

NFON Intelligent Assistant

Anruf Transkription & Zusammenfassung – Technische Transparenznotiz zu KI

Mai 2025 – V1.0 – Final

Inhalt

Inhalt.....	1
Einführung.....	1
Übersicht über die Funktionen	2
Modellübersicht.....	2
Modellarchitektur	2
Verwendungsrichtlinien.....	3
Modelleingaben und -ausgaben.....	3
Datenquellen für Training und Bewertung.....	3
Modellbewertung und -leistung	4
Sicherheits- und ethische Überlegungen.....	4
Datenschutz und Sicherheit	4
Referenzen	4

Einführung

NFON hat sich verpflichtet, künstliche Intelligenz (KI) verantwortungsbewusst einzusetzen, um die Zugänglichkeit und Benutzerfreundlichkeit der Geschäftskommunikation zu verbessern. Mit der **Anruf-Transkription und -Zusammenfassung** erhalten Nutzer durch **automatische, KI-generierte Transkripte und Zusammenfassungen** nach dem Anruf einen besseren Einblick in den Inhalt ihrer Anrufe.

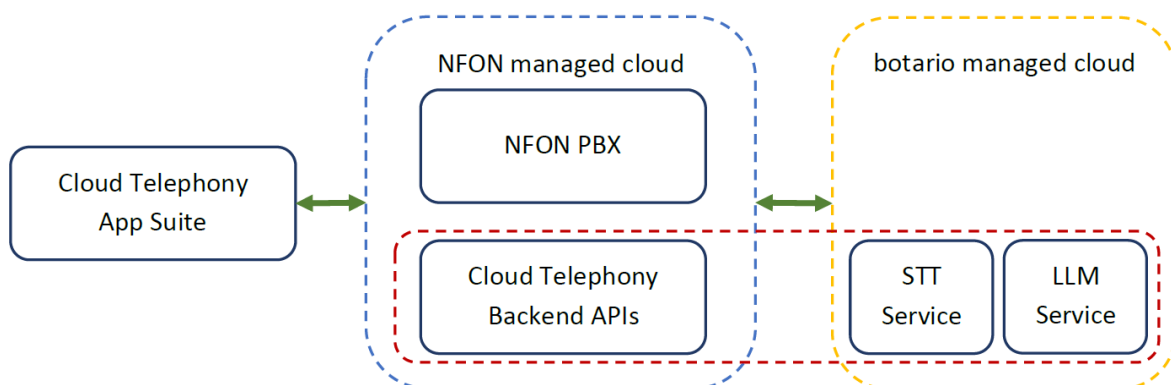
Diese Funktion wurde im Einklang mit den Datenschutz- und Sicherheitsgrundsätzen von NFON entwickelt und gewährleistet Transparenz, Kontrolle und die Einhaltung der einschlägigen Datenschutzbestimmungen. Dieses Dokument beschreibt den Aufbau, die Funktionalität und die Sicherheitsvorkehrungen der Funktion „Anruf-Transkription und -Zusammenfassung“.

Übersicht über die Funktionen

Die Anruf-Transkription und -Zusammenfassung ist eine **abonnementbasierte Funktion**, die für Nutzer des Cloud-Telefoniedienstes von NFON verfügbar ist. Nach dem Kauf können Administratoren die Funktion einzelnen Nutzern und deren zugehörigen Nebenstellen zuweisen.

Nachdem alle Anrufteilnehmer den **Bedingungen der Transkription zugestimmt** haben, wird der Prozess während eines Anrufs automatisch ausgelöst und läuft in Echtzeit im Hintergrund ab. Nach Beendigung des Anrufs initiiert das System eine Zusammenfassungsaufgabe, die **ausschließlich auf dem Transkript des Anrufs** basiert. Sowohl das vollständige Transkript als auch die Zusammenfassung werden dem Nutzer im Abschnitt „Rufliste“ der NFON-Client-Oberfläche zur Verfügung gestellt. Der Zugriff ist **streng** auf den Nutzer **beschränkt**, der an dem Anruf teilgenommen hat.

Die Funktion ist so konzipiert, dass sie unabhängig von der Anrufaufzeichnung funktioniert. Audioaufzeichnungen werden nur gespeichert, wenn andere Aufzeichnungsdienste separat aktiviert wurden.



- HTTPS-gesicherte Verbindungen
- Zustandslose Vorgänge: Es werden keine Kunden- oder personenbezogenen Daten gespeichert.

Modellübersicht

Modellarchitektur

Die Transkription von Anrufen erfolgt mit demselben **Whisper-basierten Speech-to-Text-Modell (STT)**, das auch für die Transkription von Voicemails verwendet wird. Dieses Modell wird von botario (einem Unternehmen von NFON) verwaltet und optimiert und läuft in einer dedizierten, GPU-basierten Cloud-Umgebung **in der EU**.

