



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

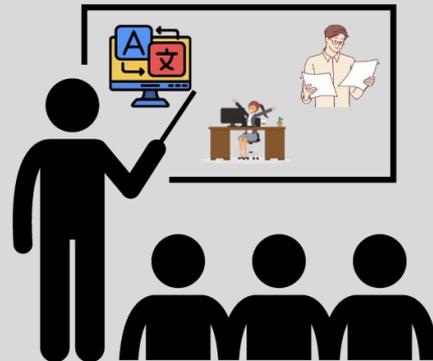
# MT-Literacy als Qualitätssicherung

Der Kundschaft eine geschickte Nutzung  
von maschineller Übersetzung vermitteln

Jessica Selinger, GS-WBF

Lidia Calabrese, BLV

équivalences, 15. November 2024





# Grundsätzliche Fragen

1

Welche Situationen sind geeignet, um die Kundschaft für maschinelle Übersetzung zu sensibilisieren?

2

Wie gehe ich auf meine Kundschaft zu?

3

Was genau soll thematisiert werden?

4

Warum ist es wichtig, mit der Kundschaft über maschinelle Übersetzung zu sprechen?



# Welche Situationen sind geeignet?

- Problematische Aufträge
- Anlässe (z. B. Konferenzen, Workshops)
- Social Media (z.B. LinkedIn)
- Diskussionen in der Gesellschaft





# Wie gehe ich auf meine Kundschaft zu?

- Professionell und kompetent
- Mit Hilfe von externen Expertinnen und Experten
- Der Kundschaft die nötigen Mittel für eine kritische Nutzung an die Hand geben
- Interaktion schaffen

2





# Wie arbeitet man mit MÜ-Systemen?

- Hauptfunktionen des MÜ-Systems, das von der Kundschaft genutzt wird
- Funktionsweise von MÜ-Systemen (5 Schlüsselkonzepte):
  - Trainingsdaten
  - Neuronale Netzwerke
  - Word embeddings
  - Aufmerksamkeitsmechanismus
  - Transformer

3



# Beispiel von Trainingsdaten

## 1. Abschnitt: Datensicherheit

### Art. 1 Grundsätze

<sup>1</sup> Zur Gewährleistung einer angemessenen Datensicherheit müssen der Verantwortliche und der Auftragsbearbeiter den Schutzbedarf der Personendaten bestimmen und die im Hinblick auf das Risiko geeigneten technischen und organisatorischen Massnahmen festlegen.

<sup>2</sup> Der Schutzbedarf der Personendaten wird nach den folgenden Kriterien beurteilt:

a. Art der bearbeiteten Daten;

b. Zweck, Art, Umfang und Umstände der Bearbeitung.

<sup>3</sup> Das Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person wird nach den folgenden Kriterien beurteilt:

a. Ursachen des Risikos;

b. hauptsächliche Gefahren;

c. ergriffene oder vorgesehene Massnahmen, um das Risiko zu verringern;

d. Wahrscheinlichkeit und Schwere einer Verletzung der Datensicherheit trotz der ergriffenen oder vorgesehenen Massnahmen.

<sup>4</sup> Bei der Festlegung der technischen und organisatorischen Massnahmen werden zudem die folgenden Kriterien berücksichtigt:

a. Stand der Technik;

b. Implementierungskosten.

<sup>5</sup> Der Schutzbedarf der Personendaten, das Risiko und die technischen und organisatorischen Massnahmen sind über die gesamte Bearbeitungsdauer hinweg zu überprüfen. Die Massnahmen sind nötigenfalls anzupassen.

### Art. 2 Ziele

Der Verantwortliche und der Auftragsbearbeiter müssen technische und organisatorische Massnahmen treffen, damit die bearbeiteten Daten ihrem Schutzbedarf entsprechend:

## Section 1 Sécurité des données

### Art. 1 Principes

<sup>1</sup> Pour assurer une sécurité adéquate des données, le responsable du traitement et le sous-traitant établissent le besoin de protection des données personnelles et déterminent les mesures techniques et organisationnelles appropriées à prendre par rapport au risque encouru.

<sup>2</sup> Le besoin de protection des données personnelles est évalué en fonction des critères suivants:

a. le type de données traitées;

b. la finalité, la nature, l'étendue et les circonstances du traitement.

<sup>3</sup> Le risque pour la personnalité ou les droits fondamentaux de la personne concernée est évalué en fonction des critères suivants:

a. les causes du risque;

b. les principales menaces;

c. les mesures prises ou prévues pour réduire le risque;

d. la probabilité et la gravité d'une violation de la sécurité des données, malgré les mesures prises ou prévues.

<sup>4</sup> Lors de la détermination des mesures techniques et organisationnelles, les critères suivants sont de plus pris en compte:

a. l'état des connaissances;

b. les coûts de mise en œuvre.

<sup>5</sup> Le besoin de protection des données personnelles, le risque encouru, ainsi que les mesures techniques et organisationnelles sont réévalués pendant toute la durée du traitement. En cas de besoin, les mesures sont adaptées.

