Stand: September, 2025

Beschreibung von und Anleitung für die Weiterentwicklung des Forensic WhisperDeskZH

ForMaLTeC

leo.engelberger@uzh.ch

Forensic WhisperDeskZH

Developer Manual

Contents

[1. Übersicht 2](#_Toc207199913)

[Grundidee 2](#_Toc207199914)

[Voraussetzungen 2](#_Toc207199915)

[Minimale Systemanforderungen 2](#_Toc207199916)

[Empfohlene Systemanforderungen 2](#_Toc207199917)

[2. Wichtige Begriffe & Symbole 2](#_Toc207199918)

[3. Techstack 3](#_Toc207199919)

[Grundlegende Eigenschaften: 3](#_Toc207199920)

[Nuget Packete: 3](#_Toc207199921)

[4. Grundlegende Architektur der Solution 3](#_Toc207199922)

[Überblick 3](#_Toc207199923)

[ForensicWhisperDeskZH-Projekt 3](#_Toc207199924)

[ModelDownloader-Projekt 4](#_Toc207199925)

[5. Architektur ForensicWhisperDeskZH 4](#_Toc207199926)

[Überblick 4](#_Toc207199927)

[Kernkomponenten 5](#_Toc207199928)

[Bird Eye View 6](#_Toc207199929)

[Klassenstruktur 6](#_Toc207199930)

[6. Architektur ModelDownloader 6](#_Toc207199931)

[Überblick 6](#_Toc207199932)

[Bird Eye View 6](#_Toc207199933)

[7. Flow ForensicWhisperDeskZH 6](#_Toc207199934)

[8. Flow ModelDownloader 6](#_Toc207199935)

[9. GitHubActions Pipeline 6](#_Toc207199936)

# Übersicht

## Grundidee

Der Forensic WhisperDeskZH ist ein Add-In für Microsoft Word, das die Transkription von Sprache in Text in Echtzeit ermöglicht. Es verwendet fortschrittliche KI-Modelle (Wisper), um Ihre gesprochenen Worte direkt als Text in Word-Dokumenten einzufügen.

## Voraussetzungen

### Minimale Systemanforderungen

**Betriebssystem:** Windows 10 oder höher

**Microsoft Word:** 2016 oder höher

**Arbeitspeicherpeicher:** 8GB RAM

**Speicherplatz:** 8 GB freier Speicherplatz

**GPU:**  nicht notwendig (aber empfohlen)

**Audio:** Funktionierendes Mikrofon oder Audio-Eingabegerät

### Empfohlene Systemanforderungen

**Betriebssystem:** Windows 10 oder höher

**Microsoft Word:** 2016 oder höher

**Arbeitspeicherpeicher:** 16 GB RAM

**Speicherplatz:** 16 GB freier Speicherplatz

**GPU:** Nvidia GPU mit CUDA-Kernen und 8 GB GPU Speicher

**Audio:** Funktionierendes Mikrofon oder Audio-Eingabegerät

# Wichtige Begriffe & Symbole

**VSTO:** Visual Studio Tools for Office, erlaubt die Integration von Ribbon Add-Ins direkt in Office Programme. **Es werden sämtliche Versionen von Office unterstützt**.

(Stand:August, 2025)

# Techstack

## Grundlegende Eigenschaften:

* **Framework: .Net Framework 4.8 & VSTO**
* **Programmiersprache: C# 8.0**

## Nuget Packete:

* **Whisper.net (sowie sämtliche Unterprojekte**

Whisper.net ist ein von Sandro Hanea entwickeltes Nuget package welches Platform übergreifendes Interfaces für .Net Sprachen und Frameworks zur Verfügung stellt. Im ForensicWhisperDeskZH nutzt diese Interfaces als Kern der Transkriptionsengine. Whisper.Net stellt Runtimes für Cuda, Vulkan, OpenVino und CPU Backends zur Verfügung.

* **NAudio**

NAudio ist eine .Net Audiobibliothek welche der Kern der Audioaufnahmefähigkeiten des ForensicWhisperDeskZH bildet. Das AddIn nutzt die Implementation der WASAPI(Windows Audio Session API) welche von Naudio zur Verfügung gestellt wird.

* **WebRtcVadSharp**

WebRtcVadSharp ist ein .NET-Wrapper für die WebRTC Voice Activity Detection (VAD)-Bibliothek von Google, welche innerhalb dieses AddIns zur Optimierung der Audioverarbeitungs genutzt wird.

