

**KÄRNTEN**  
University of  
Applied Sciences

# VST-VOICEOVER

Präsentation von Vadim Steshkov

Projekt unterstützt von Prof. Menard Christian &  
Prof. Klinger Thomas

# EINLEITUNG

DAS PROJEKT ZIELT DARAUF  
AB,



einen Workflow für die Erstellung von Lehrvideos mit mehrsprachigen Untertiteln und Vertonung – zur Nutzung in der Software Subtitle Edit – zu entwickeln. Dieser Workflow ermöglicht es, Schritt für Schritt Untertitel und Vertonung präzise und leicht verständlich zu erstellen, unterstützt durch verschiedene Modelle, APIs und Konfigurationen.



# SCHRITTE ZUR VIDEOPRODUKTION

Bevor ein Lehrvideo erstellt wird, sollten alle Schritte des Workflows klar definiert sein. So wird ein reibungsloser Ablauf und hohe Qualität sichergestellt.



Das Video präzise in Text umwandeln. So wird der Inhalt optimal erfasst und für die weiteren Arbeitsschritte vorbereitet.



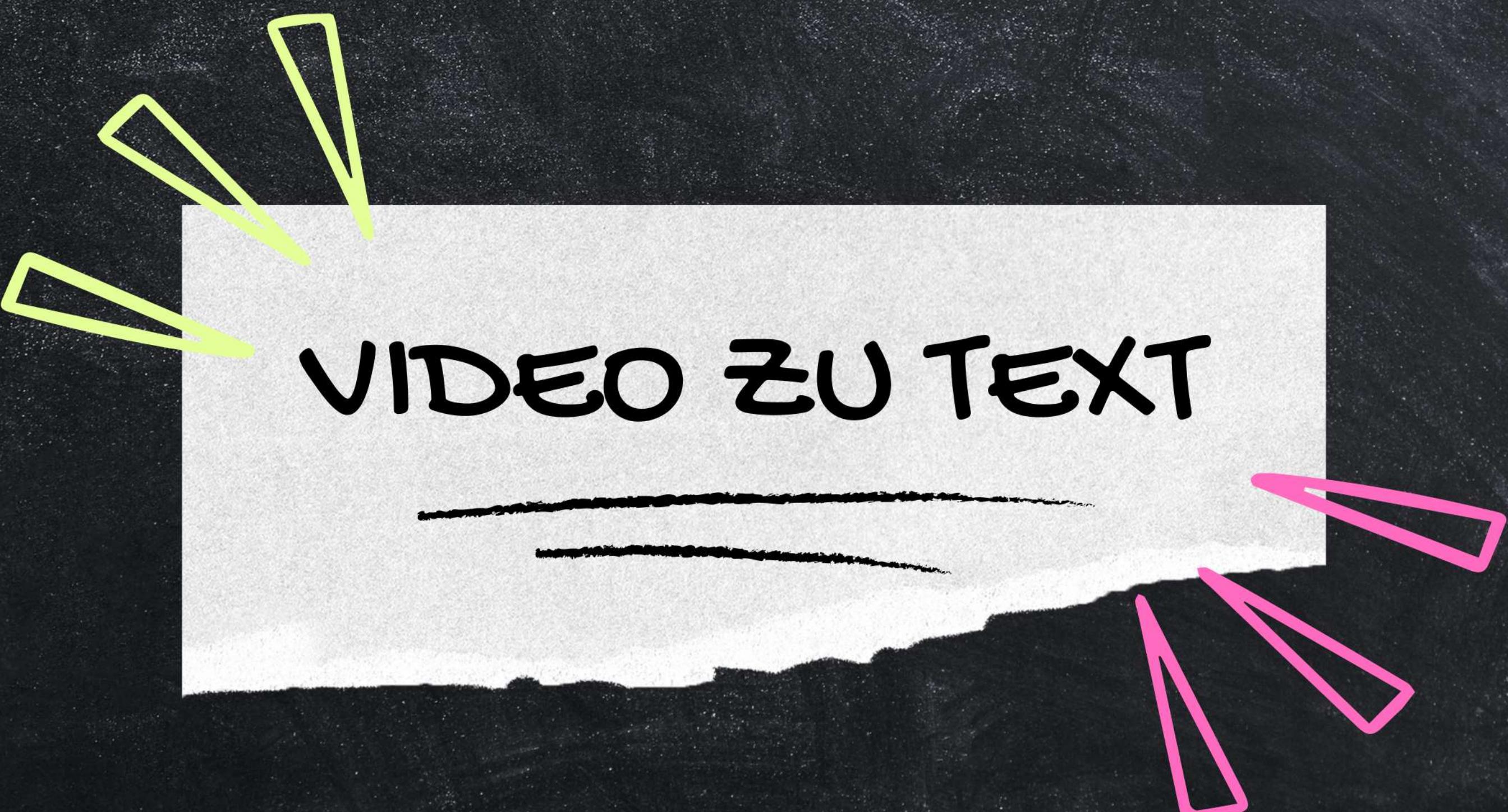
Den Text so genau wie möglich in die gewünschte Sprache übersetzen. Dabei die besten Modelle und APIs für präzise Übersetzungen verwenden.



Eine natürliche Vertonung in der Zielsprache hinzufügen. Dieser Schritt sorgt für ein angenehmes und einprägsames Lernerlebnis.