### Art. 2 Objectifs

En fonction du besoin de protection, le responsable du traitement et le sous-traitant prennent des mesures techniques et organisationnelles pour que les données traitées:





# Trainingsdaten

## Ausgangssprache

- Art der bearbeiteten Daten;

1 2 3 4

- Zweck, Art, Umfang und  
Umstände der Bearbeitung

## Zielsprache

- type de données traitées ;

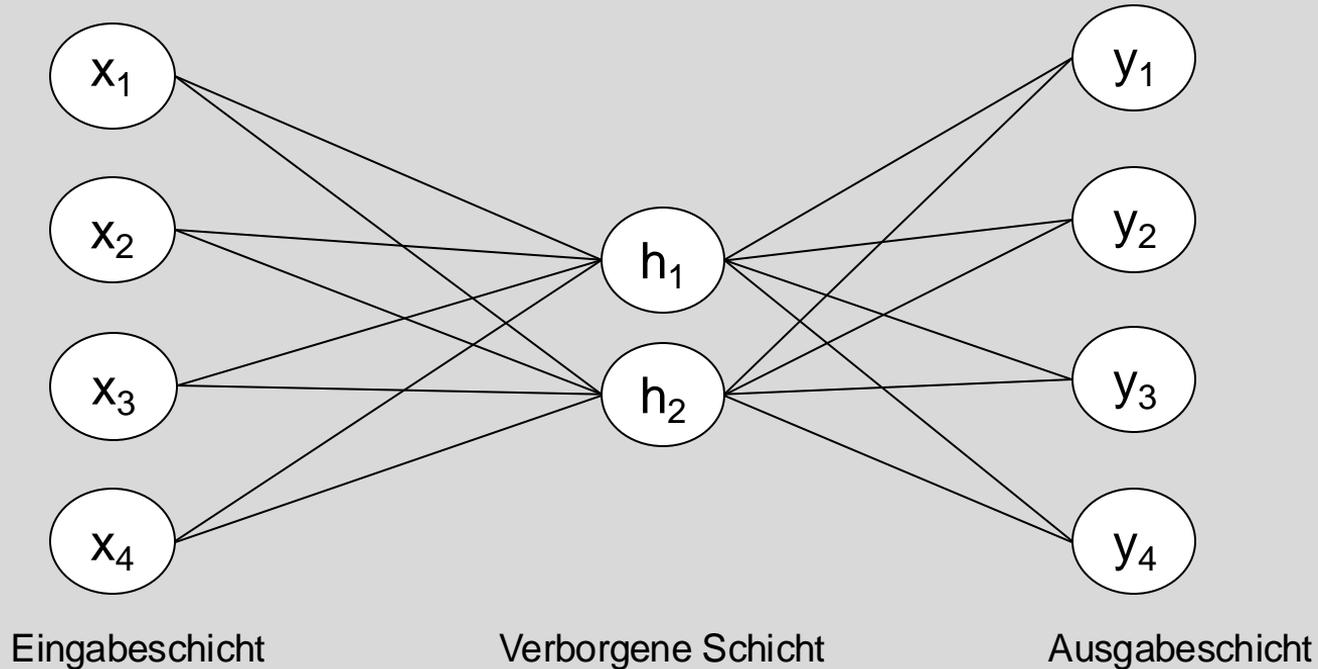
5 6 7 8

- la finalité, la nature, l'étendue  
et les circonstances du  
traitement

3



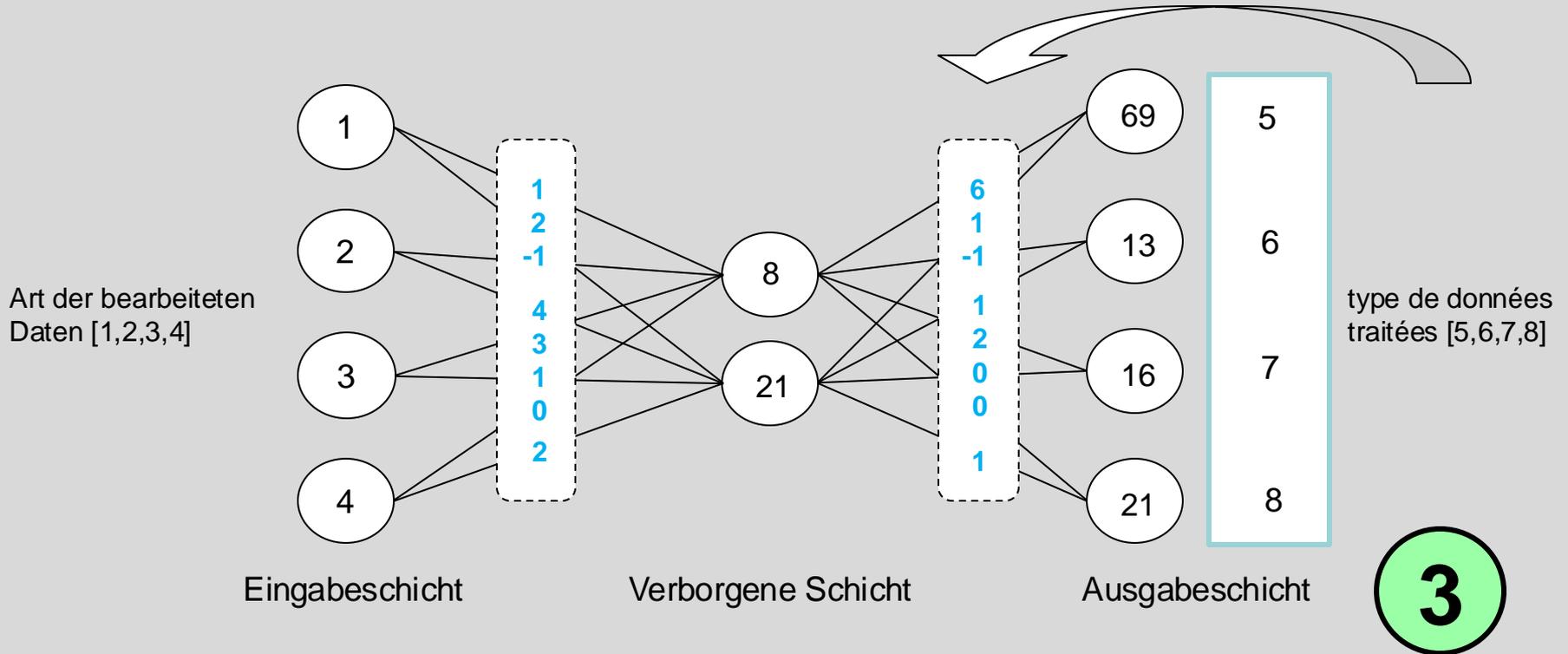
# Künstliches neuronales Netzwerk (KNN)



3



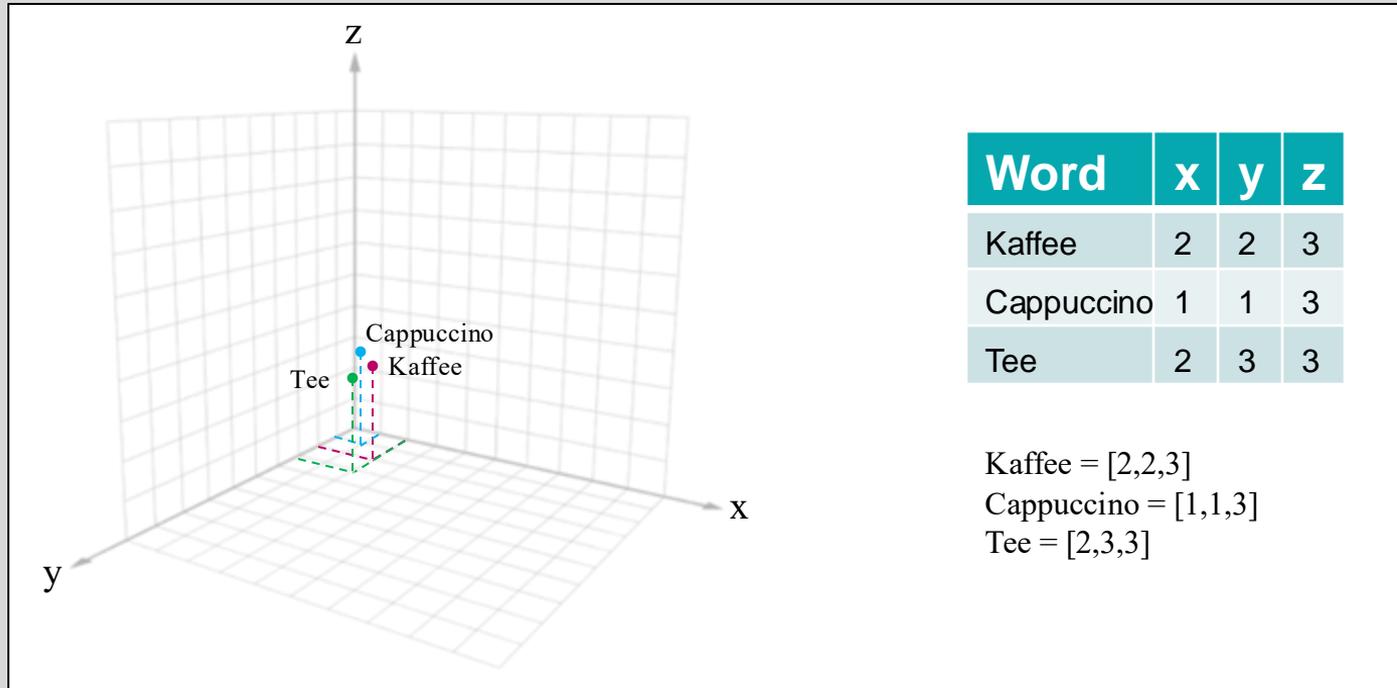
# Künstliches neuronales Netzwerk (KNN)





