

Wake-Word Testing Framework

Systematisches Testen des “Computer” Wake-Words

Inhaltsverzeichnis

1. [Test-Übersicht](#)
 2. [Erkennungsrate-Tests](#)
 3. [Falsch-Positiv-Tests](#)
 4. [Umgebungs-Tests](#)
 5. [Stress-Tests](#)
 6. [Performance-Tests](#)
 7. [Erfolgs-Kriterien](#)
 8. [Test-Protokoll](#)
-

Test-Übersicht

Dieses Dokument beschreibt ein systematisches Testing-Framework für das “Computer” Wake-Word. Alle Tests sollten durchgeführt werden, bevor das Wake-Word als produktionsreif betrachtet wird.

Warum systematisches Testing?

Ein Wake-Word-System muss in verschiedenen Szenarien zuverlässig funktionieren. Systematisches Testing stellt sicher, dass das System nicht nur unter idealen Bedingungen funktioniert, sondern auch in realen Nutzungsszenarien mit Hintergrundgeräuschen, verschiedenen Sprechern und unterschiedlichen Umgebungen.

Test-Kategorien

Die Tests sind in fünf Hauptkategorien unterteilt:

Erkennungsrate-Tests messen, wie zuverlässig das Wake-Word erkannt wird, wenn es tatsächlich gesprochen wurde. Diese Tests sind fundamental, da eine niedrige Erkennungsrate die Nutzbarkeit des Systems stark beeinträchtigt.

Falsch-Positiv-Tests messen, wie oft das System fälschlicherweise aktiviert wird, wenn das Wake-Word nicht gesprochen wurde. Zu viele Falsch-Positive führen zu Frustration und reduzieren das Vertrauen in das System.

Umgebungs-Tests stellen sicher, dass das System in verschiedenen akustischen Umgebungen funktioniert, von leisen Büros bis zu lauten öffentlichen Räumen.

Stress-Tests prüfen die Langzeit-Stabilität und Robustheit des Systems unter kontinuierlicher Nutzung.

Performance-Tests messen technische Metriken wie Latenz, CPU-Auslastung und Memory-Verbrauch.

Erkennungsrate-Tests

Test 1.1: Basis-Erkennungsrate

Ziel: Messen der grundlegenden Erkennungsrate unter idealen Bedingungen

Setup:

- Leise Umgebung (Büro oder Zimmer)

- Normale Sprechlautstärke
- Abstand zum Mikrofon: 1-2 Meter
- Keine Hintergrundgeräusche

Durchführung:

1. Starte Voice Assistant
2. Sage 100x "Computer" in normalem Tempo
3. Zähle erfolgreiche Erkennungen
4. Dokumentiere Fehlerkennungen

Checkliste:

Test 1.1: Basis-Erkennungsrate

Datum: _____ Tester: _____

Versuch	Erkannt	Notizen
1-10	___ /10	
11-20	___ /10	
21-30	___ /10	
31-40	___ /10	
41-50	___ /10	
51-60	___ /10	
61-70	___ /10	
71-80	___ /10	
81-90	___ /10	
91-100	___ /10	
GESAMT: ___ /100 (___%%)		
Erfolg: [] Ja (≥95%) [] Nein (<95%)		

Erwartetes Ergebnis: ≥95% Erkennungsrate

Test 1.2: Verschiedene Tonlagen

Ziel: Testen der Erkennungsrate bei verschiedenen Tonhöhen

Setup: Wie Test 1.1

Durchführung:

1. 20x normale Stimme
2. 20x höhere Tonlage
3. 20x tiefere Tonlage
4. 20x geflüstert
5. 20x laut

Checkliste:

Test 1.2: Verschiedene Tonlagen

Datum: _____ Tester: _____

Tonlage	Versuche	Erkannt	Rate
-----	-----	-----	-----
Normal	20	__	__%
<i>Hoch</i>	20	__	__%
<i>Tief</i>	20	__	__%
<i>Geflüstert</i>	20	__	__%
<i>Laut</i>	20	__	__%

GESAMT: __ /100 (__%)

Erfolg: [] Ja ($\geq 90\%$) [] Nein ($< 90\%$)

Erwartetes Ergebnis: $\geq 90\%$ Erkennungsrate (etwas niedriger als Basis)

Test 1.3: Verschiedene Geschwindigkeiten

Ziel: Testen der Erkennungsrate bei verschiedenen Sprechgeschwindigkeiten

Durchführung:

1. 30x normale Geschwindigkeit
2. 30x schnell ("Compu-ter")
3. 30x langsam ("Com...pu...ter")

