

# Gibt es eine Human-Machine-Parity bei der NMÜ?

**Samuel Läubli**

ASTT-Tagung 2022: Eine Welt ohne Babel?

Germersheim, 30. Juni 2022



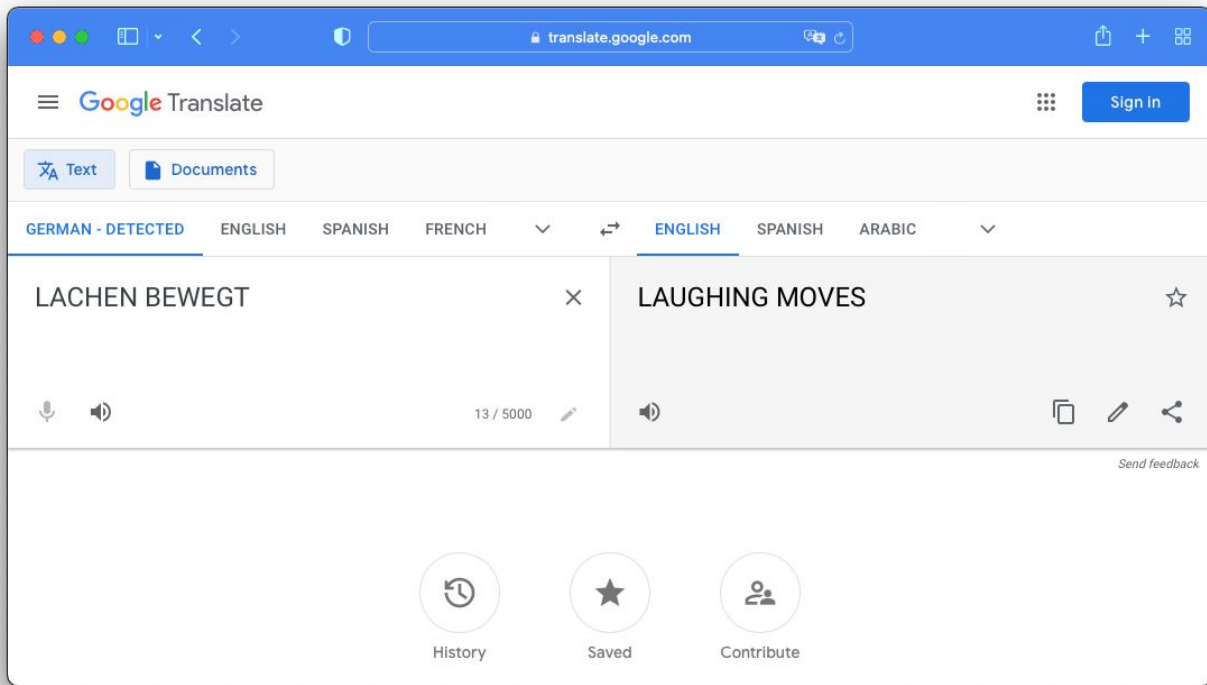
**Universität  
Zürich** <sup>UZH</sup>

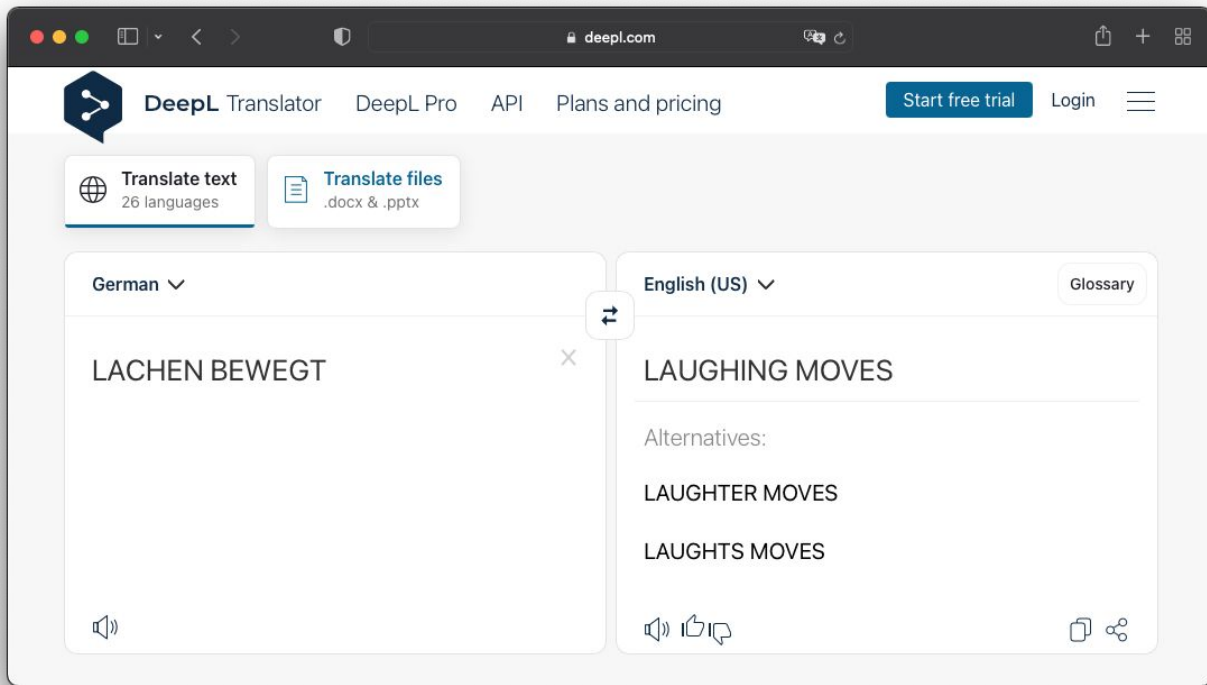
**textshuttle**

**zh  
aw**

Übersetzen Sie den folgenden Text ins Englische:


**LACHEN BEWEGT**





Lachen Online x +

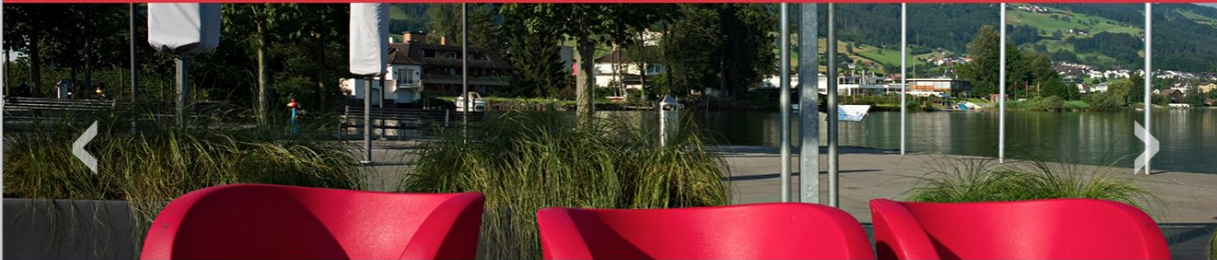
← → ↻ Not Secure | lachen.ch/de/



**GEMEINDE LACHEN SZ**

Links Index My Services

Portrait Aktuelles Verwaltung Politik Bildung Wirtschaft Soziales Freizeit / Vereine Tourismus



LACHEN BEWEGT

**Aktuelle Neuigkeiten** [Alle Neuigkeiten](#)

<p>9. Apr. 2020</p> <p><b>Veröffentlichung Verwaltungsbericht und Detailversion Rechnung 2019</b></p>	<p>3. Apr. 2020</p> <p><b>Halten Sie Abstand - auch am Wochenende</b></p> <p>Das voraussichtlich sonnige Wochenende lädt zur Bewegung</p>	<p>3. Apr. 2020</p> <p><b>Erreichbarkeit Ostern</b></p> <p>Die Telefone der Verwaltung sind am Karfreitag, 10. April 2020 sowie</p>
---	---	---

**Oft gesucht**

- Coronavirus: Erwerbsausfall
- Stellenangebote
- Fristerstreckung  
Steuererklärung 2019

# Wie unterscheidet sich die Qualität von maschineller Übersetzung von professioneller Humanübersetzung?

Samuel Läubli

ASTT-Tagung 2022: Eine Welt ohne Babel?