# Word embeddings

«You shall know a word by the company it keeps» (John Firth, 1957)



3



# Aufmerksamkeitsmechanismus



The dog didn't cross the street because **it** was too tired  
The dog didn't cross the street because **it** was too noisy

3



# Aufmerksamkeitsmechanismus



The **dog** didn't cross the street because **it** was too tired  
The dog didn't cross the **street** because **it** was too noisy

3



# Transformer

Englisch (erkannt) ▾

↔ Deutsch ▾

Einstellungen ^

automatisch ▾

Default glossary ▾

I arrived at the bank after crossing the river

×

Ich kam am Ufer an, nachdem ich den Fluss überquert hatte.

DeepL Pro, Mai 2024

3

Intro

Wann

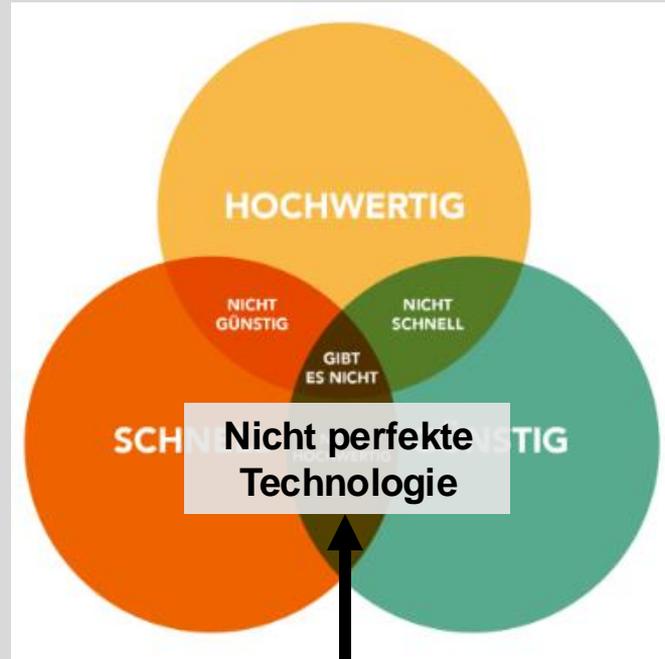
Wie

Was

Warum



# Verwendungsszenarien

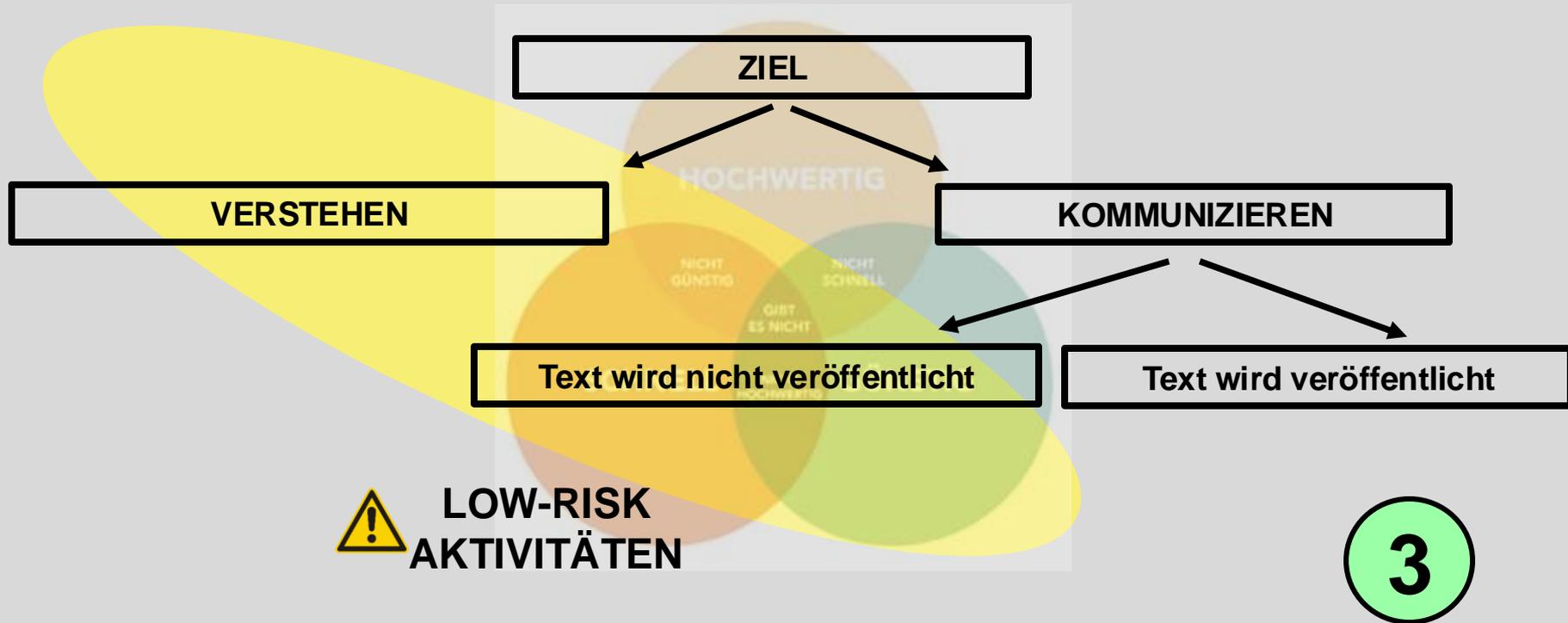


Maschinelle Übersetzung

3



# Verwendungsszenarien





# Low-Risk Aktivitäten: Risiken

- **Datenschutz**

Selbst bei Pro-Versionen!

- **Qualität**

- Unterschiedlich je nach System, Sprachkombination, Satz
- Keine Empathie, Ironie, Logik, kulturelle Komponente, kein Weltwissen
- Die Maschine kennt nur das, was in den Trainingsdaten steht
- False Fluency
- Einige Fehler sind schwer erkennbar → kognitiver Aufwand

- **Reputationsschäden und Haftungsrisiken**

Verantwortung liegt bei Userinnen und User

3



# Low-Risk Aktivitäten: Risiken

- **Datenschutz**

Selbst bei Pro-Versionen!

- **Qualität**

- Unterschiedlich je nach System, Sprachkombination, Satz
- Keine Empathie, Ironie, **Logik, kulturelle Komponente**, kein Weltwissen
- Die Maschine kennt nur das, was in der Trainingsdatensatz
- False Fluency
- Einige Fehler sind schwer erkennbar

- **Reputationsschäden und Haftung**

Verantwortung liegt bei Userinnen und User

The word "Switzerland" has 11 letters.	x	Das Wort "Schweiz" hat 11 Buchstaben.
The distance from Bern to Zurich is 59 miles	x	Die Entfernung von Bern nach Zürich beträgt 59 Meilen





# Low-Risk Aktivitäten: Risiken

- **Datenschutz**

Selbst bei Pro-Versionen!

- **Qualität**

- Unterschiedlich je nach System, Sprachkombination, Satz
