

Zur Effizienz von neuronaler maschineller Übersetzung in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch: Eine vergleichende Untersuchung am Beispiel der Rechtssprache

The Efficiency of Neuronal Machine Translation in German/English and German/Arabic Language Pairs: A Comparative Study Using the Example of Legal Language

Aimen HAMRERAS

Universität Skikda, Algerien

a.hamreras@univ-skikda.dz

Eingangsdatum: 31/07/2023

Annahmedatum: 24/01/2024

Veröffentlichungsdatum: 30/06/2024

Zusammenfassung:

Der vorliegende wissenschaftliche Beitrag untersucht die Effizienz der neuronalen maschinellen Übersetzung in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch. Ziel ist es, eine vergleichende Untersuchung durchzuführen, um die Leistung und potenzielle Herausforderungen der neuronalen Maschinellen Übersetzung in den genannten Sprachen aufzuzeigen. Die Effizienz der neuronalen maschinellen Übersetzung wird anhand der Genauigkeit und Qualität der Übersetzungen auf syntaktischer und semantischer Ebene bewertet. Es wird untersucht, wie gut die Modelle der neuronalen Maschinellen Übersetzung in der Lage sind, die grammatikalischen Strukturen und die Terminologie der Ausgangssprache zu erfassen und in die Zielsprache zu übertragen. Zudem wird die Fähigkeit der NMT zur korrekten Übertragung von Bedeutung analysiert. Die Ergebnisse dieser Untersuchung tragen dazu bei, das Verständnis der neuronalen maschinellen Übersetzung zu erweitern und potenzielle Schwachstellen aufzuzeigen, insbesondere bei einem Vergleich der Sprachpaare Deutsch/Englisch und Deutsch/Arabisch.

Schlüsselwörter: Neuronale maschinelle Übersetzung; Rechtssprache; Deutsch; Arabisch; Englisch.

Abstract:

This scientific paper studies the efficiency of neural machine translation (NMT) in the language pairs German/English and German/Arabic. The aim is to conduct a comparative study to highlight the performance and potential challenges of NMT in the aforementioned languages. The efficiency of neural machine

translation will be evaluated based on the accuracy and quality of the translations at the syntactic and semantic levels. It will be examined how well the NMT models are able to capture the grammatical structures and terminology of the source language and transfer them into the target language. In addition, it analyses the NMT's ability to transfer meaning correctly. The results of this investigation contribute to broadening the understanding of neural machine translation and highlighting potential weaknesses, especially when comparing the language pairs German/English and German/Arabic.

Keywords: Neural Machine Translation; Legal Language; German; Arabic; English.

1. Einleitung:

Bei der Diskussion über das Thema der neuronalen maschinellen Übersetzung (im Folgenden NMT aus dem Englischen *neural machine translation*), führt kein Weg an der künstlichen Intelligenz (im Folgenden KI) vorbei. Denn erst die Entwicklung der KI machte es möglich, dass die NMT heutzutage als unabdingbare Lösung in einer durch elektronische Kommunikation geprägten und vernetzten Welt etabliert hat. Doch bevor wir uns mit dem Thema im Hauptteil beschäftigen, sollte auf den Hintergrund und die Ausgangssituation der maschinellen Übersetzung eingegangen werden. Moderne Übersetzungs- und Lokalisierungstechnologien ermöglichen es verschiedenen Akteuren, die enormen Datenmengen im Rahmen der mehrsprachigen Kommunikation zu nutzen und sich schnell weltweit zu positionieren. Insbesondere gibt es zwei Arten der maschinellen Übersetzung: die neuronale maschinelle Übersetzung und die statistische maschinelle Übersetzung, die weniger leistungsfähig ist. Die neuronale maschinelle Übersetzung wurde Ende 2014 zum ersten Mal von einem Forschungsteam von Google getestet (vgl. Lionbridge 2022, Online). Zuvor basierte die maschinelle Übersetzung auf einem statistischen Modell (Statistical Machine Translation, SMT). Die Grundlage der NMT sind künstliche neuronale Netze, die die Funktionsweise des menschlichen Gehirns nachahmen. Sie bestehen aus mehreren Schichten künstlicher Neuronen, die über

Rückkopplungsschleifen miteinander interagieren. Die eingegebenen Daten werden durch Aufmerksamkeitsprozesse gewichtet. Diese Interaktion führt zu wesentlich genaueren Ergebnissen als bei älteren Methoden. Das erklärt, warum z. B. bei der neuronalen maschinellen Übersetzung die Struktur und Funktionsweise dieser Programme die Effizienz der Übersetzung erheblich steigert. Neuronale maschinelle Übersetzung bedeutet, dass man ein neuronales System konstruiert. Es ist wie beim Klavierspielen: Wenn man einen Fehler macht, geht man einen Schritt zurück, versucht es erneut und wiederholt die Sequenz, bis sie richtig ist. NMT-Systeme verwenden selbstlernende Algorithmen, um Daten zu verknüpfen und neuronale Netze zu erstellen. Die Ergebnisse von NMT sind noch nicht perfekt, z. B. bei der Übersetzung von sehr technischen Inhalten. NMT ist jedoch keine Ausnahme: Kein MT-System könnte einen Text mit vielen Fachbegriffen und Abkürzungen perfekt übersetzen. Diese Schwäche lässt sich jedoch durch gezieltes Post-Editing ausgleichen (vgl. Lionbridge, Online). Während das Thema in Sprachen wie dem Englischen, dem Französischen und dem Deutschen weitreichend erforscht wurde, bleibt der Forschungsstand für die arabische Sprache unzureichend, gemessen an dem Stellenwert, der dieser von mehr als 300 Millionen Menschen als Muttersprache gesprochenen Sprache zukommt, weswegen man annehmen könnte, dass die NMT der arabischen Sprache im Vergleich zu Englisch und Französisch noch nicht ausgereift ist und keine auf syntaktischer und semantischer Ebene zufriedenstellende Ergebnisse liefern kann. Davon ausgehend besteht das Ziel dieses Beitrags darin, die Effizienz von neuronaler maschineller Übersetzung auf syntaktischer und semantischer Ebene in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch zu erforschen, um die Leistungsfähigkeit der NMT in der arabischen Sprache ans Tageslicht zu bringen und potenzielle Schwachstellen bzw. Verbesserungsmöglichkeiten aufzuzeigen. Der vorliegende Beitrag versucht auf folgende Frage Antwort zu geben: Wie gut schneidet die NMT von verschiedenen Anbietern in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch am Beispiel der Rechtssprache ab und worauf ist dies zurückzuführen? Dabei geht man von der Hypothese aus, dass die NMT im Sprachpaar Deutsch/Englisch effizienter ist und bessere Ergebnisse liefert,