Germersheim, 30. Juni 2022



Universität  
Zürich <sup>UZH</sup>

**textshuttle**

**zh  
aw**

# Samuel Läubli

## Hintergrund

- seit 2021 **Operativer Leiter, Professur Mensch-Maschine-Kommunikation**  
Zürcher Hochschule für Angewandte Wissenschaften
- seit 2016 **Chief Technology Officer (CTO)**  
TextShuttle
- seit 2016 **Lehrbeauftragter**  
Universität Zürich
- 2020 **PhD, Machine Translation (MT)**  
Universität Zürich
- 2014–16 **Sr. Computational Linguist**  
Autodesk
- 2014 **MSc, Artificial Intelligence (AI)**  
University of Edinburgh

## Vorträge und Seminare zu AI/MT (Auswahl)

- Deutsche Übersetzungseinheit, Europäisches Parlament (LU)
- Directorate-General for Translation (DGT), Europäische Kommission (BE)
- Berufsverbände ASTTI (CH), BDÜ (DE), Universitas (AT)
- Sprachendienste (z.B. UEFA, SUVA, Swiss Life)
- CAS «Translation Technology and AI», Universität Zürich und ZHAW (CH)
- CAS «Big Data and Machine Learning», Universität Zürich (CH)
- Bachelor- and Master-Studiengänge, Universität Zürich (CH)

Nächster Start:  
Herbst 2022

# Inhalt

1. Ist maschinelle Übersetzung (MT) gleich gut wie professionelle Humanübersetzung (HT)?
2. Wie unterscheidet sich die Qualität von MT und HT?



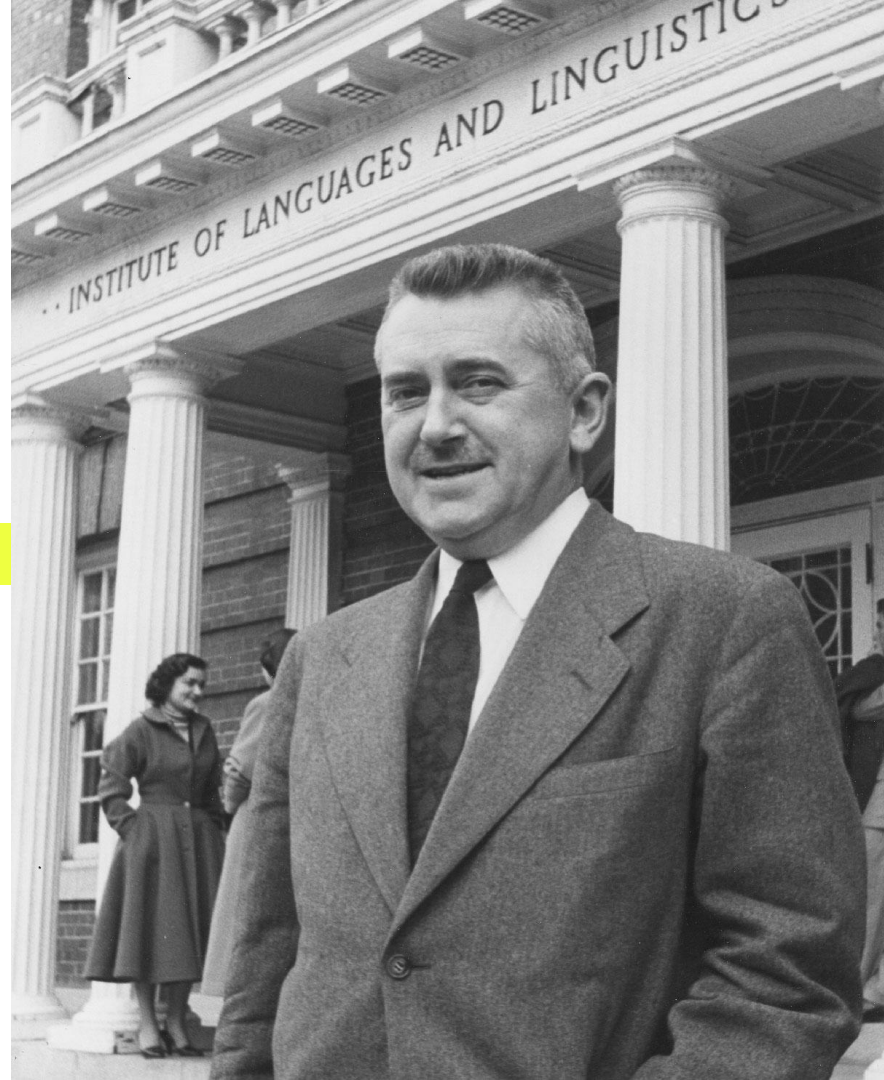


Ist MT gleich gut wie HT?

# Léon Dostert

Although he emphasized that it is not yet possible “to insert a Russian book at one end and come out with an English book at the other,” Doctor Dostert predicted that “five, perhaps three years hence, interlingual meaning conversion by electronic process in important functional areas of several languages may well be an accomplished fact.”

IBM-Pressemitteilung, 8. Januar 1954



# *Russian Is Turned Into English By a Fast Electronic Translator*

By **ROBERT K. PLUMB**

A public demonstration of what is believed to be the first successful use of a machine to translate meaningful texts from one language to another took place here yesterday afternoon.

This may be the culmination of centuries of search by scholars for "a mechanical translator." So far the system has a vocabulary of only 250 words. But there are no foreseeable limits to the number of words that the device can store or the number of languages it can be directed to translate.

Scholars and scientists who worked on it believe that within

scientists of the corporation and scholars of the Georgetown University Institute of Languages and Linguistics in Washington.

The "mechanical" part of the translation system, which is mostly electronic, is a standard commercial model of the largest International Business Machines "stock" computer. This device, called the IBM Type 701 Electronic Data Processing Machine, was put on the market last April. Since then twelve of the machines have been sold to commercial, military and university computation laboratories.

New York Times (Frontseite)

8. Januar 1954

## IN THE NEWS

*(Continued from page 40)*

### Electronic Brain Opens Future for Mechanical Translators

Russian was translated into English by an electronic "brain" recently for the first time.

Brief statements about politics, law, mathematics, chemistry, metallurgy, communications and military affairs were submitted in Russian by linguists of the Georgetown University Institute of Languages and Linguistics to the famous 701 computer of the International Business Machines Corporation. And the giant

attention was turned at brief intervals from these lightninglike numerical calculations to the altogether different consideration of logic in an entirely new and strange realm for giant electronic data processing machines: the study of human behavior—specifically, the human use of words. The result, as publicly proved today, was an unqualified success.

"The potential value of this experiment for the national interest in defense or in peace is readily seen," Prof. Leon Dostert, Georgetown language scholar who origi-

# Maschinelle Übersetzung

Georgetown-IBM, ~1962

Biological experiments, which were conducted on different cosmic LETATEL6NYX APPARATI, the astrophysical investigations of cosmic space and the flights of Soviet and also American KOSMONAVTOV with the sufficient convincingness showed, that the short-term orbital flights of below radiation belts of ground upon the absence of the increased solar activity in radiation in relation are safe.

- Acht Jahre Entwicklungszeit
- Post-editing langsamer als Humanübersetzung
- Post-editing teurer als Humanübersetzung

# SDL Cracks Russian to English Neural Machine Translation

Global Enterprises to Capitalize on Near Perfect Russian to English Machine Translation as SDL Sets New Industry Standard

**YEAH SURE**

**WHATEVS**

# Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English

Microsoft AI Blog, 14. März 2018

Hassan et al., 2018

[cs.CL] 29 Jun 2018

## Achieving Human Parity on Automatic Chinese to English News Translation

Hany Hassan<sup>\*</sup>, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Rengyan Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou

Microsoft AI & Research

### Abstract

Machine translation has made rapid advances in recent years. Millions of people are using it today in online translation systems and mobile applications in order to communicate across language barriers. The question naturally arises whether such systems can approach or achieve parity with human translations. In this paper, we first address the problem of how to define and accurately measure human parity in translation. We then describe Microsoft's machine translation system and measure the quality of its translations on the widely used WMT 2017 news translation task from Chinese to English. We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations. We also find that it significantly exceeds the quality of crowd-sourced non-professional translations.