Checkliste:

Test 1.3: Verschiedene Geschwindigkeiten

Datum: _____ Tester: _____

Geschwindigkeit	Versuche	Erkannt	Rate
Normal	30	__	__%
Schnell	/ 30	/ __	 __%
<i>Langsam</i>	/ 30	/ __	__%

GESAMT: __ / 90 (__%)

Erfolg: [] Ja ($\geq 90\%$) [] Nein ($< 90\%$)

Erwartetes Ergebnis: $\geq 90\%$ Erkennungsrate

Test 1.4: Verschiedene Abstände

Ziel: Testen der Erkennungsrate bei verschiedenen Entfernungen zum Mikrofon

Durchführung:

1. 20x aus 0.5m Entfernung
2. 20x aus 1m Entfernung
3. 20x aus 2m Entfernung
4. 20x aus 3m Entfernung
5. 20x aus 4m Entfernung

Checkliste:

Test 1.4: Verschiedene Abstände

Datum: _____ Tester: _____

Abstand	Versuche	Erkannt	Rate
-----	-----	-----	-----
0.5m	20	__	__%
1.0m	20	__	__%
2.0m	20	__	__%
3.0m	20	__	__%
4.0m	20	__	__%

GESAMT: __ /100 (__%%)

Erfolg: [] Ja ($\geq 80\%$ bei 1-3m) [] Nein

Erwartetes Ergebnis: $\geq 95\%$ bei 1-2m, $\geq 80\%$ bei 3m

Falsch-Positiv-Tests

Test 2.1: Ähnliche Wörter

Ziel: Testen, ob ähnlich klingende Wörter fälschlicherweise erkannt werden

Durchführung: Sage jedes Wort 10x und zähle Fehlerkennungen

Checkliste:

Test 2.1: Ähnliche Wörter

Datum: _____ Tester: _____

Wort	Versuche	Fehl-Erkant	Rate
-----	-----	-----	-----
Commuter	10	__	__%
Komputer	10	__	__%
Puter	10	__	__%
Commute	10	__	__%
Compute	10	__	__%
Kompute	10	__	__%
Rechner	10	__	__%
PC	10	__	__%
Laptop	10	__	__%
Tablet	10	__	__%

GESAMT: ____ Fehl-Erkennungen von 100 Versuchen (____%%)

Erfolg: [] Ja (<5% Fehl-Erkennungen) [] Nein (≥5%)

Erwartetes Ergebnis: % Falsch-Positive

Test 2.2: Normale Konversation

Ziel: Testen, ob normale Gespräche Fehlerkennungen auslösen

Setup:

- Führe normale Gespräche in der Nähe des Mikrofons
- Vermeide absichtlich das Wort “Computer”
- Dauer: 1 Stunde

Durchführung:

1. Starte Voice Assistant
2. Führe normale Gespräche (oder spiele Podcast/Video ab)
3. Zähle Fehlerkennungen

Checkliste:

Test 2.2: Normale Konversation

Datum: _____ Tester: _____

Zeit (Minuten)	Fehl-Erkennungen	Notizen
----------------	------------------	---------

-----	-----	-----
-------	-------	-------

0-10		
------	--	--

0-10		
------	--	--

10-20		
-------	--	--

10-20		
-------	--	--

20-30		
-------	--	--

20-30		
-------	--	--

30-40		
-------	--	--

30-40		
-------	--	--

40-50		
-------	--	--

40-50		
-------	--	--

50-60		
-------	--	--

50-60		
-------	--	--

GESAMT: ____ Fehl-Erkennungen in 60 Minuten

Erfolg: [] Ja (<1 pro Stunde) [] Nein (≥1 pro Stunde)

Erwartetes Ergebnis: Fehl-Erkennung pro Stunde

Test 2.3: Hintergrundgeräusche

Ziel: Testen, ob Hintergrundgeräusche Fehlerkennungen auslösen

Durchführung: Teste mit verschiedenen Hintergrundgeräuschen (je 10 Minuten)

Checkliste:

Test 2.3: Hintergrundgeräusche

Datum: _____ Tester: _____

Geräusch-Typ	Dauer	Fehl-Erkennungen	Rate/h
-----	-----	-----	-----
Musik (leise)	10min	__	__
Musik (laut)	10min	__	__
TV/Radio	10min	__	__
Straßenlärm	10min	__	__
Tippen/Klicken	10min	__	__
Ventilator	10min	__	__

GESAMT: __ Fehl-Erkennungen in 60 Minuten

Erfolg: [] Ja (<2 pro Stunde) [] Nein (≥2 pro Stunde)