weil dieses Gebiet in der deutschen bzw. der englischen Sprache besser erforscht ist und es genug Daten zum Trainieren von NMT-Systemen zur Verfügung stehen.

2. Zur maschinellen Übersetzung

Unter maschineller Übersetzung versteht man den Einsatz von automatisierter Software, die Texte ohne menschliches Zutun übersetzt. Adaptive maschinelle Übersetzung ist eine lernfähige Technologie, die sich in Echtzeit an menschliches Feedback anpasst. Maschinelle Übersetzung basiert auf Wahrscheinlichkeiten - nicht auf der Bedeutung. Sie versteht weder die Bedeutung noch den Kontext des zu übersetzenden Textes. Wenn man also nicht beide Sprachen lesen kann, wird man nie erfahren, ob die Maschine richtig geraten hat. Anwendungen der maschinellen Übersetzung können flüssige Ergebnisse liefern, aber flüssig bedeutet nicht genau. Nur weil es richtig aussieht, heißt es nicht, dass es richtig ist. Laut Microsoft handelt es sich bei maschineller Übersetzung um Anwendungen oder Online-Dienste, die sich maschineller Lerntechnologien bedienen, um große Datentexten aus und in eine ihrerunterstützten Sprachen zu übersetzen. Somit wird ein sogenannter Source-Text in eine Zielsprache übersetzt (vgl. Microsoft, Online). Microsoft setzt fort, dass die Konzepte hinter der maschinellen Übersetzung zwar einfach sind, es stecken aber dahinter äußerst komplizierte Technologien, wie das Deep Learning, Big Data, Linguistik, Cloud Computing und Web-APIs. Maschinelle Übersetzung wäre somit nicht möglich, wenn es all diese Fortschritte nicht gegeben hätte. Und mit Fortschritten ist insbesondere die sogenannte künstliche Intelligenz (KI) gemeint. Laut Kirste/Schürholz wird mit KI traditionell ein Teilgebiet der Informatik bezeichnet, das sich mit der Automatisierung von intelligentem Verhalten befasst. Die Autoren weisen aber darauf hin, dass eine genaue Begriffsbestimmung jedoch kaum möglich ist, da auch alle direkt verwandten Wissenschaften wie Psychologie, Biologie, Kognitionswissenschaft und Neurowissenschaft an einer genauen Definition von Intelligenz scheitern. Sie setzen fort, dass die Versuche zur Beschreibung und Nachbildung von Intelligenz in vier Ansätzen unterteilt werden können, und zwar nach menschlichem Denken, menschlichem

Handeln, rationalem Denken und rationalem Handeln (2019, S. 21 nach Russell et al. 2010). Als Beispiel nennen die Autoren den berühmten Turing-Test, der im Bereich des menschlichen Handelns zu klassifizieren ist, und moderne Programme zur Bilderkennung und damit verbundene Entscheidungen, die in den Bereich des rationalen Handelns gehören. Zusammengefasst lässt sich die KI als Versuch verstehen, ein System zu entwickeln, das eigenständig komplexe Probleme lösen kann. Und wenn wir diesen Aspekt auf die maschinelle Übersetzung anwenden wollen, dann brauchen wir ein ausreichend intelligentes System, das in der Lage ist, menschliches Denken und Handeln nachzubilden, sodass im Endeffekt die Übersetzungsqualität einer Maschine bzw. einer Software nicht von der eines kompetenten Übersetzers abweicht, oder diese sogar übertrifft, wenn man bedenkt, dass der Mensch keinesfalls seine maximale Performanz, im Sinne von Chomsky, erreichen kann, und dass er immer wieder bestimmten Einflussfaktoren ausgesetzt ist, wie Vergesslichkeit, Konzentrationsmangel, Müdigkeit, usw. Aber maschinelle Übersetzung ist nicht gleich maschinelle Übersetzung. Man unterscheidet hier zwischen "statistischer maschineller Übersetzung (SMT für Englisch Statistical Machine Translation) und neuronaler maschineller Übersetzung" (NMT für Englisch Neural Machine Translation). Bei der statistischen maschinellen Übersetzung wird maschinelles Lernen eingesetzt, um eine Vielzahl möglicher Übersetzungen für einen Satz zu erstellen. Anschließend wird die beste Übersetzung ausgewählt - basierend auf der Wahrscheinlichkeit, mit der Wörter und Sätze in der Ausgangs- und Zielsprache vorkommen. Im Gegensatz zur regelbasierten maschinellen Übersetzung folgt die SMT also keinen starren Regeln, die bei einer fehlerhaften Übersetzung einzeln angepasst werden müssten. Vielmehr ergeben sich die Übersetzungen aus den empirischen Gegebenheiten der jeweiligen Sprachen. Die SMT hat die Art und Weise, wie maschinelle Übersetzung Begriffe übersetzt, grundlegend verändert (vgl. Lionbridge 2022: Online).