Die Zusammenfassung wird **nach dem Anruf** mithilfe eines großen Sprachmodells (LLM) – **Gemma-3-27B-IT** – durchgeführt, das ebenfalls auf demselben von botario

verwalteten GPU-Cluster gehostet wird. Die Zusammenfassung wird **ausschließlich auf der Grundlage der Transkription des Anrufs erstellt**, ohne zusätzliche externe Datenquellen.

Die Verarbeitungsumgebung umfasst mehrere Server für Redundanz und Leistung sowie eine Infrastruktur zur Lastverteilung. **Alle Daten werden in der EU verarbeitet und in Deutschland gespeichert.**

Verwendungsrichtlinien

Benutzer werden erst nach Beendigung eines Anrufs und Abschluss der Zusammenfassung über die Verfügbarkeit von Transkripten und Zusammenfassungen benachrichtigt. Diese Ergebnisse sind im NFON-Client unter dem entsprechenden Eintrag in der **Rufliste** zugänglich.

Keine anderen Parteien haben Zugriff auf die Anruferzusammenfassungen oder Transkripte anderer Benutzer. Das Design stellt sicher, dass nur der direkt beteiligte Benutzer die generierten Inhalte einsehen kann.

Modelleingaben und -ausgaben

Die Eingabe in das System ist der **Audiostream** eines Live-Anrufs, der mithilfe der **TLS 1.3-Verschlüsselung** sicher von der NFON-Telefonanlage an den botario STT-Dienst übertragen wird. Die **Ausgabe** dieses Schritts ist die textuelle Transkription des Anrufs.

Diese Transkription dient als Eingabe für den zweiten Schritt – die Zusammenfassung –, bei dem eine strukturierte **Textzusammenfassung** des Gesprächsinhalts erstellt wird. Sowohl die Transkription als auch die Zusammenfassung werden dann in der Cloud-Infrastruktur von NFON gespeichert und mit dem jeweiligen Gesprächsverlaufselement verknüpft.

Datenquellen für Training und Bewertung

Das STT-Modell basiert auf **OpenAI's Whisper** und einem **fein abgestimmten Wav2Vec-Modell**, das unter Verwendung öffentlich zugänglicher Datensätze und zusätzlicher Daten, die in Zusammenarbeit mit dem ITSC (Institut für Technische Kommunikationssysteme) gesammelt wurden, trainiert wurde. Für das Training oder die Feinabstimmung werden keine Kundendaten verwendet.

Für die Zusammenfassungskomponente wurde das Modell **Gemma-3-27B-IT** auf einem breiten Korpus trainiert, das zwar nicht explizit detailliert beschrieben wird, aber vermutlich **öffentlich zugängliche und synthetische Datensätze** umfasst. NFON verwendet keine Kundendaten für das Training oder die Feinabstimmung dieses Modells.

Modellbewertung und -leistung

NFON / botario führt **regelmäßig Bewertungen und Evaluierungen** der KI-Modelle durch, die die Anruf-Transkription und -Zusammenfassung unterstützen, um eine hohe Leistung, Genauigkeit und Zuverlässigkeit sicherzustellen. Dazu gehören **Test- und Qualitätssicherungsprozesse**, bei denen die **menschliche Überwachung** eine wichtige Rolle bei der Validierung des Modellverhaltens, der Identifizierung potenzieller Verbesserungsmöglichkeiten und der Aufrechterhaltung der Konsistenz über verschiedene Sprachmuster und Akzente hinweg spielt.

Sicherheits- und ethische Überlegungen

Die Zusammenfassungsfunktion ersetzt nicht das menschliche Urteilsvermögen und dient lediglich der **Unterstützung** und nicht der Automatisierung der Entscheidungsfindung. Sowohl die Transkription als auch die Zusammenfassung sind **neutrale Ergebnisse**, die den Inhalt des ursprünglichen Gesprächs widerspiegeln.

Es wird darauf geachtet, dass die Modelle bei unterschiedlichen Sprachmustern, Akzenten und Inhaltstypen fair funktionieren. Es werden **keine automatisierten Entscheidungen** auf der Grundlage des Inhalts eines Anrufs getroffen, und der Benutzer behält die Kontrolle darüber, wie die Transkription und die Zusammenfassung verwendet werden.

Datenschutz und Sicherheit

Die gesamte Verarbeitung erfolgt innerhalb der Infrastruktur von NFON und seinen vertrauenswürdigen Partnern mit **Sitz in der EU**. Audiodaten werden sicher übertragen und nicht gespeichert, sofern dies nicht separat über andere Dienste konfiguriert wurde.

Die Transkriptionen und Zusammenfassungen werden sicher in der NFON-Cloud gespeichert und sind **nur für den Nutzer zugänglich, der an dem Anruf teilgenommen hat**. Dies gewährleistet Vertraulichkeit und entspricht den Grundsätzen der Datenminimierung und Zugriffskontrolle.

Referenzen

- [NFON AGB & SLA](#)
- [NFON Datenschutzhinweise](#)
- [NFON Trust Center](#)
- [OpenAI Whisper](#)
- [Gemma 3 Technical Report](#)