- Keine Empathie, Ironie, Logik, kulturelle Komponente, **kein Weltwissen**
- Die Maschine kennt nur das, was in den Trainingsdaten

- False Fluency

- Einige Fehler sind so

- **Reputationsschädigung**

Verantwortung liegt bei Us

Immer wenn Marie ihre Katze füttert, beisst sie sie.	×	Whenever Marie feeds her cat, she bites it.
Immer wenn Marie ihre Katze füttert, beisst sie sie.	×	Ogni volta che Marie dà da mangiare al suo gatto, lo morde.
Immer wenn Marie ihre Katze füttert, beisst sie sie.	×	Chaque fois que Marie nourrit son chat, elle le mord.





# Low-Risk Aktivitäten: Risiken



inkombination, Satz  
elle Komponente, kein Weltwissen

nur das, was in den Trainingsdaten steht

- False Fluency
- Einige Fehler sind schwer erkennbar → kognitiver Aufwand
- **Reputationsschäden und Haftungsrisiken**  
Verantwortung liegt bei Userinnen und User





# Low-Risk Aktivitäten: Risiken

- **Datenschutz**

Selbst bei Pro-Versionen!

- **Qualität**

- Unterschiedlich je nach System

- Keine Empathie, Ironie

- Die Maschine kennt nur

- **False Fluency**

- Einige Fehler sind schwer

- **Reputationsschäden und Haftungsrisiken**

Verantwortung liegt bei Userinnen und User

## Keine Logik

Il relève, en lien avec la société Eurochem citée dans la présente motion, que la Suisse se range parmi toute une série d'États de l'UE qui appliquent ladite pratique

Er stellt im Zusammenhang mit der in der vorliegenden Motion zitierten Firma Eurochem fest, dass die Schweiz zu einer ganzen Reihe von EU-Staaten gehört, die diese Praxis anwenden.

3



# Low-Risk Aktivitäten: Risiken

- **Datenschutz**

Selbst bei Pro-Versionen!

- **Qualität**

- Unterschiedlich je nach System

- Keine Empathie, Ironie

- Die Maschine kennt nicht

- **False Fluency**

- Einige Fehler sind schwer

- **Reputationsschäden und Haftungsrisiken**

Verantwortung liegt bei Userinnen und User

**≠ als AT**

Insbesondere «naturnahe Beutensysteme» lassen Kontrollen kaum oder nur mit grossem Aufwand für den Bieneninspektor zu.

In particolare, i "sistemi di alveari quasi naturali" consentono ispezioni con uno sforzo minimo o nullo da parte dell'ispettore apistico.