# Evaluation von Mensch-Maschine-Parität

Basis: Menschen bewerten die Qualität von (a) maschinell und (b) professionell übersetzten Texten. Sie sehen nicht, ob ein Text von (a) oder (b) produziert wurde.

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。

Members of the public who find their cars obstructed by unfamiliar vehicles during their daily journeys can use the "Twitter Move Car" feature to address this distress when the driver of the unfamiliar vehicle cannot be reached.

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue.

# Evaluation von Mensch-Maschine-Parität

Basis: Menschen bewerten die Qualität von (a) maschinell und (b) professionell übersetzten Texten. Sie sehen nicht, ob ein Text von (a) oder (b) produziert wurde.

Evaluationsprotokoll:

---

[Hassan et al., 2018](#)

---

Bewerter*innen	Bilinguale Crowdworkers
Einheit	Satz
Translationese?	Ja

---

# Evaluation von Mensch-Maschine-Parität

Basis: Menschen bewerten die Qualität von (a) maschinell und (b) professionell übersetzten Texten. Sie sehen nicht, ob ein Text von (a) oder (b) produziert wurde.

Evaluationsprotokoll:

---

	<a href="#">Hassan et al., 2018</a>	<a href="#">Läubli et al., 2018, 2020</a>
Bewerter*innen	Bilinguale Crowdworkers	Professionelle Übersetzer*innen
Einheit	Satz	Satz / Dokument
Translationese?	Ja	Nein

---

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。

Members of the public who find their cars obstructed by unfamiliar vehicles during their daily journeys can use the "Twitter Move Car" feature to address this distress when the driver of the unfamiliar vehicle cannot be reached.

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue.

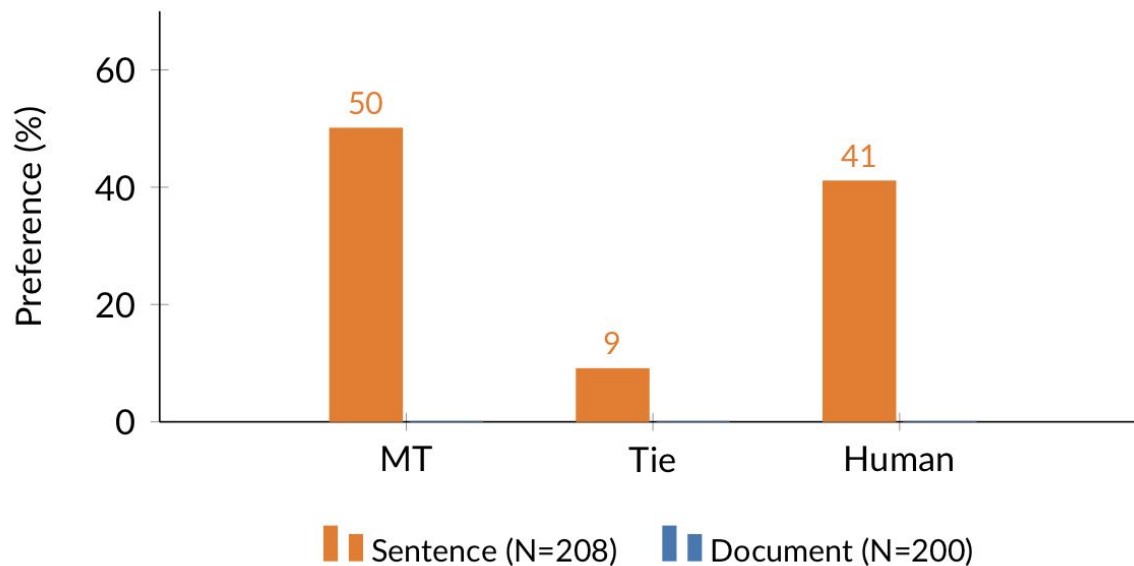
市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

Members of the public who find their cars obstructed by unfamiliar vehicles during their daily journeys can use the "Twitter Move Car" feature to address this distress when the driver of the unfamiliar vehicle cannot be reached. On August 11, Xi'an traffic police WeChat service number "Xi'an traffic police" launched "WeChat mobile" service. With the launch of the service, members of the public can tackle such problems in their daily lives by using the "WeChat Move" feature when an unfamiliar vehicle obstructs the movement of their vehicle while the driver is not at the scene. [...]

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]

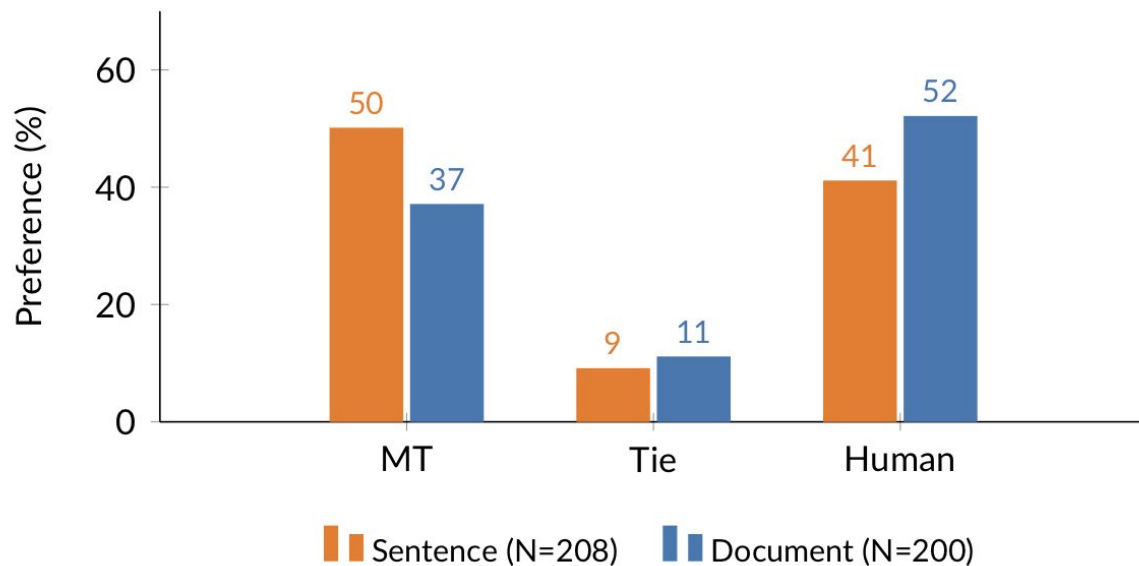
# Adäquatheit

Bewerter\*innen sehen Ausgangs- und Zieltext



# Adäquatheit

Bewerter\*innen sehen Ausgangs- und Zieltext





# Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English

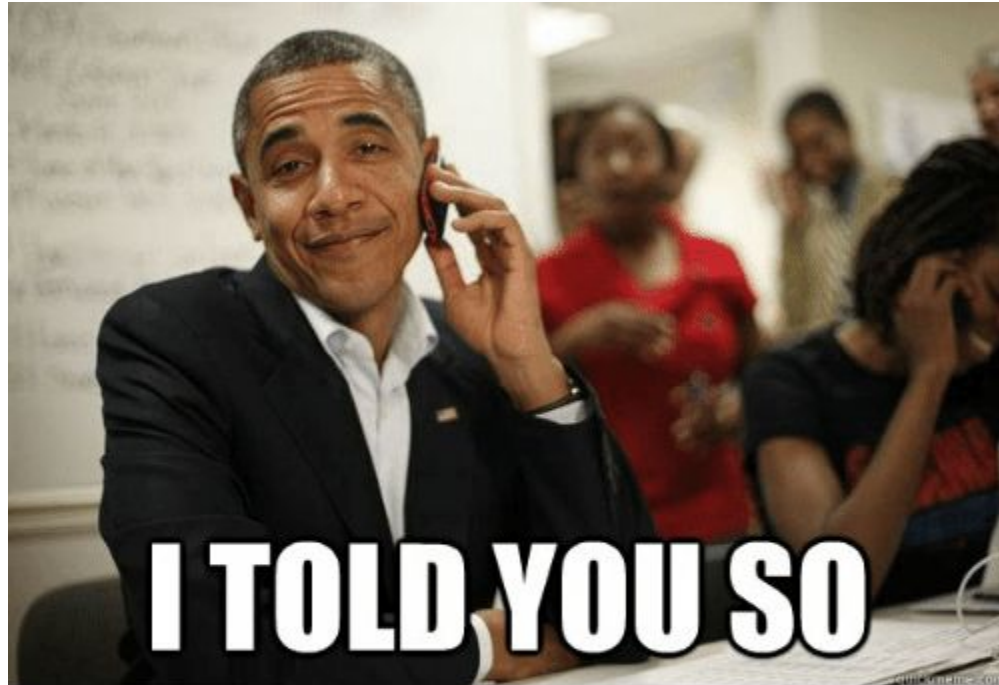
Microsoft AI Blog, 14. März 2018

*Nope.*

Hassan et al., 2018



**PROFESSIONAL TRANSLATOR:**



[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 01 September 2020](#)

## Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

[Martin Popel](#) , [Marketa Tomkova](#), [Jakub Tomek](#), [Lukasz Kaiser](#), [Jakob Uszkoreit](#), [Ondřej Bojar](#) & [Zdeněk Žabokrtský](#)

*Nature Communications* **11**, Article number: 4381 (2020) | [Cite this article](#)

**31k** Accesses | **34** Citations | **164** Altmetric | [Metrics](#)

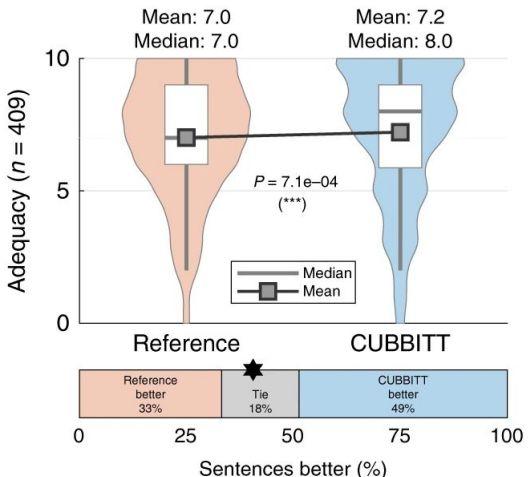
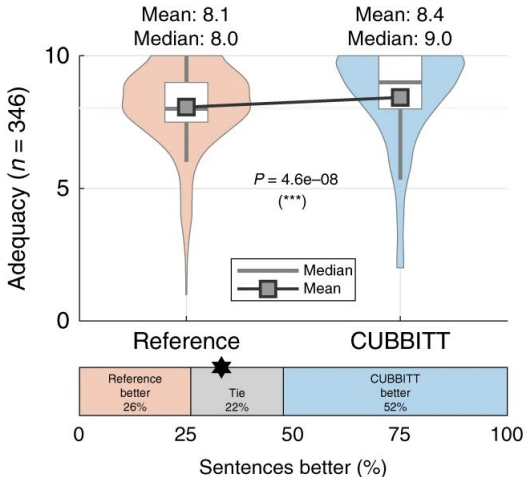
### Abstract

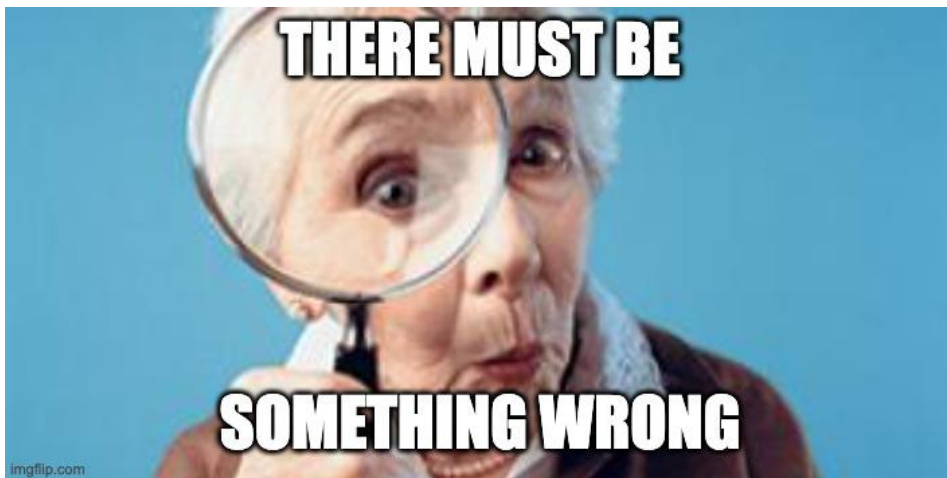
The quality of human translation was long thought to be unattainable for computer translation systems. In this study, we present a deep-learning system, CUBBITT, which challenges this view. In a **context-aware blind evaluation** by human judges, CUBBITT significantly outperformed professional-agency English-to-Czech news translation in preserving text meaning (translation adequacy). While human translation is still rated as more fluent, CUBBITT is shown to be substantially more fluent than previous state-of-the-art systems. Moreover, most participants of a Translation Turing test struggle to distinguish CUBBITT translations from human translations. This work approaches the quality of human translation and even surpasses it in adequacy in certain circumstances. This suggests that deep learning may have the potential to replace humans in applications where conservation of meaning is the primary aim.

# Evaluation mit Kontext: Adäquatheit

6 Nicht-Profis:

6 Profis:





**THERE MUST BE**

**SOMETHING WRONG**



Ist MT gleich gut wie HT?

*...ist das die richtige Frage?*

#	Ave %	Ave $z$	System
1	69.0	0.237	COMBO-6
	68.5	0.220	REFERENCE-HT




#	Ave %	Ave $z$	System
1	69.0	0.237	COMBO-6
	68.5	0.220	REFERENCE-HT



# OB vs. WIE

- Wir haben bislang gemessen, **ob** sich die Qualität von MT und HT unterscheidet.
- Interessanter: **Wie** unterscheidet sich MT von HT?



Wie unterscheidet  
sich MT von HT?

# Manuelle Fehler-Annotation: Resultate

MT enthält signifikant mehr

- Falsche Wörter: Semantik, Grammatik (\*\*\*)
- Weglassungen (\*)
- Wortstellungsfehler (\*\*\*)
- Falsch übersetzte Eigennamen (\*)

Läubli et al., 2020

Fisher's exact test (two-tailed). Significance levels: \* ( $p \leq .05$ ), \*\* ( $p \leq .01$ ), \*\*\* ( $p \leq .001$ ).

# Manuelle Fehler-Annotation: Resultate

MT enthält signifikant mehr

- Falsche Wörter: Semantik, Grammatik (\*\*\*)
- Weglassungen (\*)
- Wortstellungsfehler (\*\*\*)
- **Falsch übersetzte Eigennamen (\*)**

Läubli et al., 2020

Fisher's exact test (two-tailed). Significance levels: \* ( $p \leq .05$ ), \*\* ( $p \leq .01$ ), \*\*\* ( $p \leq .001$ ).

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

MT (Hassan et al., 2018):

Members of the public who find their cars obstructed by unfamiliar vehicles during their daily journeys can use the "Twitter Move Car" feature to address this distress when the driver of the unfamiliar vehicle cannot be reached. On August 11, Xi'an traffic police WeChat service number "Xi'an traffic police" launched "WeChat mobile" service. With the launch of the service, members of the public can tackle such problems in their daily lives by using the "WeChat Move" feature when an unfamiliar vehicle obstructs the movement of their vehicle while the driver is not at the scene. [...]

Human translation (Hassan et al., 2018):

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

MT (Hassan et al., 2018):

Members of the public who find their cars obstructed by unfamiliar vehicles during their daily journeys can use the "Twitter Move Car" feature to address this distress when the driver of the unfamiliar vehicle cannot be reached. On August 11, Xi'an traffic police WeChat service number "Xi'an traffic police" launched "WeChat mobile" service. With the launch of the service, members of the public can tackle such problems in their daily lives by using the "WeChat Move" feature when an unfamiliar vehicle obstructs the movement of their vehicle while the driver is not at the scene. [...]

Human translation (Hassan et al., 2018):

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

MT (Microsoft Bing, 29. Juni 2022):

In their daily travel, citizens find that their cars are hindered by strange vehicles, and in the case of not being able to contact the driver of unfamiliar vehicles, they can use the "WeChat car transfer" function to solve this problem. Since August 11, the WeChat service account of The Xi'an Traffic Police "Xi'an Traffic Police" has launched the "WeChat Car Transfer" service. After the launch of this service, in daily life, if a strange vehicle obstructs the driving of its own vehicle without the driver being present, the public can solve such problems by using the "WeChat Car Move" function. [...]