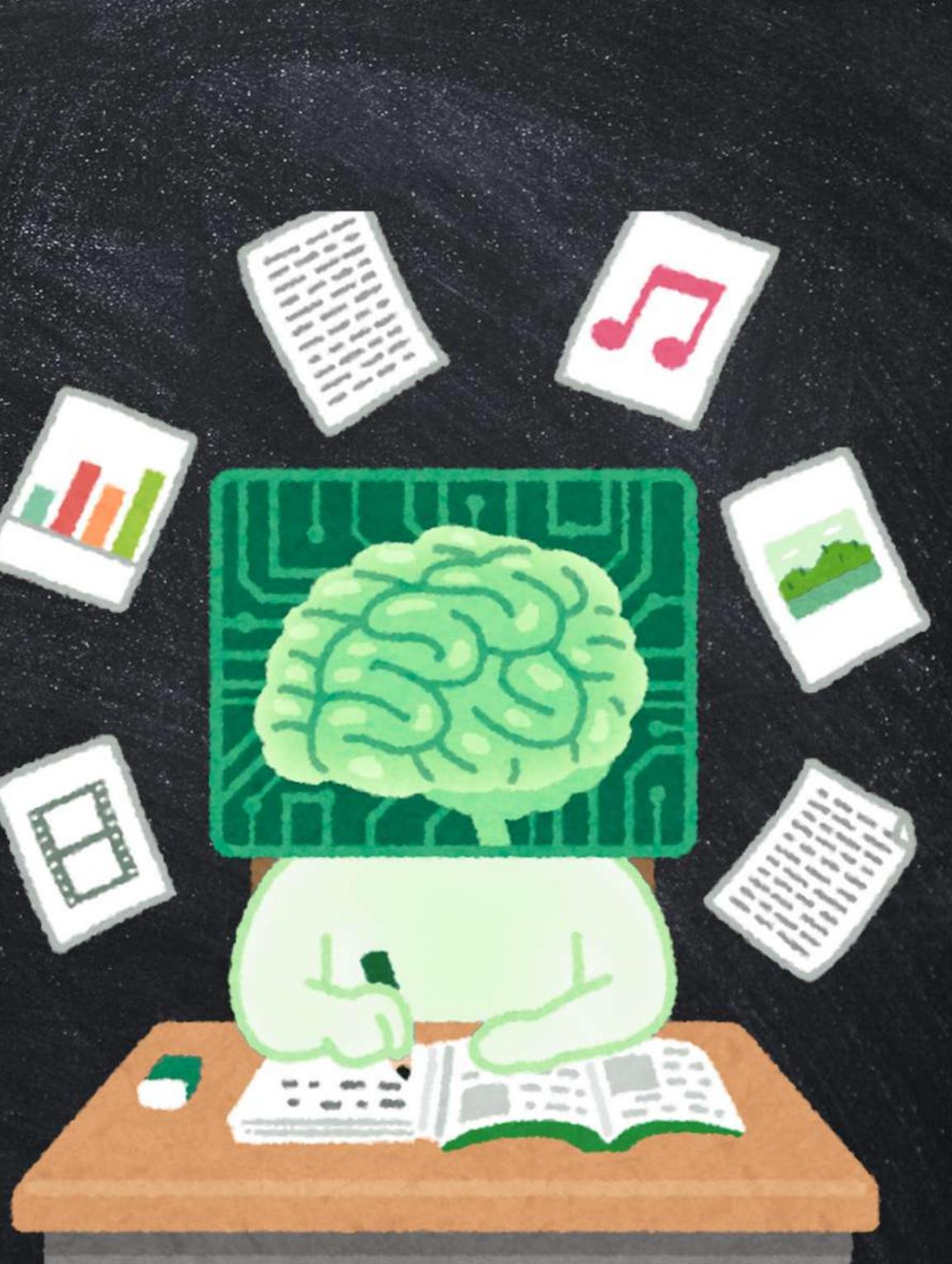
Den übersetzten Text als Untertitel in das Video einfügen. Dies macht das Video barrierefrei und fördert das Verständnis für alle Zuschauer.

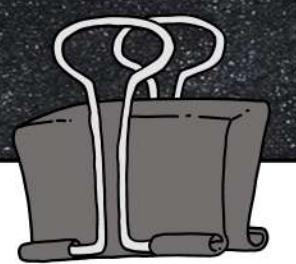


**VIDEO ZU TEXT**

# Methodik

Für die Umwandlung von Video in Text in Subtitle Edit stehen verschiedene Modelle zur Verfügung: von tiny, base, small, medium bis hin zu large\_v1, large\_v2, large\_v3 und large-turbo-v3. Diese Modelle unterscheiden sich in ihrer Genauigkeit, Geschwindigkeit und natürlichen Satzstruktur. Durch die Auswahl des passenden Modells kann der Workflow gezielt an die jeweiligen Anforderungen angepasst werden – von schnellen und einfachen Transkriptionen bis hin zu besonders präzisen und natürlich klingenden Ergebnissen.





# Detaillierte Anpassungen



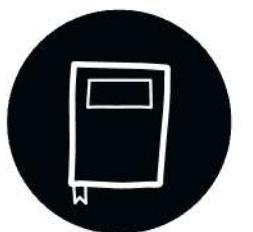
## Zusätzliche Parameter

- Feineinstellungen wie „Wörter pro Sekunde“ oder „Framerate“ verbessern die Erkennung und Lesbarkeit des Textes.
- Durch präzise Anpassung dieser Parameter kann die Qualität der automatischen Transkription erheblich gesteigert werden.



## Manuelle Anpassungen

- Trotz Automatisierung ist manuelles Editieren wichtig: zum Beispiel das Splitten oder Mergen von Sätzen für mehr Präzision.
- Dies ermöglicht eine individuellere Anpassung an den tatsächlichen Sprachfluss und erhöht die Verständlichkeit.



## Externe Skripte

- Mithilfe externer Skripte können erweiterte Analysen durchgeführt und die Genauigkeit der Erkennung überprüft werden. Sie bieten zusätzliche Flexibilität und erweitern die Möglichkeiten der Nachbearbeitung und Qualitätskontrolle.