3. Begriffsbestimmung: Neuronale maschinelle Übersetzung

Neuronale maschinelle Übersetzung ist ein Ansatz zur maschinellen Übersetzung, der auf der Verwendung von neuronalen Netzen basiert (vgl.

Lionbridge 2022, Online). Dieser Ansatz hat in den letzten Jahren aufgrund seiner Leistung und Effektivität an Bedeutung gewonnen. Im Gegensatz zu anderen Ansätzen zur maschinellen Übersetzung basiert die neuronale maschinelle Übersetzung nicht auf einer expliziten Programmierung von Regeln oder auf der Verwendung von statistischen Modellen. Stattdessen nutzt sie ein künstliches neuronales Netzwerk, das auf der Verarbeitung von Sprachpaaren basiert, um die Übersetzung von Texten zwischen verschiedenen Sprachen zu ermöglichen. Das neuronale Netzwerk besteht aus Schichten von Neuronen, die Informationen aufnehmen und verarbeiten, um eine Übersetzung auszugeben. Das Netzwerk wird zuvor auf großen Mengen von parallelisierten Korpora trainiert, um eine optimale Leistung zu erzielen. Während des Trainings lernt das neuronale Netzwerk, die Beziehungen zwischen den Wörtern und Sätzen in den beiden Sprachen zu verstehen und eine effektive Übersetzung zu produzieren. Die neuronale maschinelle Übersetzung hat gegenüber anderen Ansätzen mehrere Vorteile. Zum einen ist sie in der Lage, kontextuelle Informationen zu berücksichtigen, um eine genauere Übersetzung zu erzielen. Zum anderen kann sie sich an verschiedene Arten von Texten anpassen, einschließlich Fachterminologie und umgangssprachlichen Ausdrücken. Darüber hinaus hat sie eine höhere Leistung als andere Ansätze zur maschinellen Übersetzung, insbesondere bei der Übersetzung komplexer Sätze und Texte. Obwohl die neuronale maschinelle Übersetzung eine vielversprechende Technologie ist, gibt es immer noch Herausforderungen, die angegangen werden müssen. Eine dieser Herausforderungen ist die Verfügbarkeit von hochwertigen, parallelisierten Korpora, die für das Training von neuronalen Netzwerken verwendet werden können. Ein weiteres Problem ist die Beibehaltung der Kohärenz und des Kontextes bei der Übersetzung von langen Texten. Insgesamt ist die neuronale maschinelle Übersetzung ein vielversprechender Ansatz zur maschinellen Übersetzung, der dazu beitragen kann, die Sprachbarriere zwischen verschiedenen Kulturen und Ländern zu überwinden. Durch weitere Forschung und Entwicklung können die Leistung und Effektivität von neuronalen maschinellen Übersetzungssystemen weiter verbessert werden. Bei der Diskussion über die neuronale maschinelle Übersetzung kommen einem vielleicht die

verschiedenen Online-Tools zur sofortigen Übersetzung wie Google translate oder Deepl in den Sinn, und zwar zu recht. Denn im Jahr 2016 führte Google das erste auf einem neuronalen Netz basierende Übersetzungssystem ein, das eine große Medienresonanz hatte. Der Internetriese hat sich dabei zum Ziel gesetzt, die Lücke zwischen menschlicher und maschineller Übersetzung zu schließen und eine möglichst natürliche Übersetzung zu liefern (Schmalz 2019:194), zit. nach Yonghui Wu et al. 2016). Laut Cohen (2017:5) kann man die Entwicklung von neuronaler maschineller Übersetzung bis in die 80er Jahre zurückverfolgen, und zwar während der letzten Forschungswelle in Sachen neuronaler Netzwerke. Dazu schreibt er:

In fact, the models proposed by Forcada and Ñeco (1997) and Castaño et al. (1997) are striking similar to the current dominant neural machine translation approaches. However, none of these models were trained on data sizes large enough to produce reasonable results for anything but toy examples. The computational complexity involved by far exceeded the computational resources of that era, and hence the idea was abandoned for almost two decades. Cohen (2017:5)

Aus vorstehendem Zitat geht hervor, dass in den 90er Jahren zwar ernsthafte Versuche im Bereich neuronaler maschineller Übersetzung unternommen wurden, die EDV-Ressourcen bzw. Hard- und Softwareressourcen jedoch weitgehend überfordert waren, sodass die Idee zwei Jahrzehnte lang auf Eis gelegt wurde.