3



# Low-Risk Aktivitäten: Gegenmassnahmen

- **Datenschutz**

Keine Personendaten, keine vertrauliche oder geheime Informationen eingeben

- **Qualität**

- Bin ich in der Lage, den Output genau zu evaluieren?
- Input und Output vergleichen UND den gesamten Output überprüfen
- Die gängigsten Fehler im Output korrigieren
- Vor der Eingabe Text vorbereiten → Garbage in, garbage out

- **Reputationsschäden und Haftungsrisiken**

- Keine Entscheidung aufgrund von einem nicht überprüfem Output treffen!
- MÜ-Output eventuell als solchen kennzeichnen  
→ Transparenz und Verantwortung

3



# Low-Risk Aktivitäten: Gegenmassnahmen

- **Datenschutz**

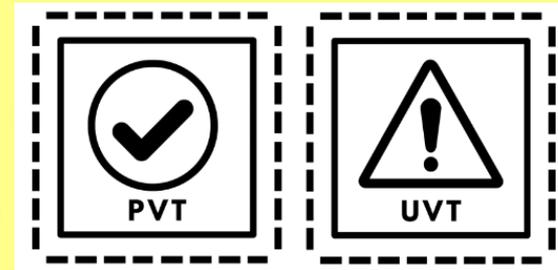
Keine Personendaten, keine vertrauliche oder geheime Informationen eingeben

- **Qualität**

- Bin ich in der Lage, den Output genau zu evaluieren?

Vorschlag von Alan K. Melby und  
Association for Machine Translation in the Americas  
(AMTA)

Labels für Konsumentenschutz



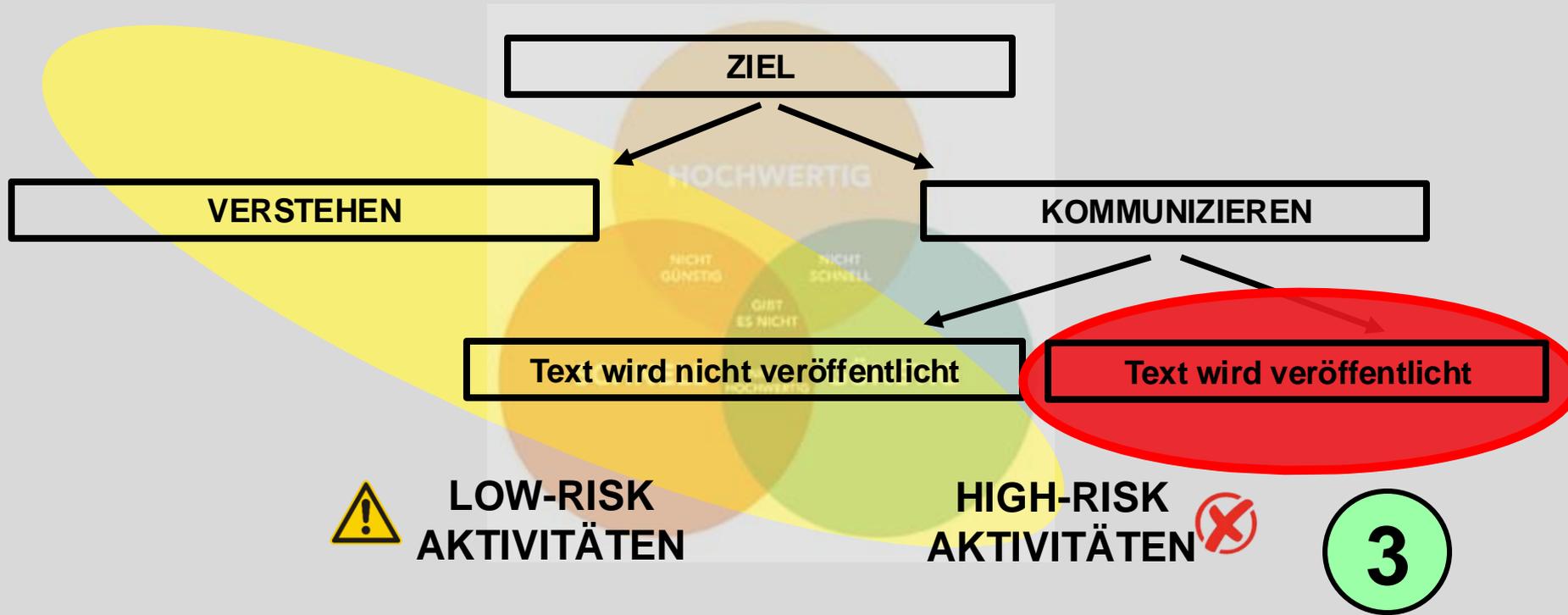
Professionally Verified Translation (PVT) / Un-Verified Translation (UVT)

→ **Transparenz** und **Verantwortung**

3



# Verwendungsszenarien





# High-Risk Aktivitäten

- Vertrauliche, geheime Texte
- Texte, die als Entscheidungsbasis dienen
- Offizielle Texte, Texte die haftungssicher sein müssen
- Texte die ein breites Publikum erreichen (Internet, Social Media usw.)
- Texte zur Veröffentlichung

3





# High-Risk Aktivitäten

Wie langlebig soll mein Text sein?