# BEWERTUNGSKRITERIEN

GENAUIGKEIT

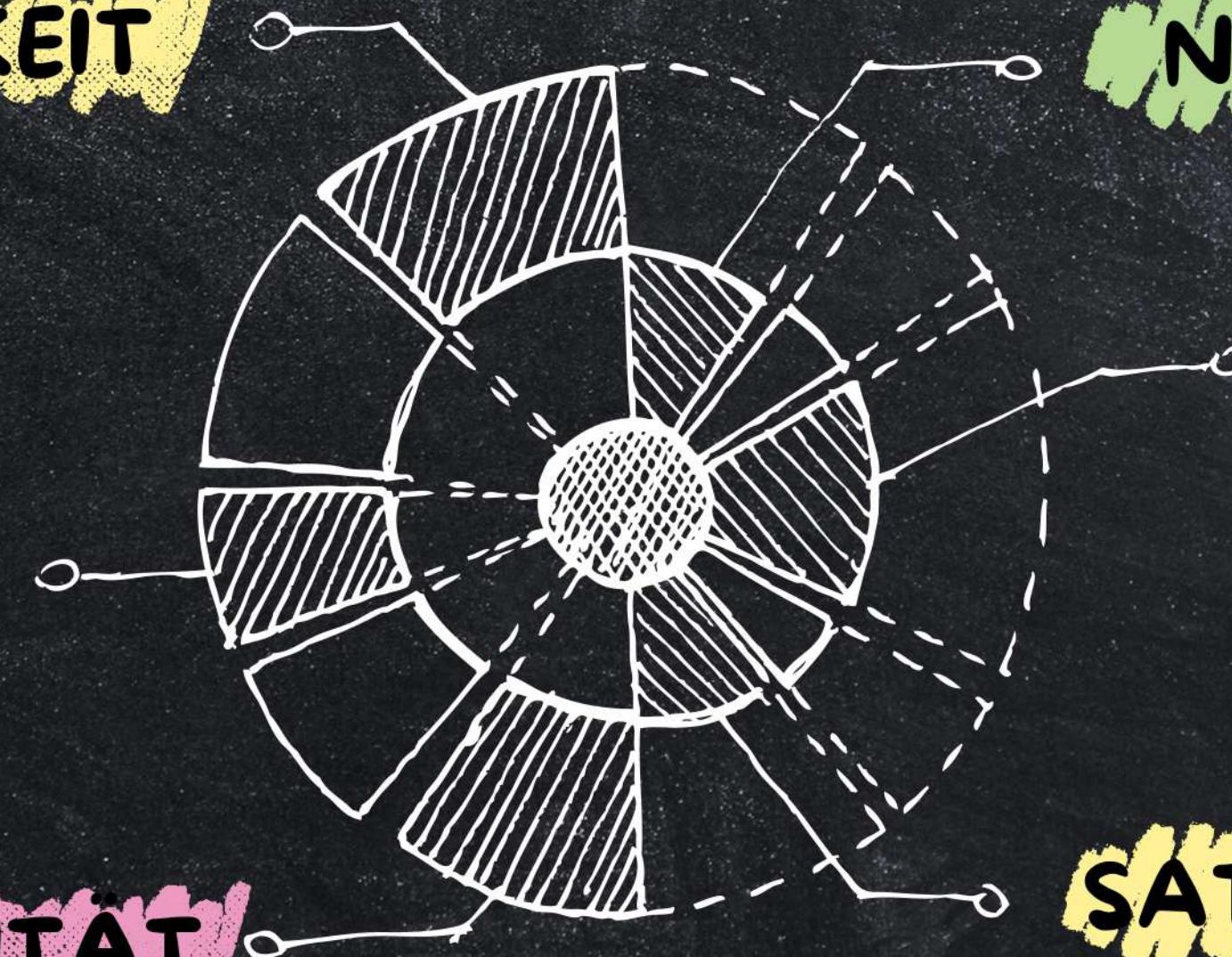
NATÜRLICHKEIT

SYNCHRONISATION

SATZBAU

ÜBERSETZUNGSQUALITÄT

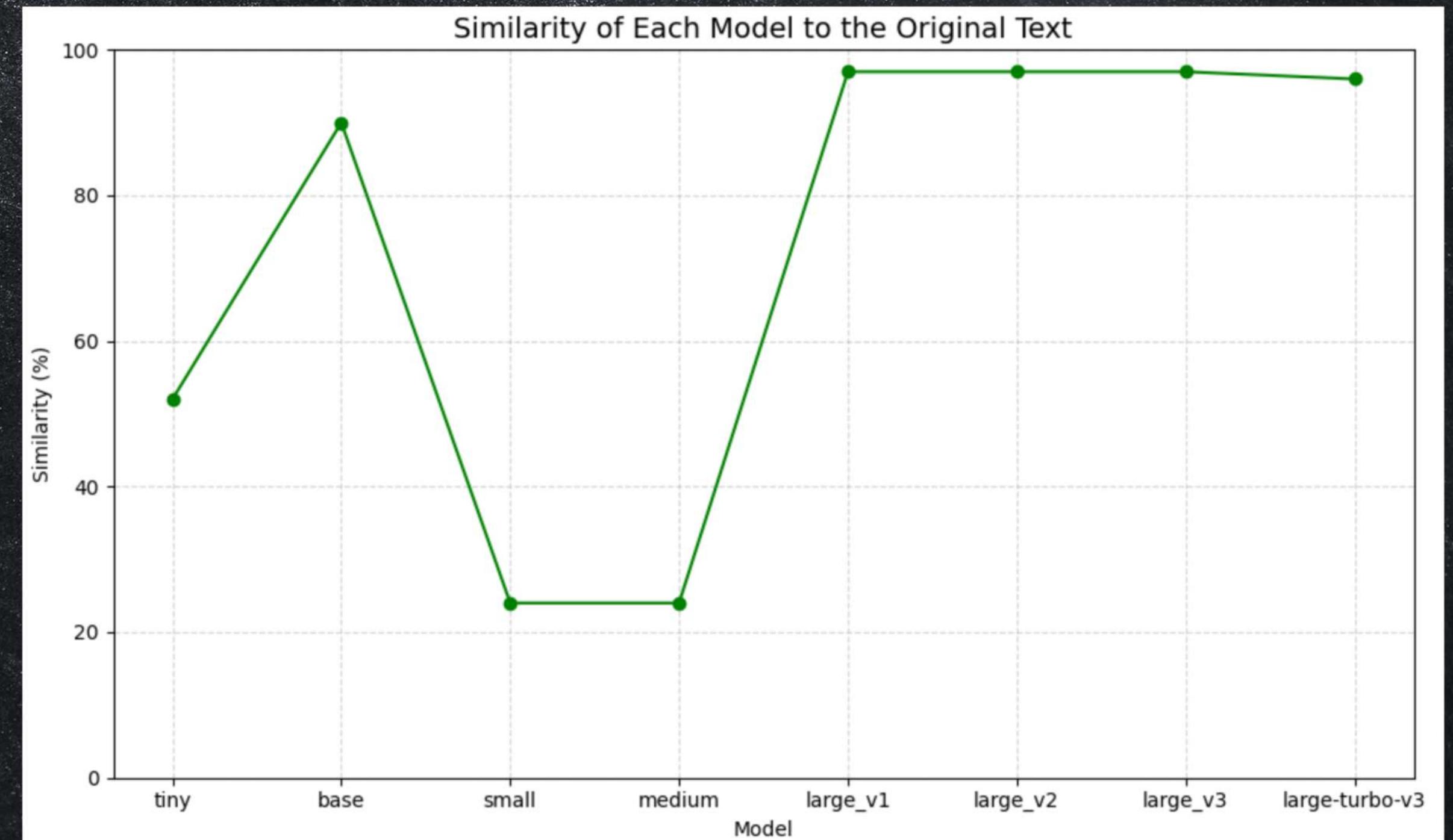
SATZTRENNUNG



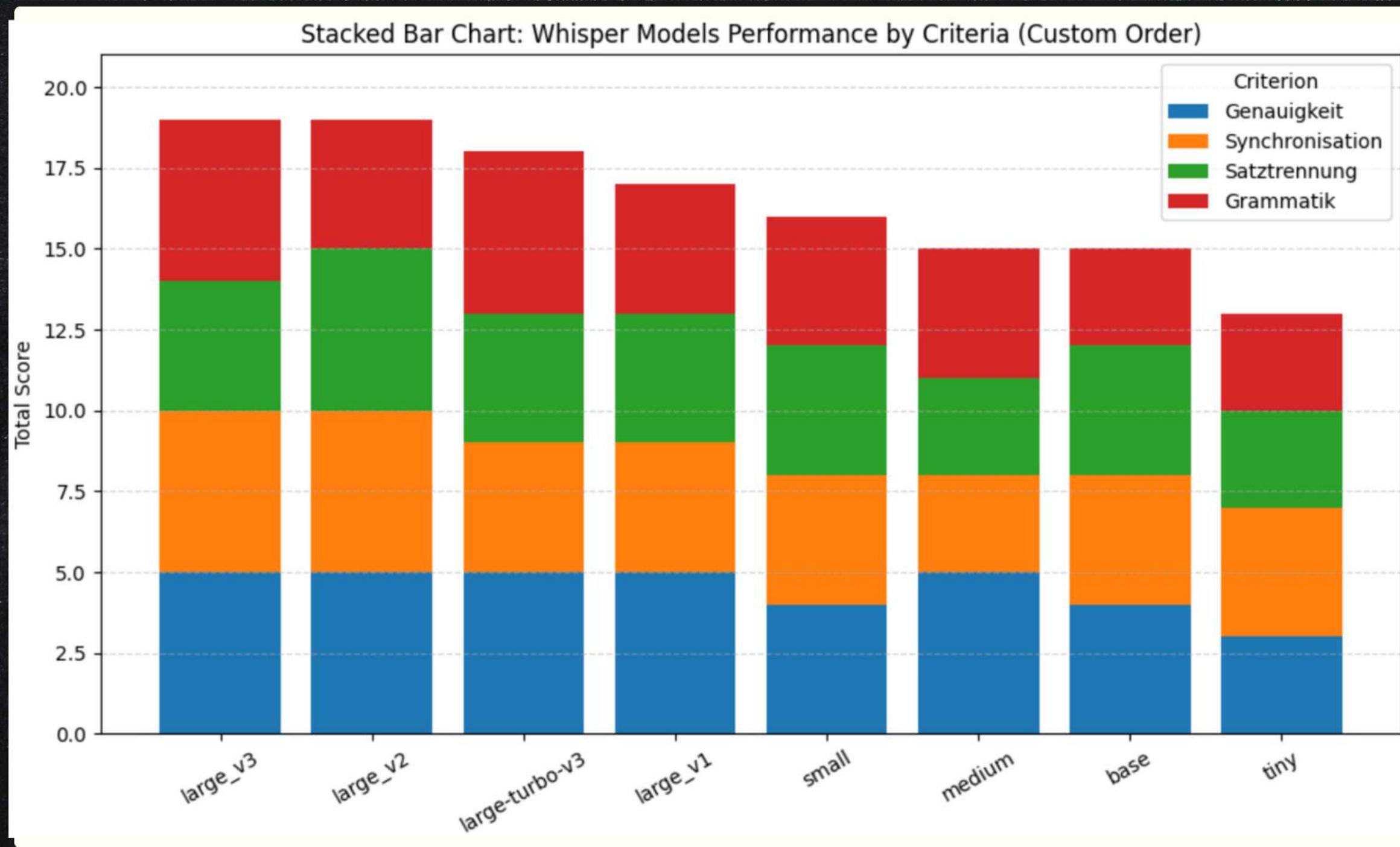
G

## VERGLEICHENDE ANALYSE DER ÄHNLICHKEIT

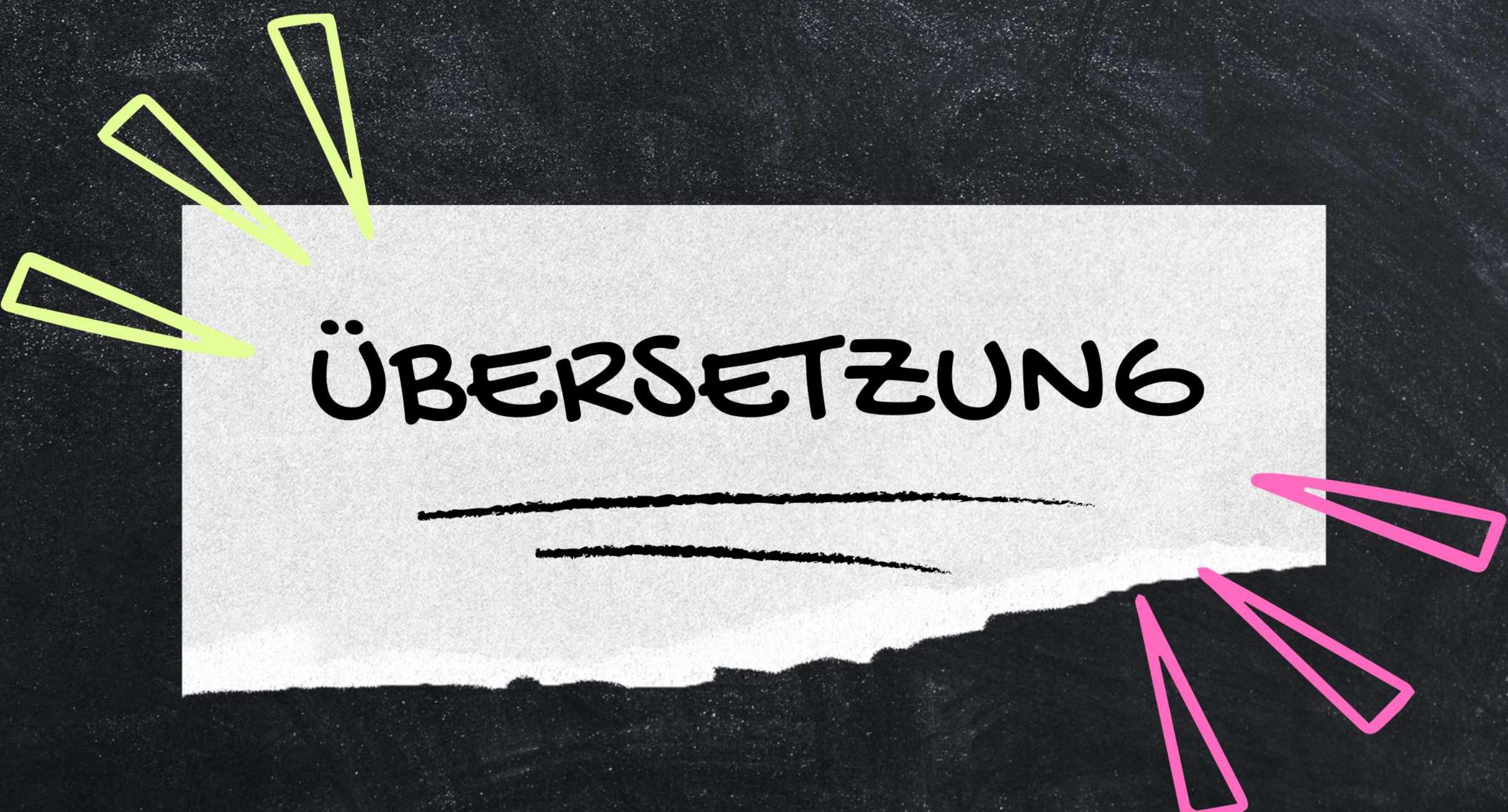
Das Diagramm zeigt die Ähnlichkeit der Modelle zum Originaltext. Die Modelle large\_v1, large\_v2 und large\_v3 erreichen dabei 97%, während small und medium nur etwa 24% erzielen. Insgesamt verdeutlichen Matrix und Diagramm die hohe Stabilität großer Modelle.



