NFON Intelligent Assistant

Voicemail Transcriptions – AI Transparency Technical Note

March 2025 – v1.0

Contents

Introduction

At NFON, we recognize the potential of artificial intelligence (AI) to improve business communications by making information more accessible and actionable. At the same time, we are committed to applying AI in a way that prioritizes transparency, privacy, and security.

Our Voicemail Transcriptions have been designed with these principles at its core. This feature enables users to receive a **text transcription of their voicemail messages** in addition to recorded audio. The transcription process is fully automated and does not involve human review.

This document provides an overview of how Voicemail Transcriptions works, the AI model that powers it, and the safeguards in place to ensure responsible AI usage.

Feature Overview

Voicemail Transcriptions is an **optional feature** that PBX administrators can enable for **Business Premium** contracts at either the system-wide level or for individual extensions without any additional fees. When activated, voicemail messages are transcribed automatically using a **Speech-to-Text (STT) model**, and the resulting text is made available alongside the original voicemail recording.

The feature operates **on demand** and does not modify the way voicemails are recorded or stored. The transcriptions are generated directly after the voicemail has been recorded. Audio recordings are securely transmitted, and transcriptions are processed without being stored on the AI system beyond the duration of the AI inference process.

Voicemail Transcriptions is **disabled by default**, ensuring that users opt in before their voicemails are transcribed.
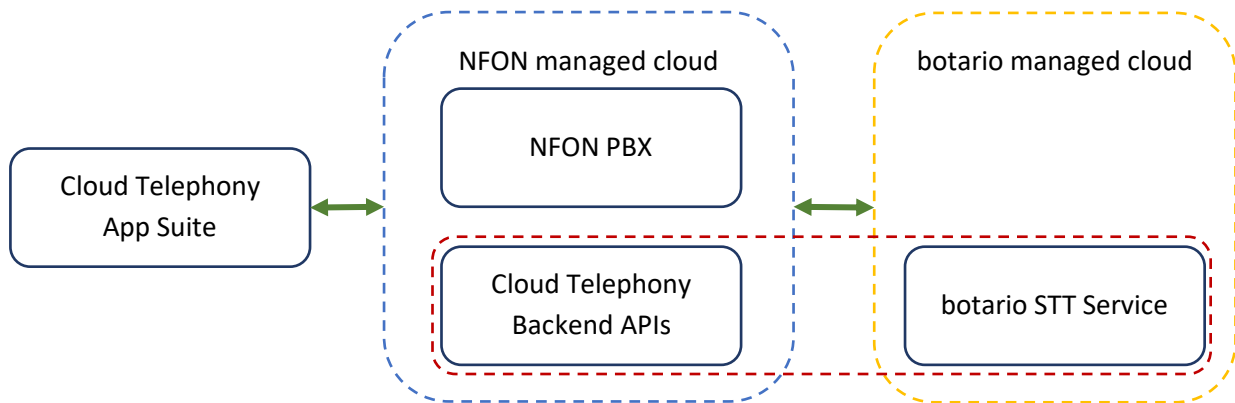
● HTTPS secured connections

● Stateless operations: no customer data is persisted in the botario STT Service

Model Overview

Model Architecture

The Voicemail Transcriptions feature is powered by a **Whisper-based AI model**, optimized by **botario**, an NFON company. The model runs on a **secure cloud infrastructure** hosted in **Germany.** In addition, ensuring compliance with European data protection regulations and the emerging EU AI Act is of priority for NFON and botario.

The AI processing environment consists of **four dedicated servers** with two additional servers for redundancy. A **load balancer** ensures optimal distribution of requests. The transcription process takes place entirely **in-memory**, meaning that neither the original audio nor the transcribed text is stored permanently. Once the transcription is completed and returned as an API response, all temporary processing data is deleted.

Usage Guidelines

Voicemail Transcriptions is available only when explicitly enabled. If the feature is disabled, voicemail messages remain unchanged and are stored as audio recordings only. When enabled, the voicemail audio is **securely transmitted** via TLS 1.3 to the transcription service, where the AI model generates the text transcription. The transcribed text is then returned to the NFON business telephony platform and made accessible to the user via an **email notification**, as well as the **telephony clients**.

Because the feature is designed to enhance accessibility, it is optimized for high accuracy across different accents and variations in spoken German. However, as with any AI-based transcription service, occasional inaccuracies may occur due to background noise, unclear speech, or complex terminology.

Model Inputs and Outputs

The Voicemail Transcriptions feature operates on a straightforward input-output basis. The **input** to the model is the voicemail **audio file**, which is transmitted securely from the NFON cloud telephony platform. The **output** is the corresponding **text transcription**, which is then displayed to the user. No additional metadata is generated by the AI system.

Data Sources for Training and Evaluation

The **Whisper** model, developed by OpenAI, was trained on a vast dataset comprising **680,000 hours of multilingual and multitask supervised data collected from the web**. This extensive dataset includes a diverse range of languages and speech patterns, enhancing the model's robustness to accents, background noise, and technical language.

In addition to utilizing Whisper, our transcription system incorporates a **Wav2Vec model**. This model has been fine-tuned using data collected in collaboration with the **ITSC (Institute for Technical Communication Systems)**. The fine-tuning process allows the model to adapt to specific speech characteristics and improve transcription accuracy for our use cases.

To ensure compliance with privacy principles, **no customer data** is used for training or fine-tuning the AI models. Instead, publicly available datasets and data collected through authorized collaborations are utilized. The models are evaluated based on metrics such as accuracy, completeness, and error handling to maintain high-quality transcription services.

Model Evaluation and Performance

NFON / botario conducts **regular assessments and evaluations** of the AI models powering Voicemail Transcriptions to ensure high performance, accuracy, and reliability. This includes **testing and quality assurance processes**, where **human oversight** plays a key role in validating model behavior, identifying potential areas for improvement, and maintaining consistency across different speech patterns and accents.

Safety and Ethical Considerations

Voicemail Transcriptions is designed to **transcribe speech as-is** without modifying, filtering, or interpreting content. The AI model does not perform sentiment analysis, content classification, or automated decision-making. The purpose of this feature is strictly **speech-to-text conversion**, ensuring that it remains a neutral and transparent tool.

The system is optimized for **fair and unbiased transcription**, minimizing potential disparities in transcription quality across different user demographics. Nevertheless, AI-based speech recognition may still exhibit minor variations in accuracy depending on pronunciation and background conditions.

NFON remains committed to **ensuring fairness and mitigating bias** in AI-generated transcriptions through ongoing model evaluation and refinement.

Privacy and Security

NFON adheres to **Privacy by Design** principles, ensuring that Voicemail Transcriptions operates in a secure and compliant manner. The **entire transcription process occurs immediately**, with no **permanent storage of audio or text on the AI system**.

All data transmissions between the NFON cloud telephony platform and the transcription service are encrypted using **TLS (HTTPS)**. The AI processing environment is **fully hosted in Germany**, with no external data transfer outside the European Economic Area (EEA).

Because the AI system does not store transcriptions or use them for further processing, **user data remains under their full control** at all times.

References

- **NFON GTC & SLA** (see Service Description Cloud Telephony)
- **NFON Privacy Policy**
- **NFON Trust Center**
- **OpenAI Whisper**