Schlechte Übersetzung = Schlechte Kommunikation!  
Was für Risiken könnte das haben?



Haftungsrisiken, finanzieller Schaden und Reputationsrisiken!

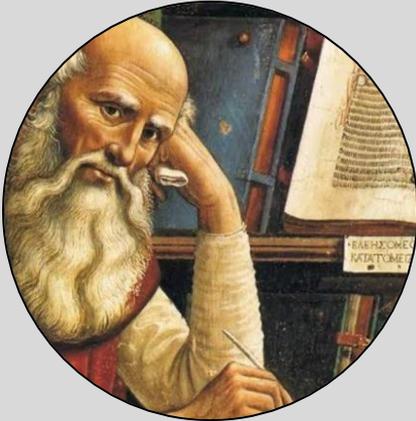


→ Professionelle Übersetzung ohne vorgängige MÜ

3



# Professionelle Übersetzung: Vorstellung



3



# Professionelle Übersetzung: Realität

CAT-Tools

Terminologie

Prozesse

MÜ und KI

Menschliches  
Know-how

3

Intro

Wann

Wie

Was

Warum



# Professionelle Übersetzung: Realität

CAT-Tools

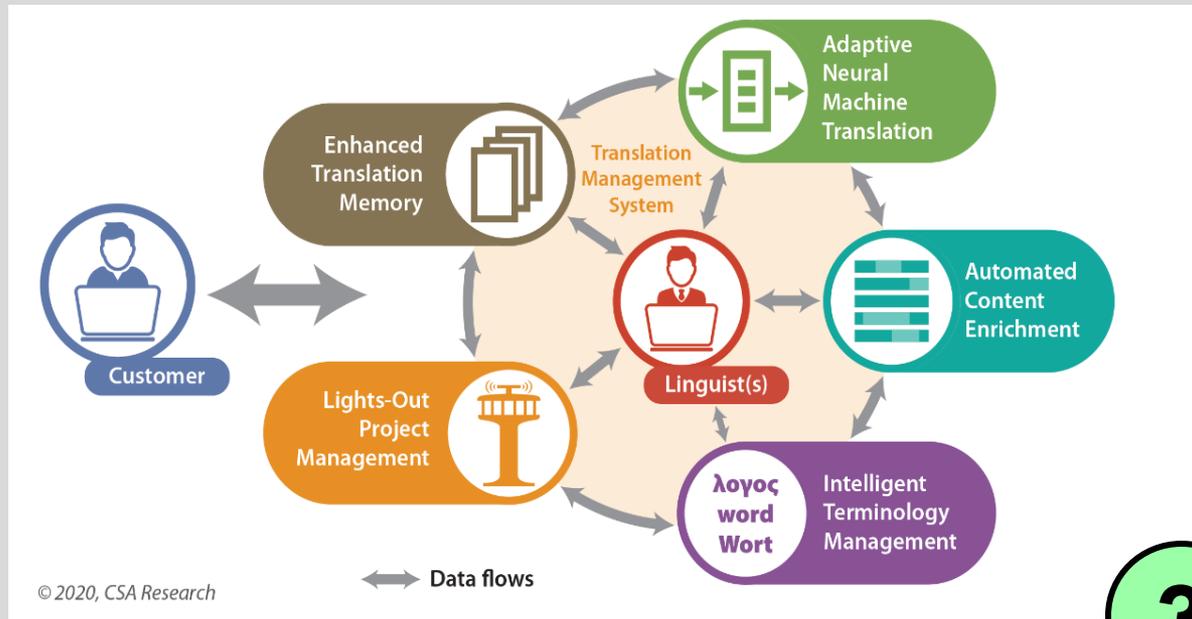
Terminologie

Prozesse

MÜ und KI

Menschliches Know-how

## AUGMENTED TRANSLATION



3



# Stand der Dinge: Debatte in der Presse

## Fremdsprachen? Bald nicht mehr!

**Künstliche Intelligenz und Sprache** Ein Test zeigt: Beim Übersetzen leistet KI Erstaunliches. Menschliche Profis werden nervös.  
Basler Zeitung, 11.3.2024

**Erstaunliche KI: Mit diesem Übersetzer spricht jeder plötzlich bis zu 9 Sprachen**

Chip.de, 11.3.2024

## Meta claims breakthrough in quest to build Hitchhiker's Guide-style language translator

Facebook owner reveals free AI tool capable of translating speech from 100 languages

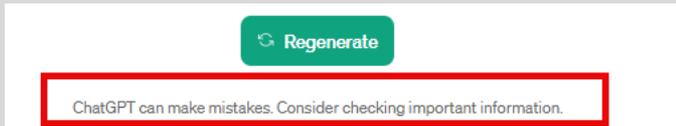
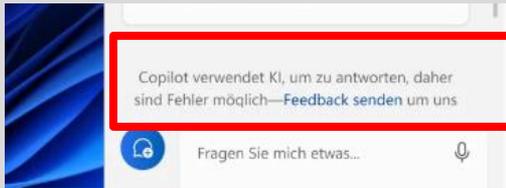
The Telegraph, 22.8.2023



