Abstract

Beschreibung und Anleitung für die Nutzung des Forensic WhisperDeskZH

ForMaLTeC

leo.engelberger@uzh.ch

Forensic WhisperDeskZH

Developer Manual

Contents

[1. Übersicht 2](#_Toc207105142)

[Wichtigste Merkmale: 2](#_Toc207105143)

[2. Voraussetzungen 2](#_Toc207105144)

[Minimale Systemanforderungen 2](#_Toc207105145)

[Empfohlene Systemanforderungen 2](#_Toc207105146)

[3. Wichtige Begriffe & Symbole 3](#_Toc207105147)

[4. Architektur 4](#_Toc207105148)

[5. Bird Eye View 4](#_Toc207105149)

[6. Klassenstruktur 4](#_Toc207105150)

[7. Flow 4](#_Toc207105151)

# Übersicht

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 (Whisper), um Ihre gesprochenen Worte direkt als Text in Word-Dokumenten einzufügen.

## Wichtigste Merkmale:

* **Echtzeit-Transkription** direkt in Word Dokumente
* **Mehrsprachige Unterstützung**
* **Verschiedene KI-Modelloptionen** für optimale Leistung

# 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

**Diktatmodus:** Im Diktatmodus, transkribiert das Add-in sämtlcihe vom Mikrofon aufgenommene Sprache und fügt den transkribierten Text direkt in das geöffnete Word Dokument an der Stelle des Cursors ein.

**Zuhörmodus:**  Der Zuhörmodus bezeichnet den Modus, in welchem das Add-in bereits mithört, den transkribierten Text aber noch nicht in das geöffnete Word Dokument einfügt. Erst wenn das Add-in Durch den Sprachbefehl «Diktat Starten» in den Diktatmodus versetzt wird, wird Text im Dokument eingefügt.

**Minimal Audio Chunk Size:** Für die Echtzeit Transkription, muss die Audioaufnahme kurz unterbrochen und ausgelesen werden. Die Minimal Audio Chunk Size bestimmt die minimale Zeitspanne in Sekunden, welche vergehen muss, dass das Add-in eine Transkription startet. Kürzere Werte führen zu häufigerer und schneller Transkription ins Word Dokument, im Gegenzug sinkt aber die Genauigkeit. Mit grösseren Werten wird die Transkription weniger häufig aktualisiert, die Genauigkeit nimmt aber zu.

**Silence Threshold:** Um die Genauigkeit der Transkription zu erhöhen versucht das Add-in Pausen zwischen Wörtern und Sätzen zu erkennen. Dies soll verhindern, das Worte abgeschnitten und somit fehlerhaft transkribiert werden. Der Silence Threshold bestimmt, wie lange es Still sein muss, dass das Plugin von einer Pause zwischen Worten ausgeht.

# Architektur

## Overview

## Bird Eye View

# Klassenstruktur

# Flow