4. Anwendungsbereiche von neuronaler maschineller Übersetzung

Auf die neuronale maschinelle Übersetzung (NMT) wird in verschiedenen Bereichen zurückgegriffen. Ihre Anwendungsbereiche reichen von der Textübersetzung in der Wirtschaft und im öffentlichen Leben wie etwas bei Unternehmen und Behörden. Zum Beispiel nutzen Unternehmen die NMT, um Unterlagen, E-Mails und Verträge in eine

beliebige Sprache zu übersetzen. So können sie schnell und effizient mit ihren Kunden und Geschäftspartnern auf der ganzen Welt kommunizieren und ihre Geschäftsaktivitäten auf globale Märkte ausweiten. Auf die NMT greifen aber auch Regierungsbehörden zurück, um z. B. offizielle Dokumente in verschiedene Sprachen zu erstellen. Besonders wenn es darum geht, schnelle und effiziente Übersetzungen zu gewährleisten, bei denen sonst menschliche Übersetzer mehr Zeit benötigen, die man nicht hat. Die Medizin stellt einen weiteren Anwendungsbereich der NMT dar. Insbesondere kann die NMT dabei helfen, Kommunikationsbarrieren zwischen Ärzten und Patienten zu überwinden. Darüber hinaus können Ärzte und Patienten durch die Übersetzung medizinischer Texte und Anweisungen in verschiedene Sprachen besser miteinander kommunizieren und sicherstellen, dass wichtige Informationen richtig verstanden werden. Auch im Bereich des Tourismus kann NMT dazu beitragen, die Kommunikation für Touristen zu erleichtern. Zum Beispiel ist es durch die NMT möglich, eine effiziente Übersetzung von Reiseinformationen, Beschilderungen und Speisekarten in verschiedene Sprachen zu erstellen. Dadurch können sich Touristen besser orientieren und ihren Aufenthalt genießen. Auch Online-Plattformen wie E-Commerce-Seiten, soziale Medien und Nachrichtenseiten können von der NMT profitieren, indem sie Inhalte in verschiedenen Sprachen anbieten und so ein größeres Publikum erreichen. Die nMT kann dazu beitragen, die Interaktion zwischen Nutzern aus verschiedenen Ländern und Sprachgruppen zu erleichtern und die Reichweite von Online-Plattformen zu erhöhen. Insgesamt gibt es viele Anwendungsbereiche für die NMT, und es ist zu erwarten, dass ihr Einsatz in Zukunft weiter zunehmen wird, da die Globalisierung fortschreitet und die verschiedenen Sprachgemeinschaften in dieser immer globaler werdenden Weltwirtschaft immer stärker aufeinander angewiesen sind.

5. Zustand der NMT für die arabische Sprache

Die neuronale maschinelle Übersetzung hat auch bei der Übersetzung der arabischen Sprache in den letzten Jahren große Fortschritte gemacht. Zumal mit der großen Migrationswelle nach Deutschland der Bedarf an schneller und effizienter Übersetzung in der Sprachkombination

Deutsch-Arabisch größer geworden ist, was u.a. Regierungsbehörden und Unternehmen vor große Herausforderungen stellte, nämlich die Kommunikation mit arabischen Muttersprachlern sicherzustellen, und zwar kurzfristig. Die arabische Sprache stellt jedoch aufgrund ihrer Besonderheiten und ihres komplexen Schreibsystems eine Herausforderung für die NMT dar. In Bezug auf den aktuellen Stand der NMT für die arabische Sprache gibt es eine wachsende Zahl von Forschungsarbeiten, die sich auf die Verbesserung der Qualität und Effizienz von NMT-Systemen für die arabische Sprache konzentrieren. Einige dieser Arbeiten beschäftigen sich mit der Integration von auf Arabisch spezialisierten Sprachmodellen, um die Genauigkeit der Übersetzungen zu verbessern. Andere arbeiten an der Verbesserung der Textvorverarbeitung, um Fehler im Schreibsystem der arabischen Sprache zu reduzieren. Zu den Forschungsarbeiten nennen wir den wissenschaftlichen Beitrag von Amjad Almahairi et al. aus dem Jahr 2016 zum Thema „First Result on Arabic Neural Machine Translation“. Darin stellen die Autoren fest, dass sich ein Großteil der Forschung im Bereich der neuronalen maschinellen Übersetzung auf europäische Sprachen konzentriert hat, obwohl NMT grundsätzlich sprachunabhängig ist. In ihrem Beitrag haben sie die neuronale maschinelle Übersetzung auf die Übersetzung in der Sprachkombination (Ar<->En) angewendet und sie mit einem standardmäßigen phrasenbasierten Übersetzungssystem verglichen. Sie führten einen umfassenden Vergleich mit verschiedenen Konfigurationen für die Vorverarbeitung der arabischen Schrift durch und zeigten, dass die phrasenbasierten und neuronalen Übersetzungssysteme vergleichbare Leistungen erbringen und dass eine angemessene Vorverarbeitung der arabischen Schrift einen ähnlichen Effekt auf beide Systeme hat. Sie stellen jedoch fest, dass die neuronale maschinelle Übersetzung das phrasenbasierte System in einem Testsatz außerhalb der Domäne deutlich übertrifft, was sie für den Einsatz in der Praxis attraktiver macht (Almahairi et al. 2016). Es gibt eine Reihe von Unternehmen und Forschungseinrichtungen, die sich auf die Entwicklung von NMT-Systemen für Arabisch spezialisiert haben. Hier sind einige Beispiele:

SYSTRAN ist ein Unternehmen, das sich auf die Entwicklung von Übersetzungssoftware und -diensten spezialisiert hat. Das Unternehmen bietet auch ein NMT-System für Arabisch an, das auf der Verarbeitung natürlicher Sprache basiert und mit einem neuronalen Netzwerk arbeitet. Das NMT-System von SYSTRAN wird kontinuierlich verbessert und optimiert, um die Qualität und Genauigkeit der Übersetzungen zu erhöhen.