Human translation (Hassan et al., 2018):

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]

市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

MT (Google Translate, 29. Juni 2022):

Citizens find that their car is blocked by a strange vehicle during their daily travel. If they cannot contact the driver of the unfamiliar vehicle, they can use the "WeChat car move" function to solve this problem. Since August 11, the WeChat service account of Xi'an Traffic Police "Xi'an Traffic Police" has launched the "WeChat Car Moving" service. After the launch of this service, in daily life, citizens can use the "WeChat Car Move" function to solve such problems when they encounter an unfamiliar vehicle obstructing their own vehicle without the driver on the scene. [...]

Human translation (Hassan et al., 2018):

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]



市民在日常出行中,发现爱车被陌生车辆阻碍了,在联系不上陌生车辆司机的情况下,可以使用"微信挪车"功能解决这一困扰。8月11日起,西安交警微信服务号"西安交警"推出"微信挪车"服务。这项服务推出后,日常生活中,市民如遇陌生车辆在驾驶人不在现场的情况下阻碍自己车辆行驶时,就可通过使用"微信挪车"功能解决此类问题。[...]

MT (DeepL, 29. Juni 2022):

The "WeChat" service was launched on August 11 by Xi'an Traffic Police's WeChat service number "Xi'an Traffic Police". The service will allow people to use the "WeChat Move Car" function to solve problems in their daily lives when unfamiliar vehicles are blocking their vehicles in the absence of the driver. [...]

Human translation (Hassan et al., 2018):

A citizen whose car is obstructed by vehicle and is unable to contact the owner of the obstructing vehicle can use the "WeChat Move the Car" function to address the issue. The Xi'an Traffic Police WeChat official account "Xi'an Jiaojing" released the "WeChat Move the Car" service since August 11. Once the service was released, a fellow citizen whose car was obstructed by another vehicle and where the driver of the vehicle was not present, the citizen could use the "WeChat Move the Car" function to address the issue. [...]

# OB und WIE kombinieren

## Probleme:

- Zahlen alleine sind nicht (sehr) vielsagend
- Evaluation auf Dokumentenebene ist zu teuer (weniger Datenpunkte mit gleichem Budget, vgl. [Graham et al., 2019](#))

## Ziel:

- Gleichzeitig untersuchen, *ob* und *wie* HT und MT (keine) Parität erreichen

# OB und WIE kombinieren

Idee: HT und MT gleichzeitig in echtem Übersetzungs-Workflow untersuchen

- *Mixed-origin pseudo documents* (ähnlich wie Arbeit mit TMs)
- Manipulation im Experiment verhindert Rater-Bias:
  - Eindrücke von veralteter MT ([Green et al., 2013](#))
  - Negative Haltung gegenüber MT ([Läubli and Orrego-Carmona, 2017](#))

When asked, “Were the machine translations useful?”, 56% responded in the affirmative, 29% were neutral, and only 15% disagreed. One user even responded,

Your machine translations are far better than the ones of Google, Babel and so on. So they wered helpfull [sic], but usually when handed over google-translated material, I find it way easier end [sic] quicker to do it on my own from unaided.

The subjects did not know that the suggestions came from Google Translate. Users may have dated perceptions of MT quality that do not account for the rapid progress in the field.

# OB und WIE kombinieren

Idee: HT und MT gleichzeitig in echtem Übersetzungs-Workflow untersuchen

- *Mixed-origin pseudo documents* (ähnlich wie Arbeit mit TMs)
- Manipulation im Experiment verhindert Rater-Bias:
  - Veraltete Eindrücke von MT ([Green et al., 2013](#))
  - Negative Haltung gegenüber MT ([Läubli and Orrego-Carmona, 2017](#))
- Evaluations-Design kombiniert Traditionen aus HT- und MT-Forschung:
  - HT: Qualifizierte Bewerter\*innen, spezifische Rückmeldungen (Fehler), kleine Stichprobe
  - MT: Laien, vage Rückmeldungen (Zahlen), grössere Stichprobe

<b>Source (DE)</b>
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.
Voraussetzungen
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.

Source (DE)	Target (EN) - Human Translation
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This working paper is limited to the necessary functionalities for inventory management.
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The chapter on user administration is still being created.
Voraussetzungen	Requirements:
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	For the later described mutations including entry/exit, the Web API returns a correlation ID.

Source (DE)	Target (EN) – Human Translation	Target (EN) – Machine Translation
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This working paper is limited to the necessary functionalities for inventory management.	This work paper is limited to the necessary functions for portfolio management.
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The chapter on user administration is still being created.	The user administration chapter is still being prepared.
Voraussetzungen	Requirements:	Requirements
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.	The persons recorded in the payroll accounting must be fully capable of working.
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	For the later described mutations including entry/exit, the Web API returns a correlation ID.	A correlation ID is returned by the Web API for the changes described later.



Source (DE)	Target (EN)
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This work paper is limited to the necessary functions for portfolio management.
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The user administration chapter is still being prepared.
Voraussetzungen	Requirements:
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	A correlation ID is returned by the Web API for the changes described later.

<b>Source (DE)</b>	<b>Target (EN)</b>
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This work paper is limited to the necessary functions for portfolio management.
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The user administration chapter is still being prepared.
Voraussetzungen	Requirements:
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	A correlation ID is returned by the Web API for the changes described later.

Source (DE)	Target (EN)	Error(s)?		
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This work paper is limited to the necessary functions for portfolio management.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The user administration chapter is still being prepared.	Terminology Typography	Omission	<input type="checkbox"/>
Voraussetzungen	Requirements:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.	Terminology Typography	Omission	<input type="checkbox"/>
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	A correlation ID is returned by the Web API for the changes described later.	Terminology Typography	Omission	<input type="checkbox"/>

Source (DE)	Target (EN)	Error(s)?		
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This work <del>ing</del> paper is limited to the necessary functions for portfolio management.	<del>X</del>	<input type="checkbox"/>	<input type="checkbox"/>
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The <del>user administration</del> chapter <del>on user administration</del> is still being prepared.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voraussetzungen	Requirements:	<input type="checkbox"/>	<input type="checkbox"/>	<del>X</del> <input type="checkbox"/>
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.	<input type="checkbox"/>	<del>Omission</del>	<input type="checkbox"/>
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	A correlation ID <del>for the later described mutations including entry/exit</del> is returned by the Web API for the changes described later.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Editierdistanz (MED)

## Anzahl Fehler

Source (DE)	Target (EN)	Error(s)?
Dieses Arbeitspapier beschränkt sich auf die notwendigen Funktionalitäten für die Bestandsführung.	This work <del>ing</del> paper is limited to the necessary functions for portfolio management.	<input checked="" type="checkbox"/> Terminology <input type="checkbox"/> Typography <input type="checkbox"/> Omission
Das Kapitel zur Benutzerverwaltung befindet sich noch in Erstellung.	The <del>user administration</del> chapter <del>on user administration</del> is still being prepared.	<input type="checkbox"/> Terminology <input type="checkbox"/> Typography <input type="checkbox"/> Omission <input checked="" type="checkbox"/>
Voraussetzungen	Requirements:	<input type="checkbox"/> Terminology <input type="checkbox"/> Typography <input type="checkbox"/> Omission
Die in der Lohnbuchhaltung erfassten Personen müssen voll arbeitsfähig sein.	The persons entered in payroll accounting must be fully capable of working.	<input type="checkbox"/> Terminology <input type="checkbox"/> Typography <input checked="" type="checkbox"/> Omission
Für die später beschriebenen Mutationen inkl. Eintritt / Austritt wird von der Web API eine Korrelations ID zurückgegeben.	A correlation ID for the later described mutations including entry/exit is returned by the Web API for the changes described later.	<input type="checkbox"/> Terminology <input type="checkbox"/> Typography <input type="checkbox"/> Omission

# OB und WIE kombinieren: Studie

Ausgangstexte:

- Produkt- und Prozessdokumentationen aus dem Versicherungswesen

Übersetzungen:

- MT: Starkes domänenspezifisches NMT-System (Terminologie via [Dinu et al., 2019](#))
- HT: Spezialisierte Fachübersetzer\*innen ([Toral, 2020](#): 'champion translators')