# TESTERGEBNISSE DER AUTOMATISCHEN UNTERTITELUNG



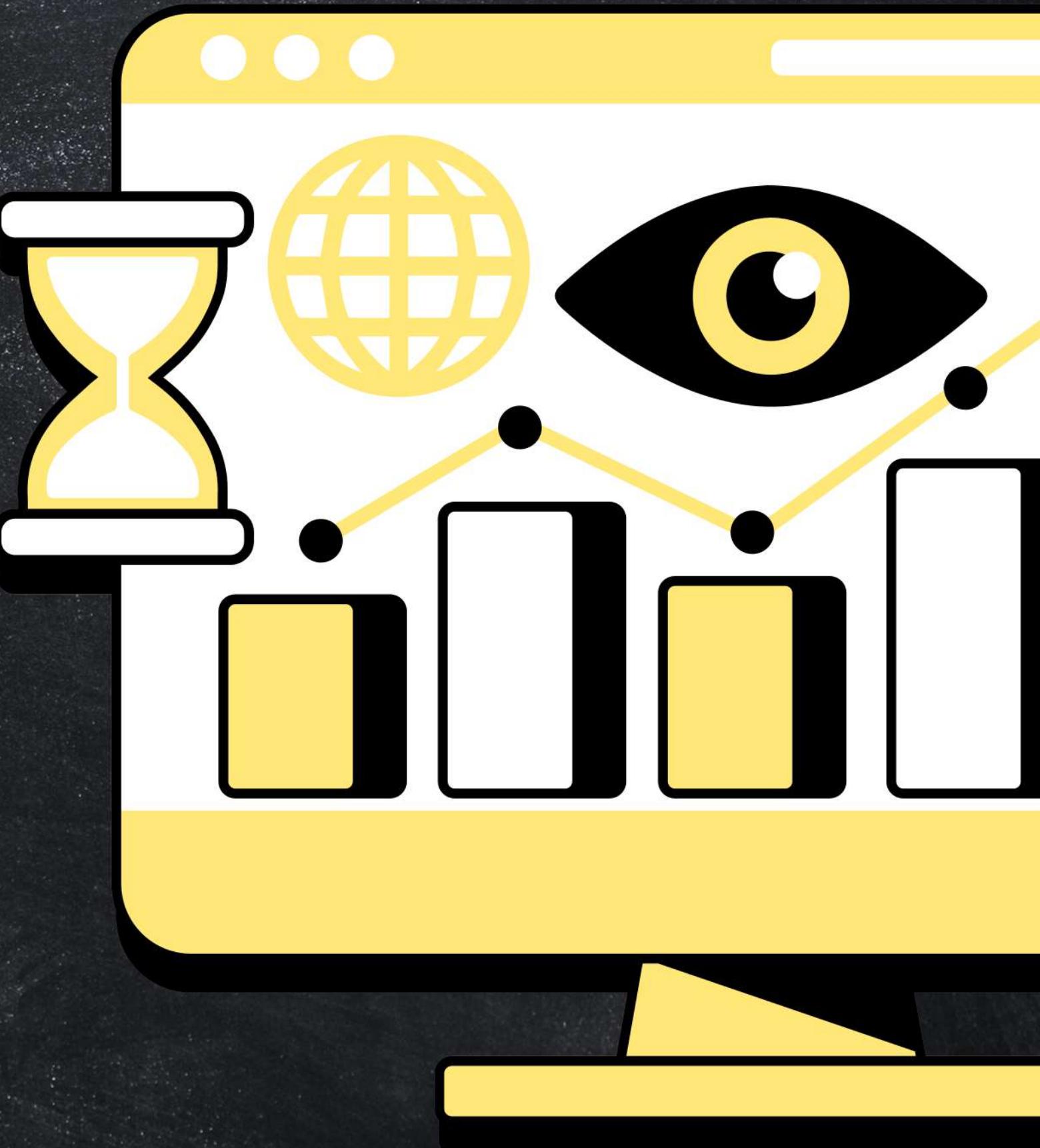
Die Balkendiagramme zeigen die Bewertung der Modelle in vier Kategorien: Genauigkeit, Synchronisation, Satztrennung und Grammatik. Große Modelle wie large\_v1, large\_v2, large\_v3 und large-turbo-v3 erzielen durchgängig Höchstwerte von 4 bis 5 Punkten, während kleinere Modelle (tiny, base) bei etwa 3 Punkten liegen. Insgesamt zeigen die Ergebnisse die klare Überlegenheit der großen Modelle in allen Kategorien.



**ÜBERSETZUNG**

# ÜBERSETZUNG: NUTZUNG VON APIs

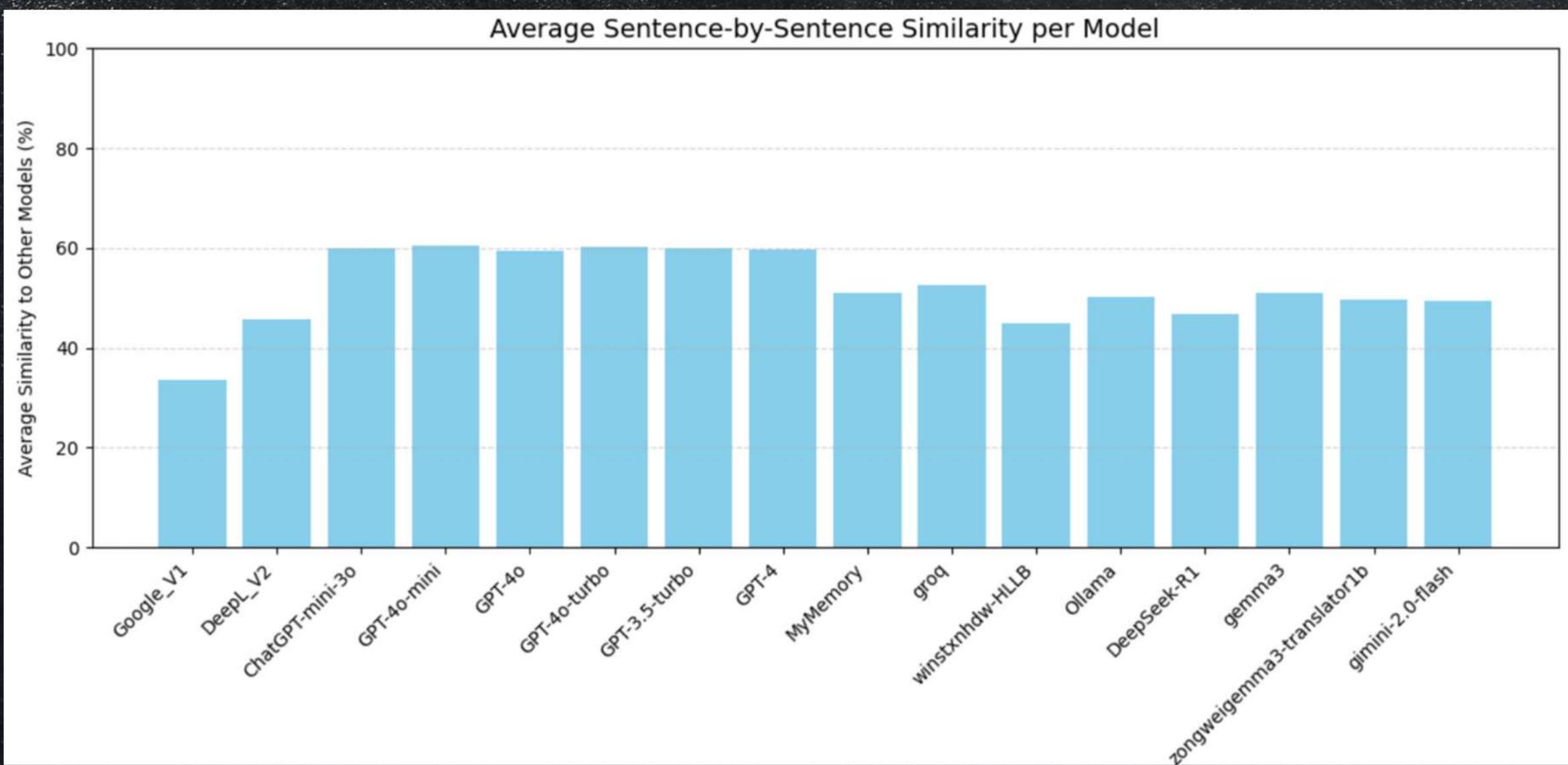
Im Schritt „Übersetzung“ wird der zuvor extrahierte Text mithilfe verschiedener APIs in die Zielsprache übertragen. Dabei kommen sowohl kostenpflichtige APIs wie DeepL und ChatGPT als auch kostenfreie wie Ollama und Google Translate API zum Einsatz. Diese Kombination ermöglicht eine präzise, natürliche Übersetzung und bietet Flexibilität je nach Projektanforderung.