Qatar Computing Research Institute (QCRI) ist eine Forschungseinrichtung, die sich auf die Entwicklung von KI-Systemen spezialisiert hat. Die Forscher bei QCRI haben verschiedene Arbeiten zur Verbesserung von NMT-Systemen für Arabisch durchgeführt. Zum Beispiel hat ein Team von QCRI-Forschern ein neues Sprachmodell entwickelt, das auf Arabisch spezialisiert ist und die Qualität der Übersetzungen verbessert.

IBM ist ein Technologieunternehmen, das auch an der Entwicklung von NMT-Systemen für Arabisch beteiligt ist. Das Unternehmen hat eine Reihe von Arbeiten zur Verbesserung der Genauigkeit und Geschwindigkeit von NMT-Systemen durchgeführt. IBM hat auch eine NMT-Plattform namens *Watson Language Translator* entwickelt, die Übersetzungen in verschiedene Sprachen, einschließlich Arabisch, anbietet.

University of Maryland hat auch Forschungsarbeiten zur Verbesserung von NMT-Systemen für Arabisch durchgeführt. Ein Beispiel ist das Werkzeug "*Arabic Diacritic Restoration*", das von Forschern der Universität entwickelt wurde, um die Schreibweise von Wörtern im Arabischen zu korrigieren und so die Genauigkeit von NMT-Systemen zu verbessern.

6. NMT in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch

Unter diesem Punkt geht es darum, NMT-Systeme bei den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch zu vergleichen und herauszufinden, ob die NMT bei Arabisch als Zielsprache genauso effizient und zuverlässig ist wie bei anderen länger etablierten NMT-Systemen, wie dem NMT-System des Englischen, da beide Sprachpaare unterschiedliche Herausforderungen für die NMT-Systeme darstellen. Aus theoretischen Erkenntnissen und Feldbeobachtungen geht hervor, dass das Sprachpaar

Deutsch/Englisch im Allgemeinen besser erforscht ist. Darüber hinaus gibt es hier mehr Trainingsdaten für NMT-Systeme, z. B. aus dem technischen, dem wissenschaftlichen und dem rechtlichen Bereich. Das bedeutet, dass NMT-Systeme für Deutsch/Englisch in der Regel eine höhere Genauigkeit und Effizienz aufweisen. Es gibt auch mehr Unternehmen und Forschungseinrichtungen, die sich auf die Entwicklung von NMT-Systemen für das Sprachpaar Deutsch/Englisch spezialisiert haben, was die Verfügbarkeit und Qualität dieser Systeme verbessert. Das Sprachpaar Deutsch/Arabisch stellt jedoch aufgrund der Besonderheiten der arabischen Sprache, wie des komplexen Schriftsystems und der Grammatik, eine größere Herausforderung für NMT-Systeme dar. Es gibt auch weniger Trainingsdaten für NMT-Systeme für Deutsch/Arabisch, was die Genauigkeit und Effizienz dieser Systeme beeinträchtigen kann. Zum Beispiel findet man wenig technische, wissenschaftliche oder rechtliche Texte, die als Produkt einer NMT-basierten Übersetzung entstanden sind. Im rechtlichen Bereich wird davon ausgegangen, dass NMT-Systeme für das Sprachpaar Deutsch/Englisch viel bessere Ergebnisse liefern können, z. B. bei der Übersetzung von Verträgen, Urteilen oder anderen juristischen Dokumenten. In der Sprachkombination Deutsch/Arabisch wird angenommen, dass man trotz kleinerer Genauigkeit zufriedenstellende Ergebnisse erwarten kann, wenn es um die Übersetzung von einfachen Rechtstexten geht, zumindest auf semantischer Ebene. Sollten aber Fachtexte mit einem höheren Grad an Fachlichkeit übersetzt werden, so muss man über entsprechende Daten verfügen, um das NMT-System entsprechend zu trainieren. Dabei muss darauf hingewiesen werden, dass die NMT im Sprachpaar DE/EN nach dem aktuellen Forschungsstand viel besser abschneidet, und zwar dank des großen Bestandes an parallelisierten Texten, über den man bereits verfügt und mit dem die NMT-Systeme besser trainiert werden können. Auf der anderen Seite und aufgrund des bescheidenen Bestandes an parallelisierten Texten im Sprachpaar DE/AR, verfügt man nicht über genug Trainingsdaten für NMT-Systeme, um eine effiziente und qualitative Übersetzung zu gewährleisten. Diese Annahme wird anhand der nachstehenden Beispiele mit authentischen Texten aus der Rechtssprache geprüft.

6.1 Praxisbeispiele aus der Rechtssprache

Im Folgenden werden anhand der NMT-basierten Übersetzungsdienste Deepl und Systran einige Sätze übersetzt, die aus Verträgen, Allgemeinen Geschäftsbedingungen und Gesetzestexten stammen. Direkt darunter wird die jeweilige menschliche Originalübersetzung zu Vergleichszwecken mit der neuronalen maschinellen Übersetzung aufgeführt. Im Anschluss werden dieselben Sätze mithilfe des NMT-Systems Systran sowie der KI ChatGPT ins Arabische übersetzt, weil die Sprachkombination Deutsch-Arabisch noch nicht von Deepl angeboten wird.