Sprachpaare:

- DE→EN, DE→FR, DE→IT

# Resultate

DE → EN

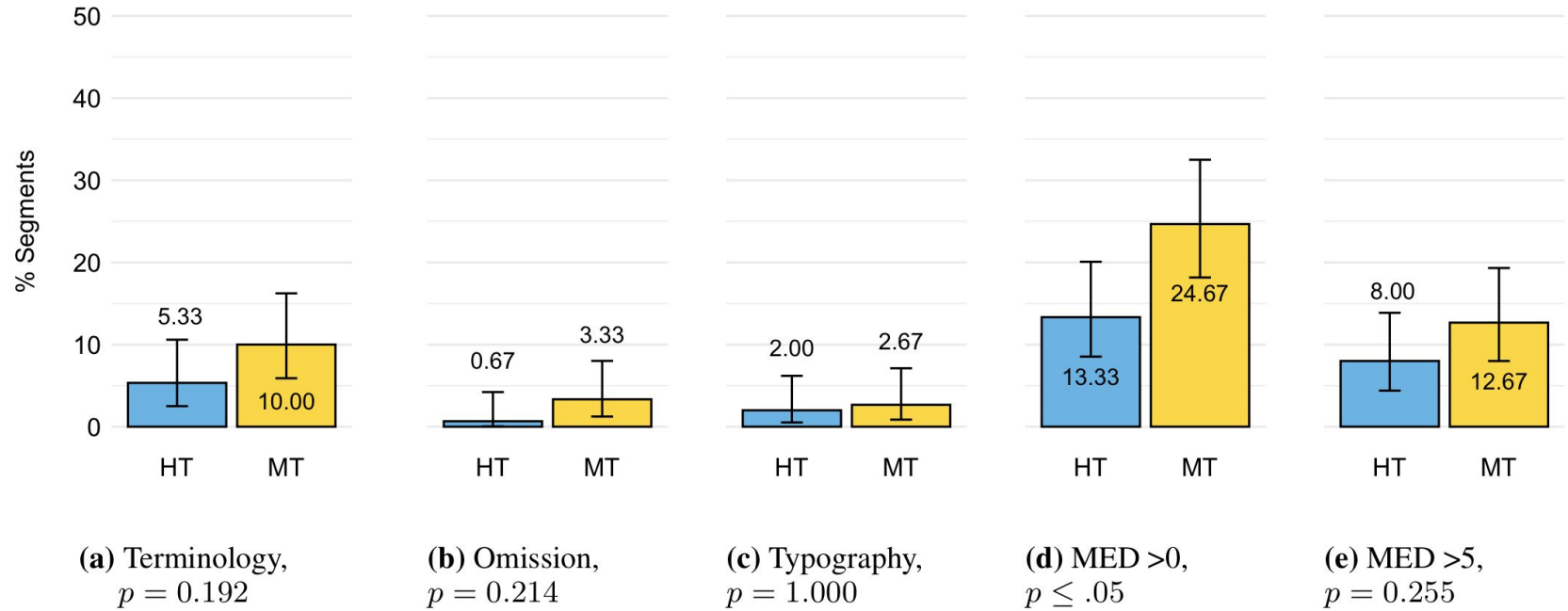


Figure 1: German–English

Quelle/Referenz: [Fischer und Läubli, 2020](#)

# Resultate

DE → FR

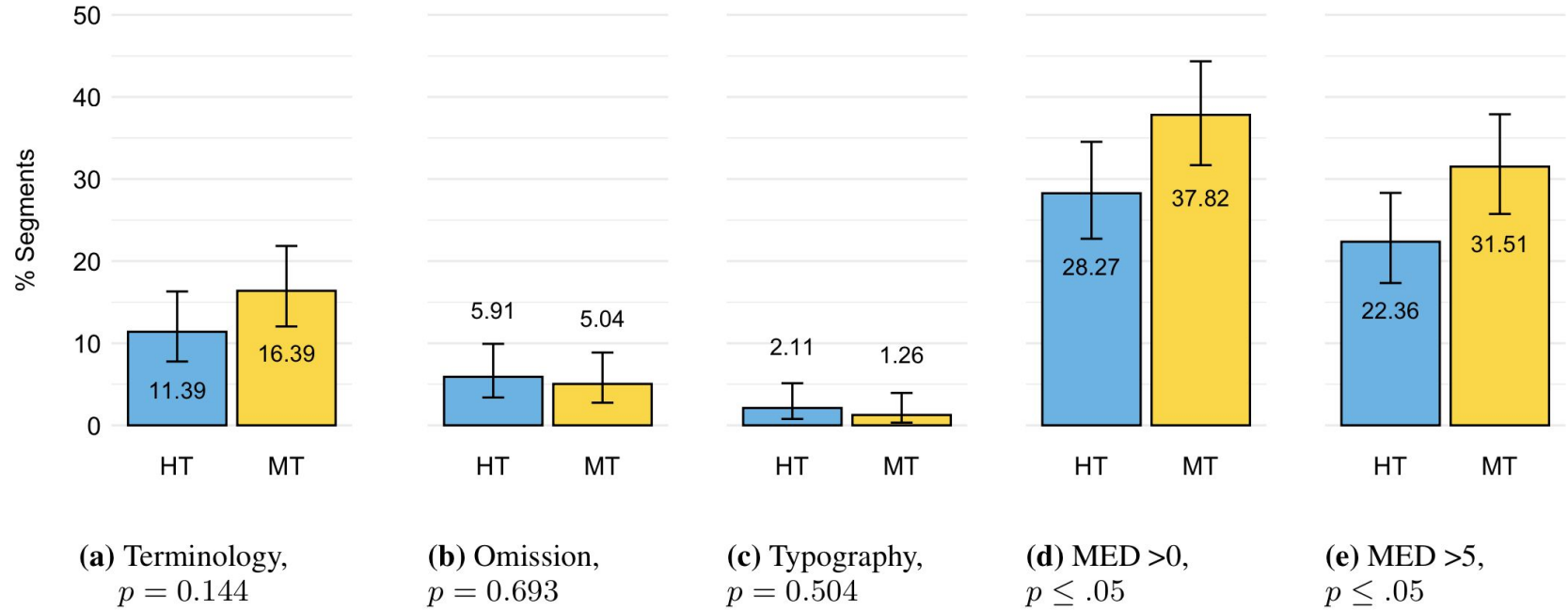


Figure 2: German–French

Quelle/Referenz: [Fischer und Läubli, 2020](#)