# MODELLVERGLEICH: ÄHNLICHKEIT

Durchschnitt =  $\frac{\text{Ähnlichkeit zu B} + \text{Ähnlichkeit zu C} + \dots}{\text{Anzahl anderer Modelle}}$

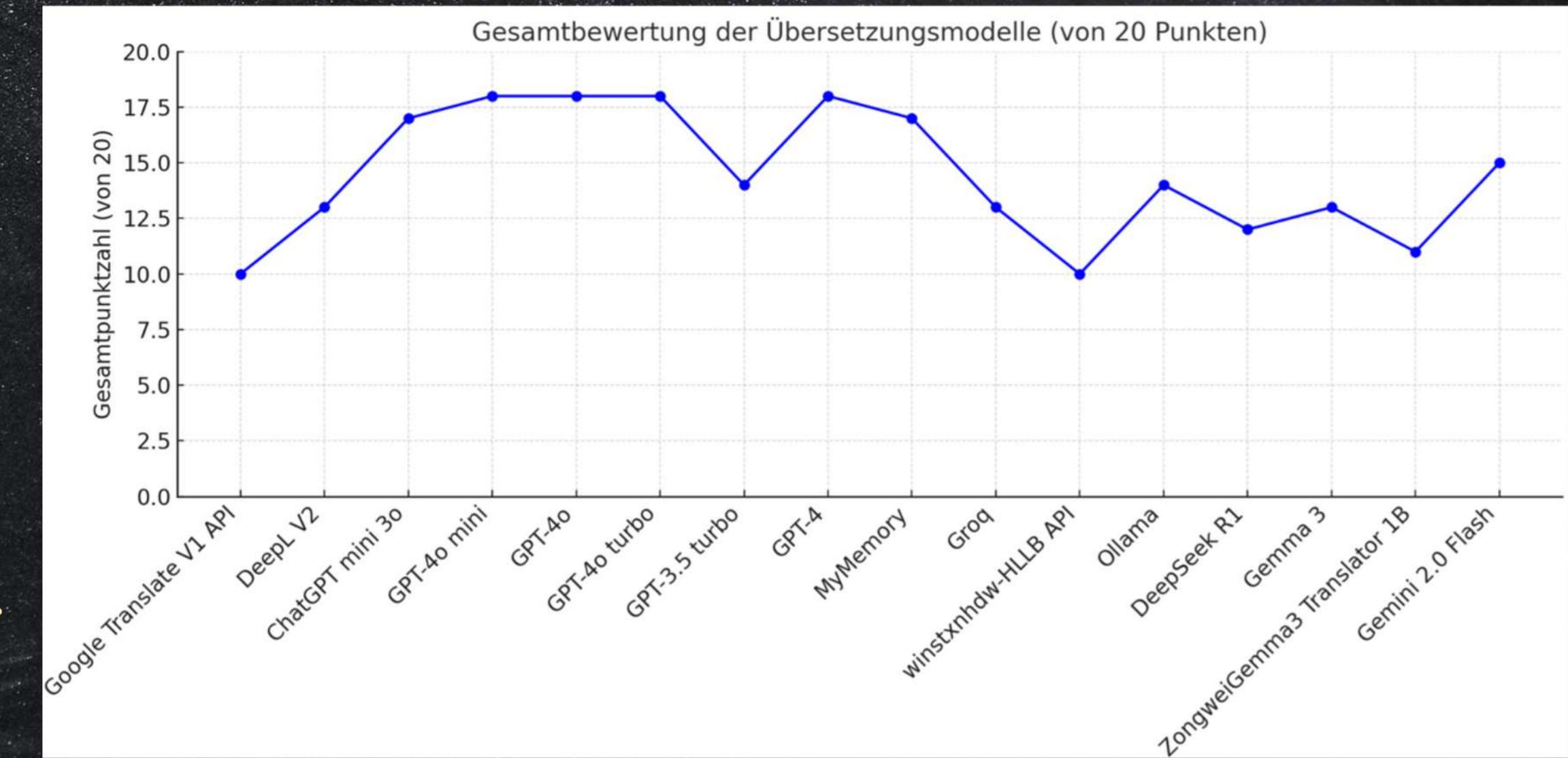
Das Balkendiagramm zeigt die durchschnittliche Satz-für-Satz-Ähnlichkeit der Modelle mit Werten zwischen etwa 50% und 60%, wobei GPT-4 und GPT-4o die höchste Ähnlichkeit von ca. 60% erreichen.



# QUALITÄTSBEWERTUNG DER MODELLE

Das Liniendiagramm zeigt die Gesamtpunktzahl der Übersetzungsmodelle basierend auf vier Bewertungskriterien: Übersetzungsqualität, Satzbau, Natürlichkeit und Grammatik.

Modelle wie GPT-4 und GPT-4o erreichen Spitzenwerte von ca. 18–19 Punkten, während andere etwas darunter liegen. Diese Daten helfen, die am besten geeigneten Modelle für verschiedene Einsatzbereiche und Zielgruppen zu identifizieren.





VERTONUNG



Für bestimmte Inhalte bietet sich das Modell „Ryan (high)“ an – eine männliche Stimme mit klarem, energischem Klang. Sie eignet sich gut für technische Erklärungen und schafft einen professionellen, präzisen Eindruck.



2

Für andere Kontexte ist das Modell „libritts\_r (medium)“ mit seiner weichen und natürlichen Klangfarbe eine ideale Wahl. Diese weibliche Stimme sorgt für ein freundliches und angenehmes Hörerlebnis.

## STIMMENAUSWAHL

# MODELLKOMBINATIONEN IM ÜBERBLICK

Basierend auf den zuvor erläuterten Analysen lassen sich sowohl kostenpflichtige als auch kostenfreie Modelle und Übersetzungs-APIs identifizieren, die jeweils unterschiedliche Stärken für die Umwandlung von Video zu Text und für die Übersetzung/Vertonung mitbringen.

## ABSCHLIESSENDE BEWERTUNG:

Für die Umwandlung von Video zu Text liefern die large-v3-Varianten eine sehr hohe Präzision und Stabilität. Bei der Übersetzung und Vertonung zeigen sich kostenpflichtige Optionen wie ChatGPT-4o oder GPT-4o-turbo als besonders flexibel und natürlich klingend. Kostenlose Alternativen wie DeepL v2 oder Ollama bieten solide Qualität und hohe Zugänglichkeit. Diese Vielfalt ermöglicht eine individuelle Anpassung je nach Projektanforderungen und Budget.