Ausgangssprache Deutsch	Zielsprache Englisch
<p>AllezwischendemHerstellerunddemKundenimZusammenhangmitdemReservierungsvertraggetroffenen Vereinbarungergebensichinsbesondereausdiesen Reservierungsbedingungen in Ihrer zum Zeitpunkt der Reservierung gültigen Fassung, unserer schriftlichen Eingangsbestätigung und unserer Annahmeerklärung. (AGBs der Evetta GmbH).</p>	<p>Systran: All agreements between the manufacturer and the customer in connection with the reservation contract are based in particular on these reservation conditions in your version valid at the time of the reservation, our written confirmation of receipt and our acceptance declaration. (General Terms and Conditions of Evetta GmbH).</p> <p>Deepl: All agreements made between the manufacturer and the customer in connection with the reservation contract are based in</p>

	<p>particular on these terms and conditions of reservation in the version valid at the time of the reservation, our written confirmation of receipt and our declaration of acceptance. (General Terms and Conditions of Evetta GmbH).</p> <p>Human translation: All agreements made between the Manufacturer and the Customer in connection with the reservation contract result in particular from these reservation conditions in their version valid at the time of the reservation, our written confirmation of receipt and our declaration of acceptance. (General Terms and Conditions of Evetta GmbH)</p>
<p>Abweichende allgemeine Vertragsbedingungen des Kunden finden auf die Reservierung keine Anwendung, es sei denn, wir haben deren Anwendbarkeit schriftlich zugestimmt. (AGBs der Evetta GmbH).</p>	<p>Systran: Different general terms and conditions of the customer shall not apply to the reservation, unless we have agreed to their applicability in writing. (General Terms and Conditions of Evetta</p>

	<p>GmbH).</p> <p>Deepl/Human translation: Deviating general terms and conditions of the customer shall not apply to the reservation unless we have agreed to their applicability in writing. (General Terms and Conditions of Evetta GmbH).</p>
<p>Verbraucher ist jede natürliche Person, die ein Rechtsgeschäft zu Zwecken abschließt, die überwiegend weder ihrer gewerblichen noch ihrer selbständigen beruflichen Tätigkeit zugerechnet werden können. (§ 13 BGB)</p>	<p>Systran: Consumers shall mean any natural person who enters into a transaction for purposes which, for the most part, cannot be attributed to his trade or to his self-employed professional activity. (§ 13 BGB)</p> <p>Deepl: A consumer is any natural person who enters into a legal transaction for purposes that can predominantly be attributed neither to his or her commercial nor to his or her independent professional activity. (§ 13 BGB)</p>

	<p>Human translation: A consumer means every natural person who enters into a legal transaction for purposes that predominantly are outside the consumer's trade, business or profession.</p>
<p>Ein Verein, dessen Zweck auf einen wirtschaftlichen Geschäftsbetrieb gerichtet ist, erlangt in Ermangelung besonderer bundesgesetzlicher Vorschriften Rechtsfähigkeit durch staatliche Verleihung. Die Verleihung steht dem Land zu, in dessen Gebiet der Verein seinen Sitz hat. (§ 22 BGB)</p>	<p>Systran: In the absence of specific federal legislation, an association whose purpose is to conduct a commercial business shall acquire legal capacity through the award of a public contract. The award is to be given to the country in whose territory the association is established. (§ 22 BGB)</p> <p>Deepl: In the absence of special federal regulations, an association whose purpose is directed towards economic business operations acquires legal capacity through state conferral. The award shall be granted to the Land in whose territory the association has its seat. (§ 22 BGB)</p>

	<p>Human translation: An association the object of which consists of commercial business operations acquires legal personality, for lack of special provisions under federal law, by state grant. The grant is in the power of the Land in the territory of which the association has its seat.</p>
<p>Diese Verordnung gilt für die ganz oder teilweise automatisierte Verarbeitung personenbezogener Daten sowie für die nichtautomatisierte Verarbeitung personenbezogener Daten, die in einem Dateisystem gespeichert sind oder gespeichert werden sollen. (Artikel 2 der Datenschutz-Grundverordnung der Europäischen Union)</p>	<p>Systran: This Regulation shall apply to the processing of personal data in whole or in part by means of automated processing and to the non-automated processing of personal data stored or intended to be stored in a file system. (Article 2 of the General Data Protection Regulation of the European Union)</p> <p>Deepl: This Regulation applies to the processing of personal data wholly or partly by automatic means and to the processing otherwise than by automatic means of personal data which are</p>

	<p>stored or are intended to be stored in a filing system. (Article 2 of the General Data Protection Regulation of the European Union)</p> <p>Human translation: This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system. (Article 2 of the General Data Protection Regulation in the European Union)</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Ausgangssprache Deutsch	Zielsprache Arabisch
<p>AllezwischendemHerstellerunddemKundenimZusammenhangmitdemReservierungsvertraggetroffenen Vereinbarungergebensichinsbesondereausdiesen Reservierungsbedingungen in Ihrer zum Zeitpunkt der Reservierung gültigen Fassung, unserer schriftlichen Eingangsbestätigung und unserer Annahmeerklärung. (AGBs eines Automobilherstellers)</p>	<p>Systran</p> <p>وجميع الاتفاقات المبرمة بين الشركة المصنعة والعميل فيما يتعلق بعقد الحجز هي بوجه خاص نتيجة لشروط الحجز هذه في نسختها صالحة وقت الحجز، وتأكيدنا الكتابي بالاستلام، وإعلاننا بقبول العقد.</p> <p>Chatgpt</p> <p>جميع الاتفاقات التي تمت بين الشركة المصنعة والعميل فيما</p>