Erwartetes Ergebnis: Fehl-Erkennungen pro Stunde

Umgebungs-Tests

Test 3.1: Verschiedene Räume

Ziel: Testen in verschiedenen akustischen Umgebungen

Checkliste:

Test 3.1: Verschiedene Räume

Datum: _____ Tester: _____

Raum-Typ	Versuche	Erkannt	Rate	Notizen
-----	-----	-----	-----	-----
Kleines Zimmer	20	__	__% /	
Großes Zimmer	20	__	__% /	
Büro	20	__	__% /	
Küche	20	__	__% /	
Badezimmer (Hall)	20	__	__% /	
Draußen	20	__	__% /	

GESAMT: __ /120 (__%%)

Erfolg: [] Ja ($\geq 85\%$ in allen Räumen) [] Nein

Erwartetes Ergebnis: $\geq 85\%$ in allen Räumen

Test 3.2: Verschiedene Lautstärke-Umgebungen

Ziel: Testen bei verschiedenen Umgebungslautstärken

Checkliste:

Test 3.2: Verschiedene Lautstärke-Umgebungen

Datum: _____ Tester: _____

Umgebung	Versuche	Erkannt	Rate	Fehl-Pos
-----	-----	-----	-----	-----
Sehr leise (<30dB)	20	__	__% / __	
Leise (30-40dB)	20	__	__% / __	
Normal (40-50dB)	20	__	__% / __	
Laut (50-60dB)	20	__	__% / __	
Sehr laut (>60dB)	20	__	__% / __	

GESAMT: __ /100 (__%%)

Erfolg: [] Ja ($\geq 80\%$ bei 30-60dB) [] Nein

Erwartetes Ergebnis: $\geq 80\%$ bei normalen Lautstärken (30-60dB)

Stress-Tests

Test 4.1: Schnelle Wiederholungen

Ziel: Testen bei schnellen aufeinanderfolgenden Aktivierungen

Durchführung: Sage "Computer" 100x hintereinander mit nur 2 Sekunden Pause

Checkliste:

Test 4.1: Schnelle Wiederholungen

Datum: _____ Tester: _____

Versuch	Erkannt	Doppel-Erkennung	Verzögerung
-----	-----	-----	-----
1-10	__ /10	__	[] Ja/Nein
11-20	__ /10	__	[] Ja/Nein
21-30	__ /10	__	[] Ja/Nein
31-40	__ /10	__	[] Ja/Nein
41-50	__ /10	__	[] Ja/Nein
51-60	__ /10	__	[] Ja/Nein
61-70	__ /10	__	[] Ja/Nein
71-80	__ /10	__	[] Ja/Nein
81-90	__ /10	__	[] Ja/Nein
91-100	__ /10	__	[] Ja/Nein

GESAMT: __ /100 (__%%)

Doppel-Erkennungen: __

Erfolg: [] Ja ($\geq 90\%$, keine Doppel-Erkennungen) [] Nein

Erwartetes Ergebnis: $\geq 90\%$, keine Doppel-Erkennungen

Test 4.2: Langzeit-Stabilität

Ziel: Testen der Stabilität über längere Zeit

Durchführung: Lasse Voice Assistant 1 Stunde laufen und aktiviere alle 5 Minuten

Checkliste:

Test 4.2: Langzeit-Stabilität

Datum: _____ Tester: _____

Zeit (Min)	Aktivierung	Erkannt	CPU%	Memory(MB)
-----	-----	-----	-----	-----
5	Computer	[]	__	__
10	Computer	[]	__	__
15	Computer	[]	__	__
20	Computer	[]	__	__
25	Computer	[]	__	__
30	Computer	[]	__	__
35	Computer	[]	__	__
40	Computer	[]	__	__
45	Computer	[]	__	__
50	Computer	[]	__	__
55	Computer	[]	__	__
60	Computer	[]	__	__

GESAMT: __ / 12 (__ %)

Durchschn. CPU: __ %

Durchschn. Memory: __ MB

Erfolg: [] Ja (≥95%, stabil) [] Nein

Erwartetes Ergebnis: ≥95%, keine Memory Leaks, CPU stabil

Performance-Tests

Test 5.1: Latenz-Messung

Ziel: Messen der Zeit von Wort-Ende bis Erkennung

Checkliste:

Test 5.1: Latenz-Messung

Datum: _____ Tester: _____

Versuch	Latenz (ms)	Notizen
-----	-----	-----
1	_	
2	_	
3	_	
4	_	
5	_	
6	_	
7	_	
8	_	
9	_	
10	_	