**large-v3 + DeepL v2 + Ryan (high)**

**large-v3 + ChatGPT-4o + Ryan (high)**

**large-v3 + Ollama + Ryan (high)**

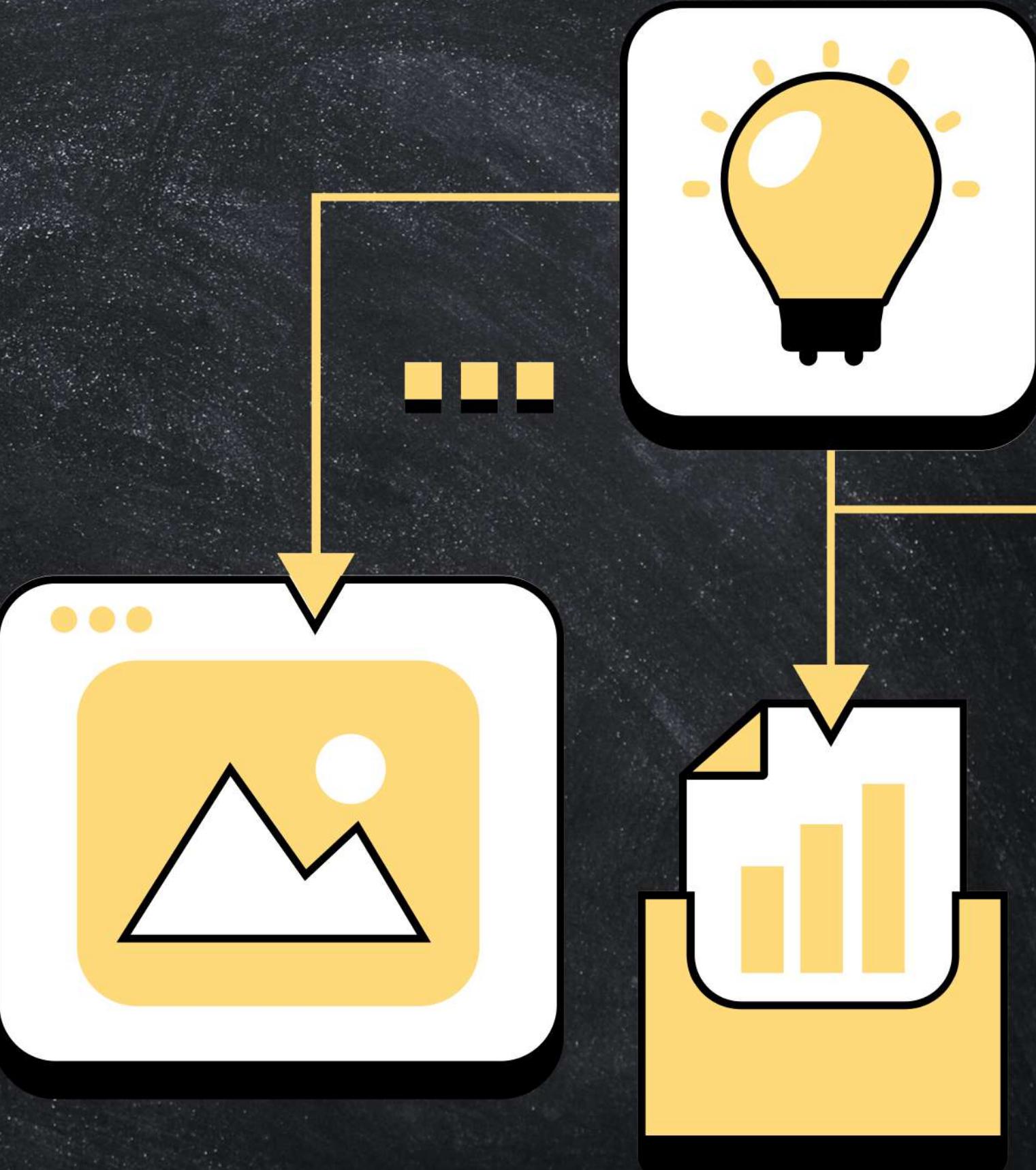
**large-turbo-v3 + GPT-4o-turbo + Ryan (high)**



**UNTERTITEL**

# UNTERTITEL EINFÜGEN

Dank der neuen Version 4.0.12 von Subtitle Edit ist es nun möglich, Untertitel direkt in Videos einzubetten. Dieser Schritt basiert auf den bereits in den vorherigen Schritten erstellten Outputs und gestaltet sich dadurch besonders einfach.



# » Probleme «

Unabhängig von den gewählten Modellen sind in bestimmten Fällen manuelle Korrekturen erforderlich. Die automatische Transkription kann nicht alle sprachlichen und inhaltlichen Feinheiten eines Videos vollständig abbilden, sodass eine nachträgliche Bearbeitung notwendig ist.

Eine vollständige Synchronisation ist aufgrund der unterschiedlichen Satzstrukturen in verschiedenen Sprachen nicht möglich. Daher ist es wichtig, für jedes Video individuelle Einstellungen und Zeitbereiche sorgfältig zu definieren, um den manuellen Korrekturaufwand zu minimieren.

# ZUSAMMENFASSUNG DER PROJEKTERGEBNISSE

Ein strukturierter und leicht nachvollziehbarer Arbeitsablauf, um den gesamten Prozess zu optimieren.

Detaillierte Beschreibung der Projektumgebung sowie die Analyse der verwendeten Tools, Methoden und deren Ergebnisse und Argumente.

Nicht alle APIs sind frei zugänglich oder leicht zu implementieren, was eine sorgfältige Auswahl und Anpassung erforderlich macht.