	<p>يتعلق بعقد الحجز تتبع بشكل خاص من هذه الشروط والأحكام للحجز بالنسخة السارية في وقت الحجز، وتأكيد الاستلام الخطي الخاص بنا وإشعار قبولنا.</p> <p>الترجمة البشرية:</p> <p>جميع الاتفاقيات التي تم إبرامها بين المصنع والعميل والمتعلقة بعقد الحجز تخضع بشكل خاص إلى شروط الحجز في النسخة السارية وقت القيام بالحجز، بالإضافة إلى التأكيد الكتابي لاستلام الحجز وإشعار القبول من طرفنا.</p>
<p>Abweichende allgemeine Vertragsbedingungen des Kunden finden auf die Reservierung keine Anwendung, es sei denn, wir haben deren Anwendbarkeit schriftlich zugestimmt. (AGBs eines Reisebüros)</p>	<p>Systran</p> <p>ولا ينطبق انحراف الشروط والأحكام العامة للعميل على التحفظ، ما لم نتفق على انطباقه كتابة.</p> <p>Chatgpt</p> <p>لا تطبق الشروط العامة للعقد المختلفة للعميل على الحجز، ما لم نوافق على تطبيقها بشكل خطي.</p> <p>الترجمة البشرية:</p> <p>لا يخضع الحجز لشروط العميل التي تختلف عن الشروط المنصوص عليها هنا، إلا إذا تم الموافقة عليها كتابيا.</p>
<p>Verbraucher ist jede natürliche Person, die ein Rechtsgeschäft zu Zwecken abschließt, die überwiegend weder ihrer gewerblichen noch ihrer selbständigen beruflichen Tätigkeit zugerechnet werden können. (§ 13 BGB)</p>	<p>Systran</p> <p>والمستهلكون هم أي شخص طبيعي يدخل في معاملة لأغراض لا تعزى إلى حد كبير إلى حرفته أو مهنته والمستهلكون هم أي شخص طبيعي يدخل في معاملة لأغراض لا تعزى إلى حد كبير إلى حرفته أو مهنته.</p>

	<p>(§ ١٣ جيجابايت).</p> <p>Chatgpt</p> <p>المستهلك هو أي شخص طبيعي يقوم بإبرام صفقة قانونية لأغراض لا يمكن تصنيفها بشكل رئيسي ضمن نشاطه التجاري أو المهني المستقل. (المادة 13 من قانون المدني الألماني)</p> <p>الترجمة البشرية:</p> <p>يعرف المستهلك على انه كل شخص طبيعي يقوم بإبرام صفقة قانونية لأغراض لا تسنف بشكل أساسي ضمن نشاطه التجاري أو عمله الحر. (المادة 13 من قانون المدني الألماني)</p>
<p>Ein Verein, dessen Zweck auf einen wirtschaftlichen Geschäftsbetrieb gerichtet ist, erlangt in Ermangelung besonderer bundesgesetzlicher Vorschriften Rechtsfähigkeit durch staatliche Verleihung. Die Verleihung steht dem Land zu, in dessen Gebiet der Verein seinen Sitz hat. (§ 22 BGB)</p>	<p>Systran</p> <p>وفي حال عدم وجود لوائح اتحادية خاصة، فإن الرابطة التي يتمثل غرضها في إدارة الأعمال التجارية الاقتصادية تكتسب أهلية قانونية بمنح العقد للدولة. وتمنح هذه الجائزة للدولة التي يوجد في أراضيها مقر النادي المسجل. (§ ٢٢ جيجابايت)</p> <p>Chatgpt</p> <p>يكتسب الجمعية التي تهدف إلى القيام بنشاط تجاري اقتصادي، في غياب تشريعات فيدرالية خاصة، شخصية اعتبارية من خلال منحة حكومية. تتم المنحة من قبل الدولة التي يتخذ فيها الجمعية مقرها. (المادة 22 من قانون المدني الألماني)</p> <p>الترجمة البشرية:</p> <p>في غياب لوائح اتحادية خاصة،</p>

	<p>فإن الجمعية التي تكمن أهدافها في القيام بنشاط تجاري تكتسب الصفة القانونية من خلال اعتماد تمنحه الدولة. يتم منح هذا الاعتماد من قبل الولاية التي يقع مقر الجمعية ضمن نطاقها الجغرافي. (المادة 22 من القانون المدني الألماني)</p>
<p>Diese Verordnung gilt für die ganz oder teilweise automatisierte Verarbeitung personenbezogener Daten sowie für die nichtautomatisierte Verarbeitung personenbezogener Daten, die in einem Dateisystem gespeichert sind oder gespeichert werden sollen. (Artikel 2 der Datenschutz-Grundverordnung der Europäischen Union)</p>	<p>Systran:</p> <p>تطبق هذه اللائحة على كل أو جزء من المعالجة الآلية للبيانات الشخصية وعلى المعالجة غير الآلية للبيانات الشخصية المخزنة أو المزمع تخزينها في نظام ملفات. (المادة ٢ من النظام الأساسي المتعلق بحماية بيانات الاتحاد الأوروبي)</p> <p>Chatgpt:</p> <p>تتطبق هذا اللائحة على معالجة البيانات الشخصية التي يتم معالجتها بشكل كامل أو جزئي بطرق آلية، وكذلك على معالجة البيانات الشخصية التي يتم تخزينها أو يجب تخزينها في نظام ملفات. (المادة 2 من لائحة حماية البيانات العامة للاتحاد الأوروبي)</p> <p>الترجمة البشرية:</p> <p>تتطبق هذا اللائحة على معالجة البيانات الشخصية التي تتم بشكل آلي كلياً أو جزئياً، وكذلك على المعالجة غير الآلية للبيانات المخزنة أو التي سيتم تخزينها في نظام ملفات. (المادة 2 من اللائحة العامة لحماية البيانات للاتحاد الأوروبي)</p>

