

NFON Intelligent Assistant

Voicemail Transkription – Technische Transparenznotiz zu KI

August 2026

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	5
Referenzen	5

Einführung

Wir bei NFON erkennen das Potenzial künstlicher Intelligenz (KI) zur Verbesserung der Geschäftskommunikation, indem Informationen leichter zugänglich und besser verwertbar gemacht werden. Gleichzeitig verpflichten wir uns, KI so einzusetzen, dass Transparenz, Datenschutz und Sicherheit im Vordergrund stehen.

Unsere Voicemail-Transkriptionen wurden unter Berücksichtigung dieser Grundsätze entwickelt. Mit dieser Funktion können Benutzer zusätzlich zu den aufgezeichneten Audio-Nachrichten eine **Text-Transkription ihrer Voicemail-Nachrichten** erhalten. Der Transkriptionsprozess ist vollständig automatisiert und erfordert keine Überprüfung durch Menschen.

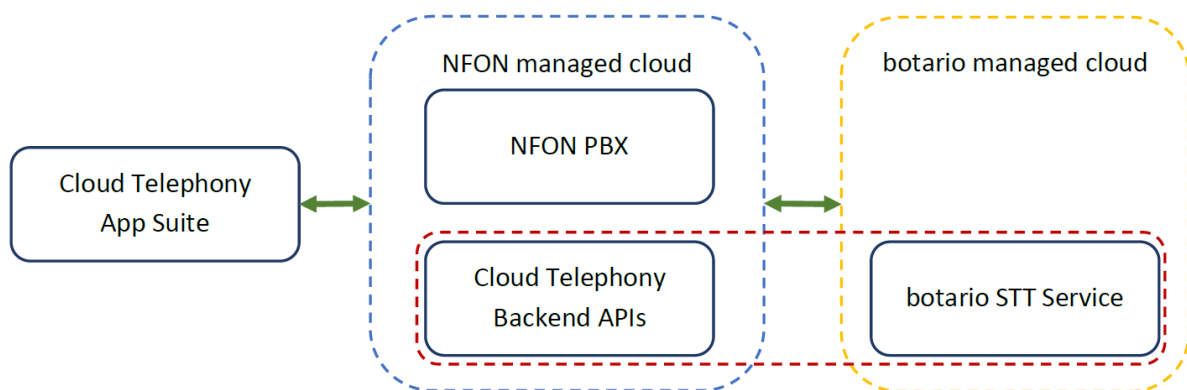
Dieses Dokument bietet einen Überblick über die Funktionsweise von Voicemail-Transkriptionen, das zugrunde liegende KI-Modell und die Sicherheitsvorkehrungen, die eine verantwortungsvolle Nutzung der KI gewährleisten.

Übersicht über die Funktionen

Voicemail-Transkriptionen sind eine **optionale Funktion**, die PBX-Administratoren für **Business Premium**-Verträge entweder systemweit oder für einzelne Nebenstellen ohne zusätzliche Gebühren aktivieren können. Bei Aktivierung werden Voicemail-Nachrichten automatisch mithilfe eines **Speech-to-Text-Modells (STT)** transkribiert, und der resultierende Text wird neben der ursprünglichen Voicemail-Aufzeichnung zur Verfügung gestellt.

Die Funktion wird **bei Bedarf** ausgeführt und ändert nichts an der Art und Weise, wie Voicemails aufgezeichnet oder gespeichert werden. Die Transkriptionen werden direkt nach der Aufzeichnung der Voicemail erstellt. Die Audioaufzeichnungen werden sicher übertragen, und die Transkriptionen werden verarbeitet, ohne über die Dauer des KI-Inferenzprozesses hinaus im KI-System gespeichert zu werden.

Voicemail-Transkriptionen sind **standardmäßig deaktiviert**, sodass Benutzer sich vor der Transkription ihrer Voicemails dafür entscheiden müssen.



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

Modellübersicht

Modellarchitektur

Die für „Voicemail-Transkriptionen“ zugrundeliegende Speech-to-Text Funktion basiert auf einem **Whisper-basierten KI-Modell (whisper-large-v3-turbo)**, das von NFON zusammen mit einer eigenen Implementierung mit spezifischen Schutzmechanismen sowie unseren eigenen Fine-Tunings der NVIDIA-Modelle **Canary (canary-1b-v2)** und **Parakeet (parakeet-tdt-0.6b-v3)** optimiert wurde. Aufgrund des unternehmenseigenen Fine-Tunings sind die eingesetzten Modelle nicht vollständig mit der von NVIDIA bereitgestellten Standardversion vergleichbar.

Das Modell läuft auf einer **sicheren Cloud-Infrastruktur**, die in **Deutschland** gehostet wird. Darüber hinaus hat die Einhaltung der europäischen Datenschutzbestimmungen und des neuen EU-KI-Gesetzes für NFON oberste Priorität.

Die KI-Verarbeitungsumgebung besteht aus **vier dedizierten Servern** mit zwei zusätzlichen Servern für Redundanz. Ein **Load Balancer** sorgt für eine optimale Verteilung der Anfragen. Der Transkriptionsprozess findet vollständig im **Arbeitsspeicher** statt, sodass weder die Original-Audiodatei noch der transkribierte Text dauerhaft gespeichert werden. Sobald die Transkription abgeschlossen ist und als API-Antwort zurückgegeben wird, werden alle temporären Verarbeitungsdaten gelöscht.