# Resultate

DE → IT

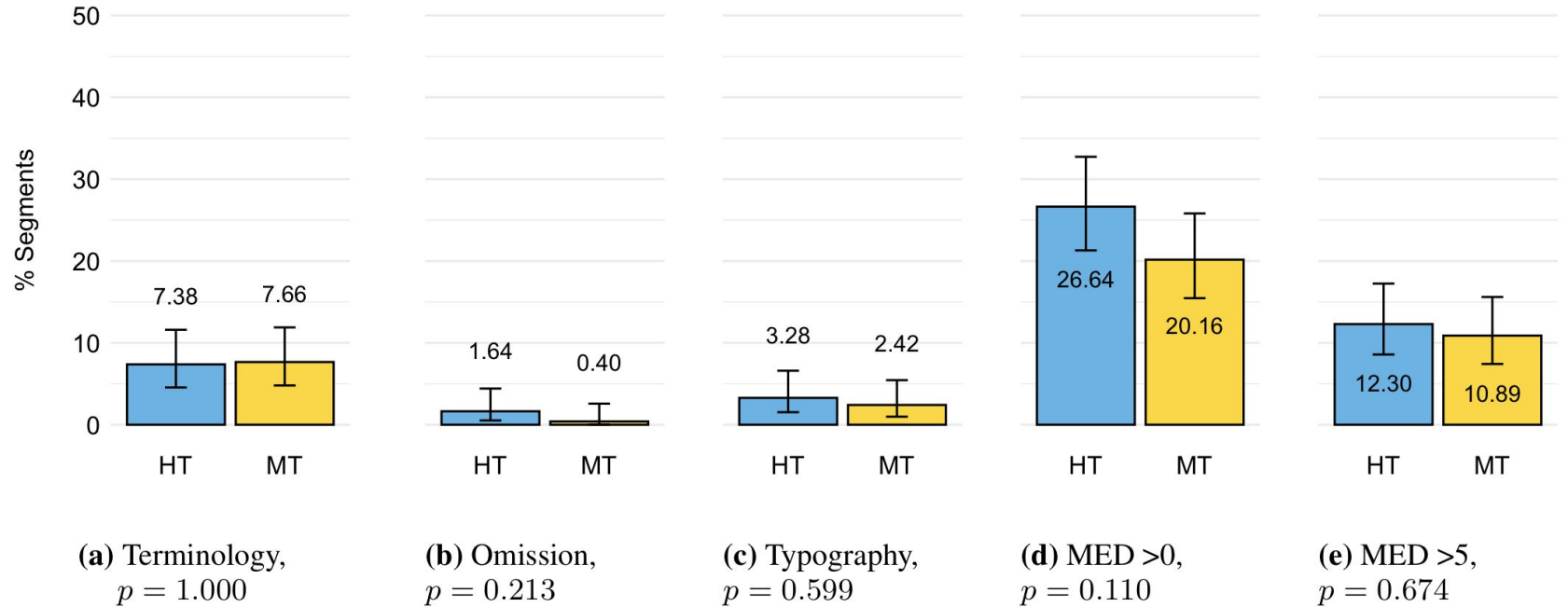


Figure 3: German-Italian

Quelle/Referenz: Fischer und Läubli, 2020

In one particular instance, a rater commented ‘NMT hat überkorrigiert’ (‘NMT has overcorrected’), when in fact the segment in question originated from HT.

# OB und WIE kombinieren: Erkenntnisse (bis dato)

- Unsere Evaluationsmethode zielt ab auf:
  - Gleiche Rigorosität bei Bewertung von HT und MT
  - Resultate produzieren, die aussagekräftiger als Werte zu Flüssigkeit und Adequatheit sind (s. auch [Popović, 2020](#))
- Unser Vergleich von domänenspezifischer MT und HT zeigt:
  - MT benötigt mehr Nachbearbeitung als HT – in 2 von 3 Sprachpaaren
  - MT enthält tendenziell mehr, aber stellenweise ähnliche viele Weglassungen, Terminologiefehler und typographische Probleme wie HT

# Vielen Dank für Ihre Aufmerksamkeit

Samuel Läubli

TextShuttle

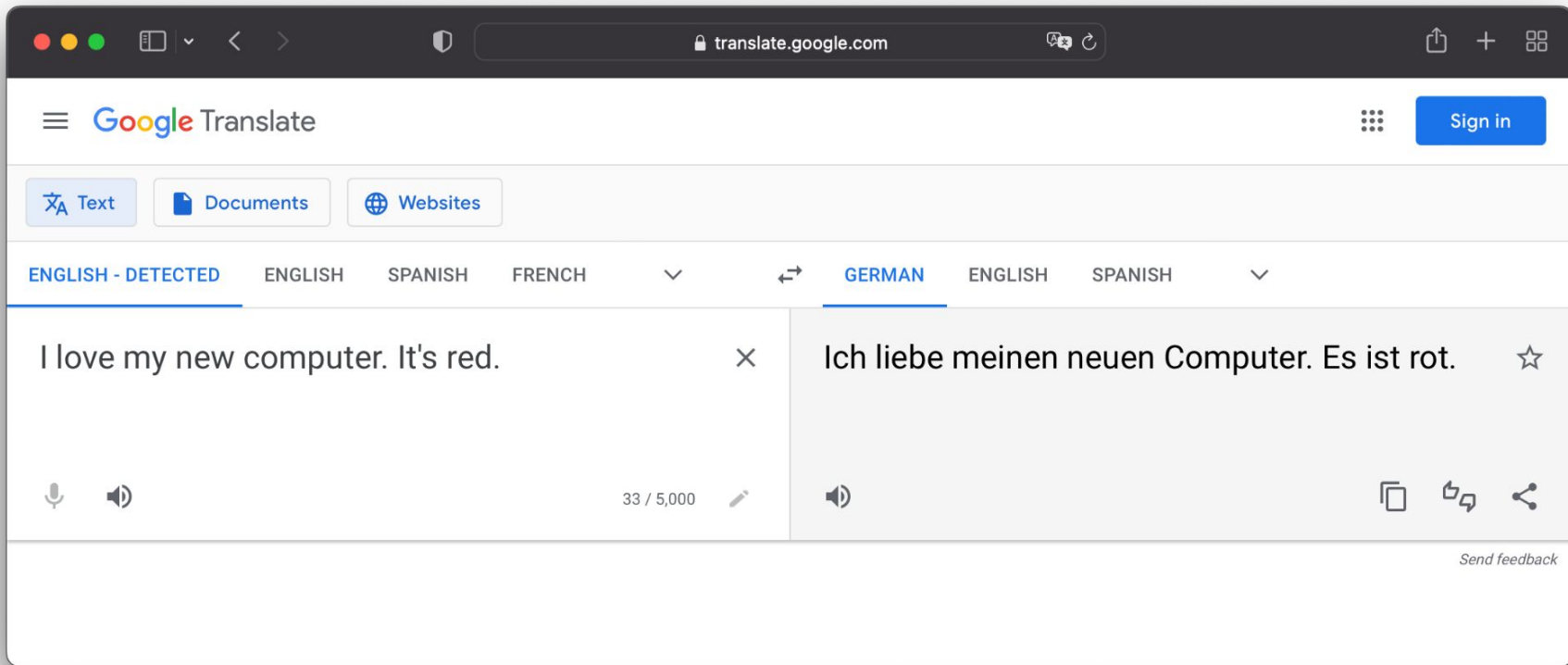
laebli@textshuttle.ai

ZHAW

laeb@zhaw.ch



@samlaeubli



Text

Documents

Websites

ENGLISH - DETECTED

ENGLISH

SPANISH

FRENCH



GERMAN

ENGLISH

SPANISH



I love my new computer. It's red.



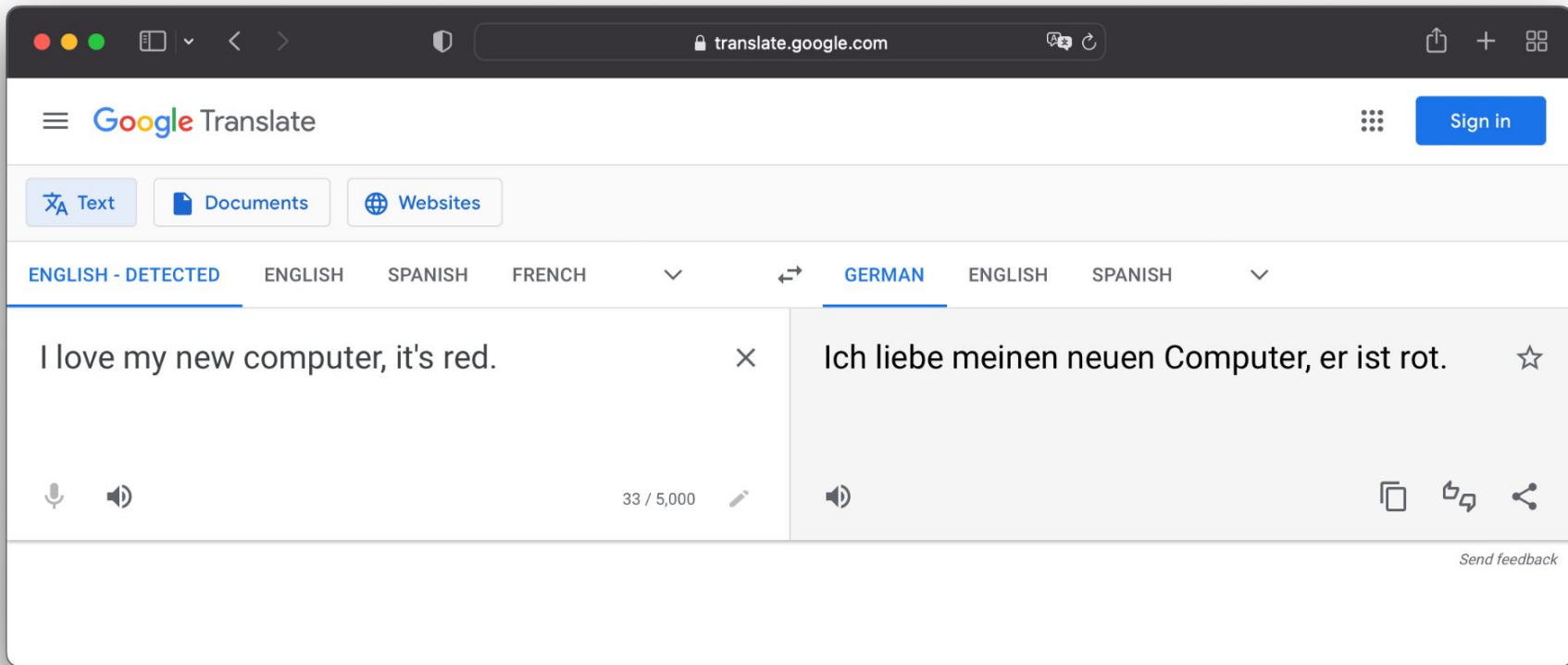
Ich liebe meinen neuen Computer. Es ist rot.