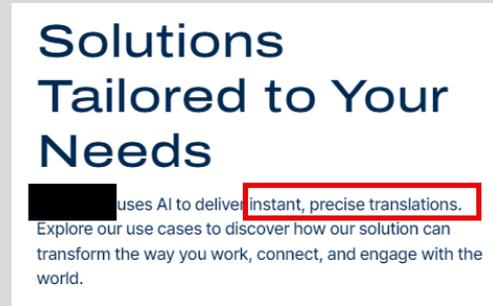


# Stand der Dinge: Transparenz der Tools

## Generative KI-Tools



## MÜ-Tools





# Stand der Dinge: Einfache Bedienung

**MT ist extrem einfach zu verwenden**

Idee, dass Übersetzen einfach sei



4



# MT-Literacy ist nötig!

## Neue digitale Kompetenz, eher kritische als technische

- Ob, wann und warum sollte man MÜ verwenden
- Bewusstsein über Risiken und Folgen
- Anwendungsszenarien definieren
- Wie soll der Output angepasst / weiterverwendet werden
- Reale Kosten → Ethik!



4



# MT-Literacy ist auch unsere Aufgabe!

- **Wer, wenn nicht wir**  
Das können wir nicht delegieren
- **Wir sollten an der Debatte teilnehmen**  
Den Dialog mit den anderen Akteuren der Sprachindustrie suchen
- **MT-Literacy ist je nach Situation anpassbar**  
Auf vielen Ebenen gibt es noch viel zu tun!
- **Wir sind die Expertinnen und Experten**  
Chance, sich als Profis neu zu positionieren



**ES IST ZEIT, SICHTBAR ZU WERDEN!**

**4**



# Literaturverzeichnis

- AG «KI in der Bundesverwaltung» (2024), [Merkblatt zur Verwendung von generativen KI-Werkzeugen in der Bundesverwaltung](#) (Zugriff: September 2024)
- Allamar, J. (2019), [Visualizing A Neural Machine Translation Model \(Mechanics of Seq2seq Models With Attention\)](#) (Zugriff: Januar 2023)
- Allamar, J. (2021), [The Illustrated Transformer](#) (Zugriff: Mai 2023)
- Bowker, L. , and J. Buitrago Ciro (2019), [Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community](#) (Zugriff: September 2024)
- Bowker L. (2019), [Machine Translation Literacy as a Social Responsibility](#) (Zugriff: September 2024)
- Bowker L. (2021), [Machine Translation Literacy: Stop. Think. Act](#) (Zugriff: September 2024)
- Bowker L. (2021), [Machine Translation: 5 ways to be a savvy user](#) (Zugriff: September 2024)
- Bowker L. (2023), [De-mystifying Translation. Introducing Translation to Non-translators](#), Routledge, London (Zugriff: Oktober 2024)
- Bundeskanzlei (2024), [Merkblatt DeepL Pro Translate und DeepL Write](#) (Zugriff: September 2024)
- Bundesrat (2020), [Leitlinien «Künstliche Intelligenz» für den Bund. Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung](#) (Zugriff: Januar 2024)
- Caprasi, C. (2023) Artificial Intelligence, Machine Learning , Deep Learning, GenAI and more, [Artificial Intelligence. Machine Learning . Deep Learning. GenAI and more | by Chiara Caprasi | Women in Technology | Medium](#) (Zugriff: Oktober 2024)
- De Palma D. (2017), [Augmented Translation Powers up Language Services](#). (Zugriff: September 2024)
- Financial Times, [Generative AI exists because of the transformer](#) (Zugriff: November 2023)
- Fleuret, F. (2021), [Deep learning - 12.3 Word embeddings and translation](#) (Zugriff: Januar 2023)
- Forcada, M. (2017), [Making sense of neural machine translation](#) (Zugriff: Februar 2023)
- Hirschle, J. (2022), Deep Natural Language Processing, Carl Hanser Verlag München



# Literaturverzeichnis

IBM Cloud Learn Hub (2020), [Neural Networks](#) (Zugriff: Dezember 2023)

Läubli, S. (2021), What Is Machine Translation and How Does It Work, Direction générale de la traduction de la Commission européenne, pp. 88-107 (von dem Autor zur Verfügung gestellte Folien).

Melby K. A (2024), [Labels for Translation Output: a Triple Win](#) (Zugriff: Oktober 2024)

Mikolov, T. et al (2013), [Efficient estimation of Word Representations in Vector Space](#) (Zugriff: März 2023)

Mitchell, M. (2019), Intelligence artificielle: triomphes et déceptions, éditions Dunod

O'Brien, S. und Ehrensberger-Dow, M. (2020), [MT Literacy – A cognitive view](#), (Zugriff: Oktober 2024)

Uszkoreit, J. et al (2017), [Transformer: A Novel Neural Network Architecture for Language Understanding](#) (Zugriff: November 2023)