**API-Integration**

**Workflow**

**Dokumentierte Analyse**

**Skripte**

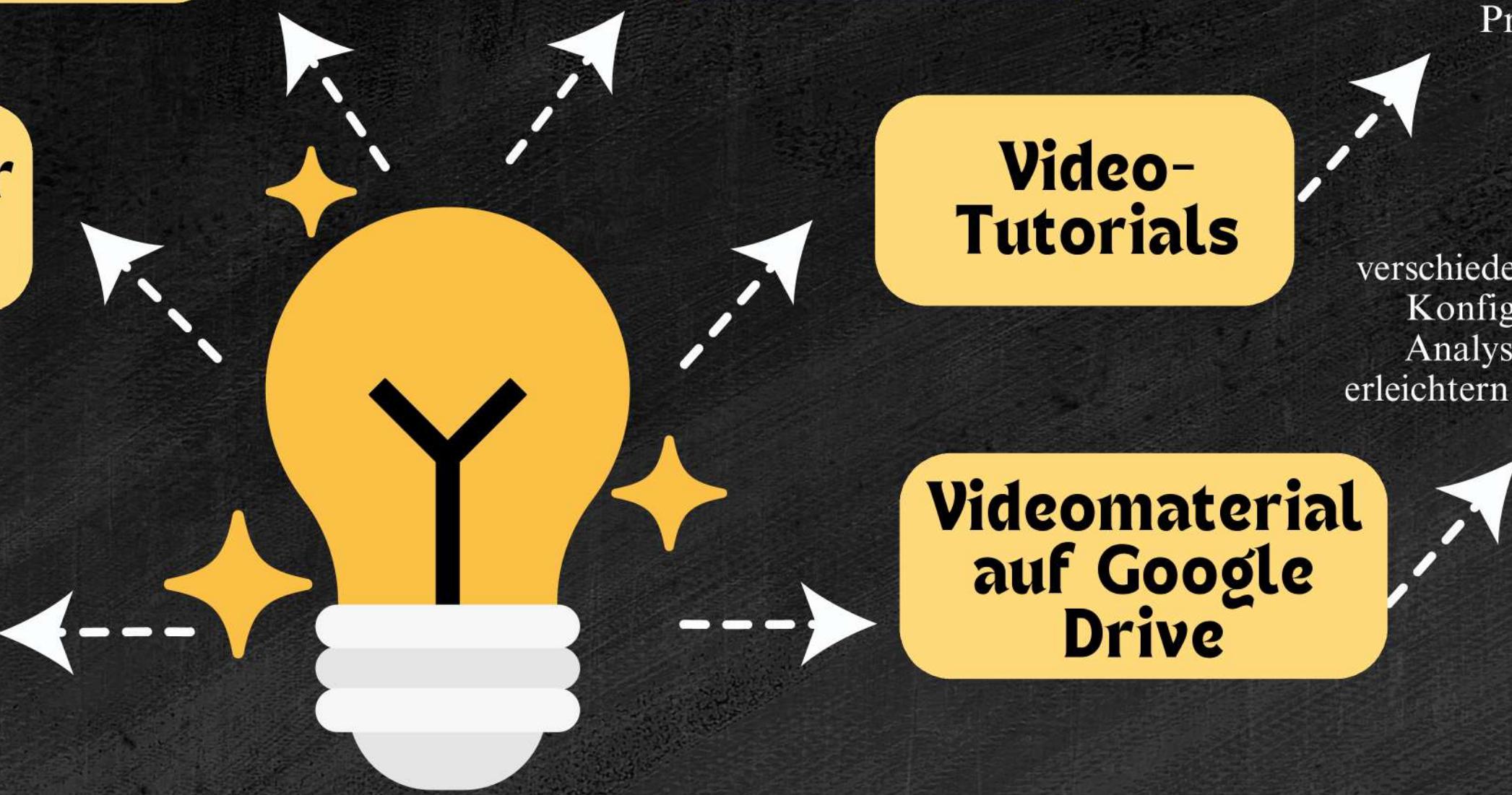
**Video-Tutorials**

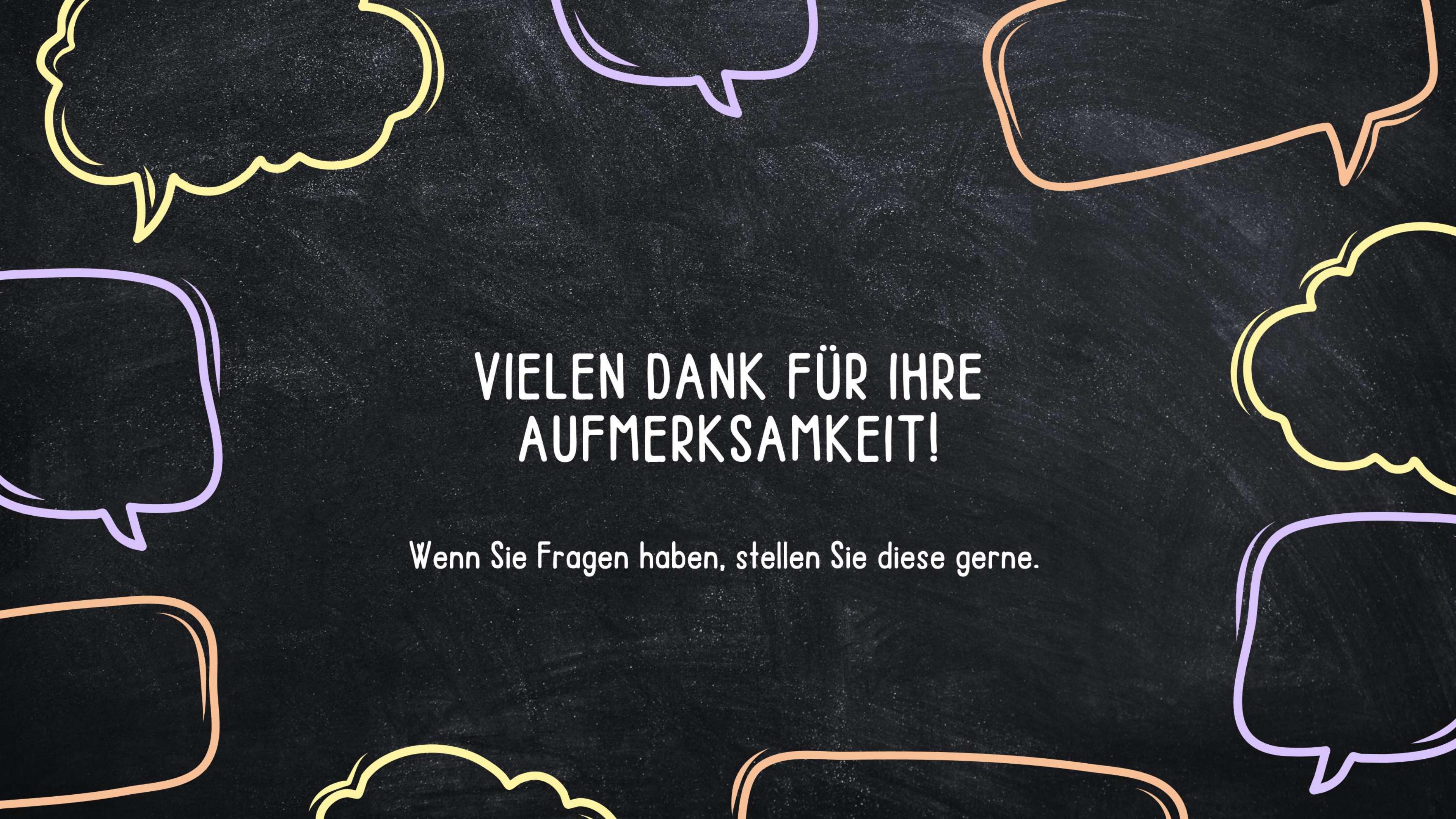
**Videomaterial auf Google Drive**

Hilfreiche Skripte für Automatisierung, Analyse oder individuelle Anpassungen.

Kurze und präzise Anleitungen, die die Nutzung und Funktionalität der Projektumgebung demonstrieren.

Kurze Videos mit verschiedenen Modellen und Konfigurationen sind zur Analyse bereitgestellt. Sie erleichtern die Bewertung der Ergebnisse.





VIELEN DANK FÜR IHRE  
AUFMERKSAMKEIT!

Wenn Sie Fragen haben, stellen Sie diese gerne.