33 / 5,000



Send feedback



Text Documents Websites

ENGLISH - DETECTED ENGLISH SPANISH FRENCH GERMAN ENGLISH SPANISH

I love my new computer, it's red. X

Ich liebe meinen neuen Computer, er ist rot. ☆

33 / 5,000

Send feedback

# The Dangers of Underclaiming: Reasons for Caution When Reporting How NLP Systems Fail

Samuel R. Bowman  
New York University  
bowman@nyu.edu

## Abstract

Researchers in NLP often frame and discuss research results in ways that serve to deemphasize the field's successes, often in response to the field's widespread hype. Though well-meaning, this has yielded many misleading or false claims about the limits of our best technology. This is a problem, and it may be more serious than it looks: It harms our credibility in ways that can make it harder to mitigate present-day harms, like those involving biased systems for content moderation or resume screening. It also limits our ability to prepare for the potentially enormous impacts of more distant future advances. This paper urges researchers to be careful about these claims and suggests some research directions and communication strategies that will make it easier to avoid or rebut them.

## 1 Introduction

Over the last few years, natural language processing has seen a wave of surprising negative results overturning previously-reported success stories about what our models can do, and showing that widely-used models are surprisingly brittle (Jia

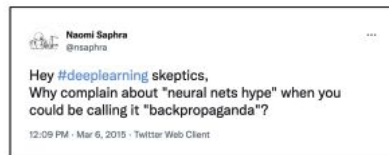


Figure 1: Hype is a problem. The opposite of hype isn't necessarily better. (Quoted with permission.)

Kiela et al., 2021; Bowman and Dahl, 2021). While we have only a limited ability to control the public narrative taking place through industry PR and the media, there's reason to be hopeful that we researchers are getting much better at avoiding the worst forms of overconfidence about our systems. Less fortunately, this pattern of disappointment seems to have led to many instances of pessimism about model performance that are ungrounded from real empirical results. This leaves room for the research community's consensus about our capabilities to fall short of our actual capabilities.

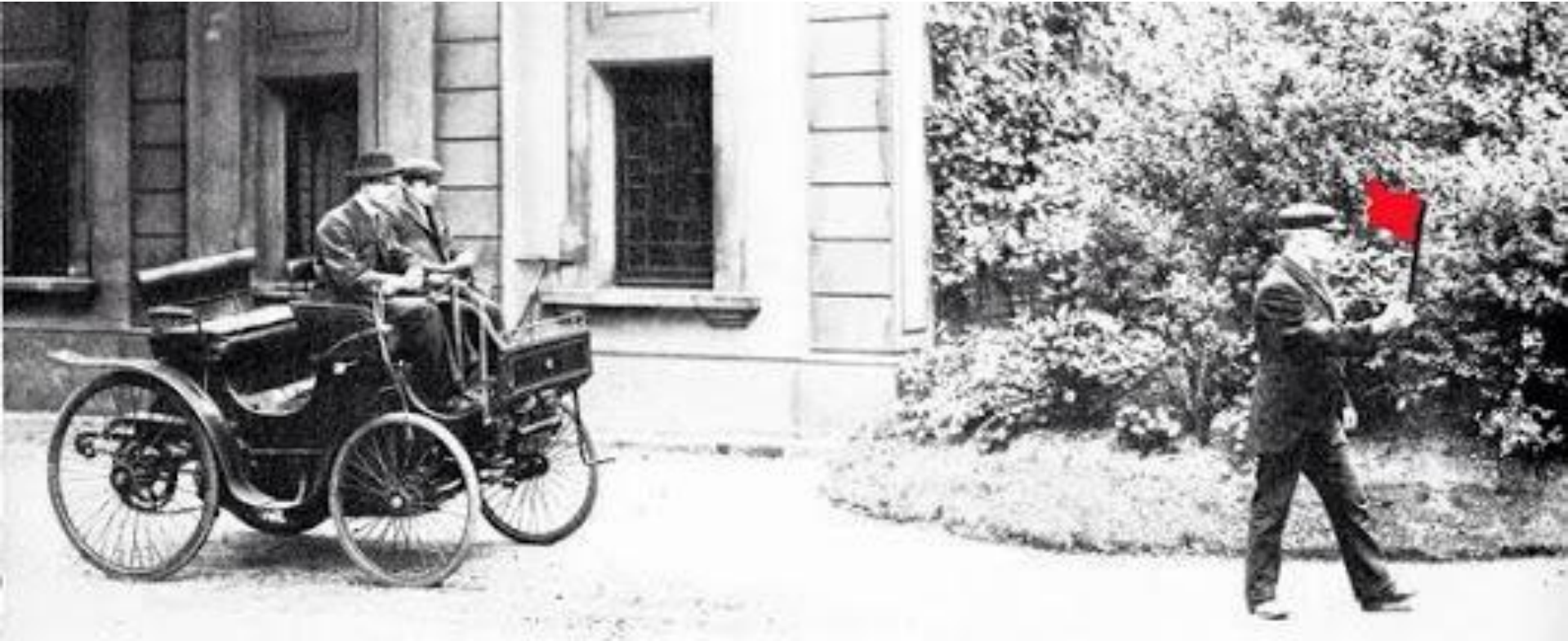
I call this issue *underclaiming*, for lack of a better term,<sup>1</sup> and argue that it is more dangerous than it might seem. It risks our credibility and thereby

“... our attempts to avoid hype often overshoot: Instead of merely correcting overly optimistic claims about our capabilities, we replace them with **overly pessimistic claims.**”

“This is a problem, and it may be more serious than it looks: It harms our credibility in ways that can make it harder to mitigate present-day harms, like those involving biased systems for content moderation or resume screening. It also **limits our ability to prepare for the potentially enormous impacts of more distant future advances.**”



# Red Flag Act (1865–1896)



# Red Flag Act (1865–1896)

- Gesetz in Grossbritannien und Irland
- Ziel: Vermeidung von Unfällen im Strassenverkehr durch «Dampfwagen»
- Bei jedem Automobil mussten zwei Personen zum Führen des Fahrzeugs anwesend sein, und ein Fussgänger hatte voraus zu laufen, der zur Warnung der Bevölkerung eine rote Flagge (red flag) tragen musste (Wikipedia).

# Pennsylvania (1896)

The most infamous of the Red Flag Laws was enacted in Pennsylvania circa 1896, when legislators unanimously passed a bill through both houses of the state legislature, which would require all motorists piloting their “horseless carriages”, upon chance encounters with cattle or livestock to

1. immediately stop the vehicle,
2. “immediately and as rapidly as possible ... disassemble the automobile”, and
3. “conceal the various components out of sight, behind nearby bushes” until equestrian or livestock is sufficiently pacified.

The law never took effect, due to a veto by the state’s governor, Daniel H. Hastings

# Pennsylvania (1896)

The most infamous of the Red Flag Laws was enacted in Pennsylvania circa 1896, when legislators unanimously passed a bill through both houses of the state legislature, which would require all motorists piloting their “horseless carriages”, upon chance encounters with cattle or livestock to

1. immediately stop the vehicle,
2. “immediately and as rapidly as possible ... disassemble the automobile”, and
3. “conceal the various components out of sight, behind nearby bushes” until equestrian or livestock is sufficiently pacified.

The law never took effect, due to a veto by the state’s governor, Daniel H. Hastings