# Grundlegende Architektur der Solution

## Überblick

Die Solution besteht hauptsächlich aus zwei Projekten, **ForensicWhisperDeskZH** (Das eigentliche AddIn) **& ModelDownloader (**Hilfsprojekt für das Builden mit Jobs). Dazukommen die **Keywords.xml** Datei und das Job Script für die Github Actions; **build-and-publish-vsto.yml**

### **ForensicWhisperDeskZH-Projekt**

Dieses Projekt ist das eigentliche VSTO AddIn. Es stellt dem User das Interface sowie die Transkriptionsfunktionalität zur Verfügung. Sämtliche Features sind hier enthalten. Zur Interaktion

### **ModelDownloader-Projekt**

Das ModelDownloader Project ist ein Utility Project und wird nicht grundsätzlich für das AddIn benötigt. Es dient dazu die Modelle automatisiert herunterladen zu können und kann lokal ausgeführt werden, ist aber hauptsächlich dazu da die WhisperAI Modelle für den Build Prozess in GitHub Actions zur Verfügung zu stellen.

# Architektur ForensicWhisperDeskZH

## Überblick

Das ForensicWhisperDeskZH-Projekt ist das Hauptprojekt der Solution und implementiert ein vollständiges VSTO (Visual Studio Tools for Office) Add-In für Microsoft Word. Es stellt eine umfassende Spracherkennungs- und Transkriptionslösung bereit, die auf OpenAI's Whisper-Technologie basiert und nahtlos in die Word-Benutzeroberfläche integriert ist.

Das Projekt folgt einer modularen Architektur mit klar getrennten Verantwortlichkeiten, die in verschiedene Namespaces organisiert sind:

* **Audio-Modul (ForensicWhisperDeskZH.Audio)**

Das Audio Modul ist zuständig für die Aufnahme und Verarbeitung von Audio. Es erlaubt das Management von Mikrofonen und der Geräte Enumeration. Ausserdem enthält dieses Modul ein Audio Diagnose Tool für Debugging Zwecke. Vor der Einführung von WASAPI wurde WaveIn Event verwendet, was zu häufigen Problemen geführt hat und das Audio Diagnostik Tool notwendig machte. Mittlerweile sollte das Audio Modul stabil genug sein ohne dieses auszukommen, vorläufig werden die Tools aber beibehalten.

* **Dokument-Modul (ForensicWhisperDeskZH.Document)**

Das Dokument Modul stellt die Word Dokument Integration dar, und ist zuständig für die Kommunikation zwischen dem AddIn und Word. Das Modul handelt das Einfügen des Textes an der Cursor Position und stellt sicher, dass das Dokument verfügbar ist.

* **Transkription-Modul (ForensicWhisperDeskZH.Transcription)**

Das Transkriptionsmodul agiert als Interface zu den WhisperModellen und ist auf Whisper.Net aufgebaut. Es erlaubt das Management und das Downloaden von Whisper Modellen

* **Text-Modul (ForensicWhisperDeskZH.Text)**

Das Textmodul ist für das Bereinigen, Formatieren und Buffern von transkribiertem Text zuständig. Dazu gehört auch, das ersetzt von Keywords mit den dazugehörigen Symbolen.

* **Common-Modul (ForensicWhisperDeskZH.Common)**

Das Common Modul ist für das Konfigurations-Management über JSON und XML, den Logging-Service sowie weiter Hilfs- und Utilities-Aufgaben zuständig.

## Kernkomponenten

**AddInViewModel**: Die zentrale Koordinationsschicht des Add-Ins, die als MVVM-Pattern implementiert ist. Orchestriert die Interaktion zwischen Transkription, Dokumentbearbeitung und UI. Unterstützt sowohl kontinuierliche Transkription als auch einen "Listening Mode" mit Trigger-Phrasen ("diktat starten"/"diktat beenden").

**TranscriptionSettings:** Umfassende Konfigurationsklasse für alle Aspekte der Spracherkennung, einschließlich Audio-Parameter, Whisper-Model-Einstellungen, Sprach-Optionen und Text-Verarbeitungsregeln.

**MainRibbon:** Word-Ribbon-Integration für die Benutzeroberfläche mit Buttons für Start/Stop, Einstellungen, Mikrofonauswahl und Model-Konfiguration.

Bird Eye View

## Klassenstruktur

# Architektur ModelDownloader

## Überblick

## Bird Eye View

# Flow ForensicWhisperDeskZH

# Flow ModelDownloader

# GitHubActions Pipeline