now:**

Simply put, when the doctor explicitly clicks on "Start encounter", Nabla captures the audio of the encounter between the doctor and the patient.

The audio is then transcribed live using our HIPAA-compliant speech-to-text external API.

When the doctor clicks on "Stop encounter", the audio capture is immediately stopped.

The transcript produced by the API is then processed to generate a structured note (method clinicians use to document patient encounters in a certain way like Subjective, Objective, Assessment & Plan). Our home-grown engine combines the best available Large Language Models (LLM) with our own customized LLMs specifically trained for this usecase. HIPAA compliance is maintained during the entire process.

Here is a neat graph that sums it up:*Overview of the Nabla data flow*

**Data storage and processing**

Because healthcare data security is so essential, Nabla takes a very conservative approach to data storage. We only store transcripts and notes **temporarily**, for a **configurable period of time** (14 days by default). The only exception is data submitted as feedback (see below). Audio recordings are **not retained at all**. Nabla only processes individual chunks of audio and discards them afterwards. The full audio recording of the encounter is never held in memory or stored on disk.

The goal of this retention period is to give enough time for physicians to review, edit and export their note to their EHR software. After expiration, data is immediately removed from the application and only remains in our backups. 7 days later, backups also expire and patient data is **entirely gone from our systems**.

All stored transcripts and notes are protected with two layers of encryption (system and application), using strong cryptography standards.

This data processing is done on Nabla's infrastructure in Google Cloud Platform (GCP) and Microsoft Azure, in strict compliance with both HIPAA and GDPR. To make sure patient data is never stored outside our control, we also have agreements to opt out of data retention for all services used to process it.

**Feedback & privacy**

To provide the best support experience and make Nabla even better over time, we allow physicians to send feedback after each encounter. This is optional and can be deactivated for organizations. When providing feedback, physicians may choose to attach the generated note and transcript. Since data submitted this way is permanently stored, we use a **deidentification algorithm** to systematically remove all portions of the transcripts and notes that contain personally identifiable information.

In practice, this algorithm masks the [18 HIPAA identifiers](https://www.luc.edu/its/aboutus/itspoliciesguidelines/hipaainformation/the18hipaaidentifiers/) including names, addresses, dates, phone and fax numbers, SSN, medical record numbers, health plan beneficiary, account, certificate or license numbers, vehicle, device or serial identifiers, URLs and IP addresses. Here's what this pseudonymisation process looks like:

**"My name is Clément, I was born on 06.16"** becomes **"My name is [masked\_name\_001], I was born on [masked\_date\_001]".**

The masked version is the one we end up using for support and AI training. This is in line with our general philosophy, which is: only store data that you actually need to deliver value to physicians. Storing only deidentified feedback data, instead of all data transiting through our servers, allows us to focus AI training on data that has the highest learning value for our models, and minimize risk for all users. It also gives more control to organizations, who may choose between reaping the benefits of better support and improved AI models over time, or restrict all feedback collection if required by their own data security policies.

**Information security**

Because Nabla retains medical data temporarily and may receive additional feedback data from physicians, we put in place a comprehensive information security program to ensure safe handling of data. This includes employee training, third-party audits and penetration testing, strict management of roles & permissions, strong authentication processes, encryption at rest and in transit, continuous vulnerability scanning, logging, monitoring, alerting and much more.

We are **SOC 2 Type 2 compliant** and **ISO 27001** certified.