Verwendungsrichtlinien

Die Transkription von Voicemails ist nur verfügbar, wenn sie ausdrücklich aktiviert ist. Ist die Funktion deaktiviert, bleiben Voicemail-Nachrichten unverändert und werden nur als Audioaufzeichnungen gespeichert. Bei Aktivierung wird die Voicemail-Audioaufnahme **sicher** über TLS 1.3 an den Transkriptionsdienst **übertragen**, wo das KI-Modell die Texttranskription erstellt. Der transkribierte Text wird dann an die NFON-Geschäftstelefonieplattform zurückgesendet und dem Benutzer über eine **E-Mail-Benachrichtigung** sowie über die **Telefonie-Clients** zugänglich gemacht.

Da die Funktion zur Verbesserung der Barrierefreiheit entwickelt wurde, ist sie für eine hohe Genauigkeit bei verschiedenen Akzenten und Variationen der gesprochenen deutschen Sprache optimiert. Wie bei jedem KI-basierten Transkriptionsdienst können jedoch aufgrund von Hintergrundgeräuschen, unklarer Sprache oder komplexer Terminologie gelegentlich Ungenauigkeiten auftreten.

Modelleingaben und -ausgaben

Die Funktion „Voicemail-Transkriptionen“ funktioniert auf einer einfachen Input-Output-Basis. Der **Input** für das Modell ist die Voicemail-**Audiodatei**, die sicher von der NFON-Cloud-Telefonieplattform übertragen wird. Der **Output** ist die entsprechende **Texttranskription**, die dann dem Benutzer angezeigt wird. Das KI-System generiert keine zusätzlichen Metadaten.

Datenquellen für Training und Bewertung

Das von OpenAI entwickelte **Whisper**-Modell wurde anhand eines umfangreichen Datensatzes trainiert, der **680.000 Stunden mehrsprachiger und multitaskingfähiger überwachter Daten** umfasst, die aus dem Internet gesammelt wurden. Dieser umfangreiche Datensatz enthält eine Vielzahl von Sprachen und Sprachmustern, wodurch die Robustheit des Modells gegenüber Akzenten, Hintergrundgeräuschen und Fachsprache verbessert wird.

Zusätzlich zur Verwendung von OpenAIs **Whisper-Modell** integriert unser Transkriptionssystem die Nvidia Modelle **Canary** und **Parakeet**. Diese Modelle wurden unter Verwendung von öffentlich verfügbaren, offen lizenzierten Sprachdatensätze feinabgestimmt. Durch die Kombination von starken Open-Source-Grundlagen und unserem eigenen Feinabstimmungsprozess kann das Modell eine genaue und robuste Transkription über viele Sprachen, Akzente und Aufnahmebedingungen hinweg bereitstellen.

Um die Einhaltung der Datenschutzgrundsätze zu gewährleisten, werden **keine Kundendaten** für das Training oder die Feinabstimmung der KI-Modelle verwendet. Stattdessen werden öffentlich zugängliche Datensätze und Daten verwendet, die im Rahmen autorisierter Kooperationen gesammelt wurden. Die Modelle werden anhand von Metriken wie Genauigkeit, Vollständigkeit und Fehlerbehandlung bewertet, um eine hohe Qualität der Transkriptionsdienste zu gewährleisten.

Modellbewertung und -leistung

NFON führt **regelmäßig Bewertungen und Evaluierungen** der KI-Modelle durch, die die Voicemail-Transkriptionen 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 Gewährleistung der Konsistenz über verschiedene Sprachmuster und Akzente hinweg spielt.

Sicherheits- und ethische Überlegungen

Die Voicemail-Transkription ist darauf ausgelegt, **gesprochene Sprache unverändert zu transkribieren**, ohne Inhalte zu verändern, zu filtern oder zu interpretieren. Das KI-Modell führt keine Stimmungsanalyse, Inhaltsklassifizierung oder automatisierte Entscheidungsfindung durch. Der Zweck dieser Funktion beschränkt sich strikt auf die **Umwandlung von Sprache in Text** und stellt sicher, dass sie ein neutrales und transparentes Werkzeug bleibt.

Das System ist auf eine **faire und unvoreingenommene Transkription** optimiert und zielt darauf ab, mögliche Unterschiede in der Transkriptionsqualität zwischen verschiedenen Nutzergruppen zu minimieren. Dennoch kann KI-basierte Spracherkennung abhängig von Aussprache und Umgebungsbedingungen geringfügige Genauigkeitsunterschiede aufweisen.

NFON verpflichtet sich weiterhin, **Fairness sicherzustellen** und **Verzerrungen** in KI-generierten Transkriptionen durch kontinuierliche Modellbewertung und -verbesserung **zu reduzieren**.

Datenschutz und Sicherheit

NFON folgt den Prinzipien von **Privacy by Design** und stellt sicher, dass die Voicemail-Transkription sicher und rechtskonform betrieben wird. Der **gesamte Transkriptionsprozess erfolgt unmittelbar, ohne dass Audio- oder Textdaten dauerhaft im KI-System gespeichert werden.**

Alle Datenübertragungen zwischen der NFON Cloud-Telefonieplattform und dem Transkriptionsdienst sind mittels **TLS (HTTPS)** verschlüsselt. Die KI-Verarbeitungsumgebung wird **vollständig in Deutschland betrieben**, ohne jegliche Datenübertragung außerhalb des Europäischen Wirtschaftsraums (EWR).

Da das KI-System weder Transkriptionen speichert noch diese für eine weitere Verarbeitung nutzt, **verbleiben die Nutzerdaten jederzeit vollständig unter der Kontrolle der Nutzer.**

Referenzen

- [NFON AGB & SLA](#)
- [NFON Datenschutzhinweise](#)
- [NFON Trust Center](#)
- [Botario TTS/STT](#)
- [OpenAI Whisper](#)
- [Nvidia Canary](#)
- [Nvidia Parakeet](#)
- [Technical Paper für Canary + Parakeet von Nvidia](#)