Durchschnitt: ____ ms

Min: ____ ms

Max: ____ ms

Erfolg: [] Ja (<100ms Durchschnitt) [] Nein

Erwartetes Ergebnis: <100ms Durchschnitt, <200ms Maximum

Test 5.2: Ressourcen-Verbrauch

Ziel: Messen von CPU und Memory Auslastung

Checkliste:

Test 5.2: Ressourcen-Verbrauch

Datum: _____ Tester: _____

Zustand	CPU%	Memory(MB)	Notizen
-----	-----	-----	-----
Idle (wartend)	__	__	
Wake-Word erkannt	__	__	
STT aktiv	__	__	
TTS aktiv	__	__	
Nach 10min	__	__	
Nach 30min	__	__	
Nach 60min	__	__	

Durchschn. CPU: __%

Durchschn. Memory: __MB

Erfolg: [] Ja (<20% CPU, <200MB Memory) [] Nein

Erwartetes Ergebnis: <20% CPU, <200MB Memory

Erfolgs-Kriterien

Minimale Anforderungen

Ein Wake-Word-System gilt als produktionsreif, wenn folgende Kriterien erfüllt sind:

ERFOLGS-KRITERIEN CHECKLISTE

Datum: _____ Tester: _____

Kategorie	Kriterium	Status
-----	-----	-----
Erkennungsrate	≥95%	[]
Erkennungsrate (Variationen)	≥90%	[]
Falsch-Positive (ähnlich)	<5%	[]
Falsch-Positive (Konv.)	<1/Stunde	[]
Umgebungs-Robustheit	≥85%	[]
Keine Doppel-Erkennungen	0 von 100	[]
Langzeit-Stabilität	≥95%	[]
Latenz	<100ms	[]
CPU-Auslastung	<20%	[]
Memory-Auslastung	<200MB	[]

GESAMT: __ /10 Kriterien erfüllt

PRODUKTIONSREIF: [] Ja (10/10) [] Nein (<10/10)

Empfohlene Kriterien (Optional)

Für exzellente Performance:

ERWEITERTE KRITERIEN (OPTIONAL)

Datum: _____ Tester: _____

Kategorie	Kriterium	Status
-----	-----	-----
Erkennungsrate	≥98%	[]
Falsch-Positive (Konv.)	<0.5/Stunde	[]
Latenz	<50ms	[]
CPU-Auslastung	<10%	[]
Memory-Auslastung	<100MB	[]
Funktioniert bis 4m Abstand	≥80%	[]
Funktioniert in allen Räumen	≥90%	[]

GESAMT: __ /7 erweiterte Kriterien erfüllt

EXZELLENT: [] Ja (≥5/7) [] Nein (<5/7)

Test-Protokoll

Vollständiges Test-Protokoll

WAKE-WORD TEST-PROTOKOLL

=====

Projekt: Computer Voice Assistant

Wake-Word: "Computer"

Methode: ☐ Porcupine ☐ OpenWakeWord

TESTER-INFORMATION

Name: _____

Datum: _____

Uhrzeit: _____

System: _____

Mikrofon: _____

KONFIGURATION

Sensitivity: _____

Cooldown: _____ Sekunden

Sample Rate: _____ Hz

Threshold: _____ (falls OpenWakeWord)

TEST-ERGEBNISSE

1. Erkennungsrate-Tests

- ☐ Test 1.1: Basis (___%)
- ☐ Test 1.2: Tonlagen (___%)
- ☐ Test 1.3: Geschwindigkeit (___%)
- ☐ Test 1.4: Abstände (___%)

2. Falsch-Positiv-Tests

- ☐ Test 2.1: Ähnliche Wörter (___%)
- ☐ Test 2.2: Konversation (___ /Stunde)
- ☐ Test 2.3: Hintergrund (___ /Stunde)

3. Umgebungs-Tests

- ☐ Test 3.1: Räume (___%)
- ☐ Test 3.2: Lautstärken (___%)

4. Stress-Tests

☐ Test 4.1: Wiederholungen (___%%)

☐ Test 4.2: *Langzeit* (___%%)

5. Performance-Tests

☐ Test 5.1: Latenz (___ms)

☐ Test 5.2: *Ressourcen* (___% CPU, ___MB)

GESAMTBEWERTUNG

Produktionsreif: ☐ Ja ☐ Nein

Empfehlung: ☐ Deployment ☐ Weitere Optimierung

NOTIZEN & BEOBACHTUNGEN

NÄCHSTE SCHRITTE

Unterschrift: _____

Dokumentende

Seite {page}