7. Schlussfolgerung

Die obengenannten Beispiele aus der Rechtssprache geben Aufschluss über die Effizienz der neuronalen maschinellen Übersetzung in den Sprachpaaren Deutsch/Englisch und Deutsch/Arabisch, sowohl auf syntaktischer als auch auf semantischer Ebene. Beim Vergleich wurden für das Sprachpaar DE/EN zwei NMT-Anbieter herangezogen, nämlich DeepL und Systran. Beide Anbieter haben bei den ausgewählten Beispielen relativ gut abgeschnitten, wobei DeepL die besseren Ergebnisse und eine der menschlichen Übersetzung nähere und präzisere Übersetzung liefert. Generell kann man sagen, dass die neuronale maschinelle Übersetzung im Sprachpaar DE/EN zufriedenstellende Ergebnisse gewährleistet, zumindest in der Rechtssprache, wenn man bedenkt, dass das sogenannte Deep Learning in der Rechtssprache relativ leichter ausfällt, zumal solche Texte sehr gebräuchlich und verbreitet sind und einen größeren Bestand an parallelisierten Texten zum Training aufweisen. Da aber der vorliegende Beitrag darauf abzielt, einen Vergleich zwischen den Sprachpaaren DE/EN und DE/AR in Sachen Effizienz zu ziehen, sollten hier die diesbezüglichen Ergebnisse ausführlich präsentiert und ausgewertet werden. Auf syntaktisch-semantischer Ebene weist die NMT-basierte Übersetzung für Arabisch viele Mängel auf, die in eine professionelle Übersetzung nicht gehören. Zum Beispiel liest man beim ersten Beispiel „*Alle zwischen dem Hersteller ... und unserer Annahmeerklärung*“ im NMT-System Systran Folgendes: *وجميع الاتفاقات المبرمة بين الشركة المصنعة والعميل فيما يتعلق بعقد الحجز هي بوجه خاص نتيجة لشروط الحجز هذه في نسختها صالحة وقت الحجز في نسختها*. In dieser Übersetzung fällt ein syntaktisch-semantischer Fehler auf. Es sollte *في نسختها* heißen, was vom KI-System ChatGPT richtig formuliert wurde. Ein anderer Fehler ist die Verwendung eines unpassenden Wortes *نتيجة*, wobei hier *خاضعة* stehen sollte. Beim Beispiel „*Abweichende ... schriftlich zugestimmt*“ ist die Übersetzung von Systran komplett danebengegangen, wobei „*abweichende Vertragsbedingungen*“ mit *انحراف العامة الشروط* übersetzt wurde. Dabei geht es um einen semantischen Fehler, der sogar das Verständnis der Übersetzung beeinträchtigen kann. Es ist aber darauf hinzuweisen, dass die neuronale maschinelle Übersetzung nicht

gleich neuronale maschinelle Übersetzung ist, weil z. B. ChatGPT ein besseres Ergebnis beim selben Beispiel liefert, wobei „*abweichende Vertragsbedingungen*“ mit „الشروط العامة للعقد المختلفة“ übersetzt wird, was aber nicht mit einer korrekten menschlichen Übersetzung gleichzusetzen ist. Bei der Betrachtung der obengenannten Beispiele und beim Vergleich von NMT-basierten Übersetzungen kommt man zum Schluss, dass NMT-Systeme im Sprachpaar DE/EN viel besser abschneiden als im Sprachpaar DE/AR. Dieses Ergebnis könnte man mit der Tatsache begründen, dass die neuronale maschinelle Übersetzung im Sprachpaar DE/EN viel ausgereifter ist, weil die entsprechenden auf Deep Learning basierten Systeme über einen viel größeren Datenbestand mit parallelisierten Texten verfügen und somit besser trainiert werden können. Diese Diskrepanz in Sachen Qualität kann nicht mit der Komplexität der arabischen Syntax erklärt werden, weil NMT-Systeme in der Lage sind, mit irgendwelchem Regelwerk trainiert zu werden. Man benötigt bloß eine große Menge an Beispieldaten, die aus bereits übersetzten Texten besteht, wie beispielsweise Rechtstexte. Diese Daten werden als „parallele“ Texte bezeichnet, da sie Sätze oder Abschnitte enthalten, die in der Quellsprache und der Zielsprache übersetzt sind. Betrachtet man die verfügbaren Daten aus dem Rechtswesen, so stellt man fest, dass das Volumen zweisprachiger Rechtstexte DE/EN immens ist, sodass das Training von NMT-Systemen viel effizienter ist. Schließlich kann geschlussfolgert werden, dass die Qualität der NMT-basierten Übersetzung im Sprachpaar DE/AR auf die Bereitstellung ausreichender Trainingsdaten für NMT-Systeme angewiesen ist. Hierzu muss ein großer Bestand an parallelisierten Fachtexten von professionellen Übersetzern und den zuständigen Stellen aufgebaut und ständig aktualisiert werden.

Literaturverzeichnis

- Almahairi, A. & al. (2016). *First Result on Arabic Neural Machine Translation*. In *Computation and Language* (cs.CL). [Unternehmen], Zugriff 19.04.2023, URL: <https://arxiv.org/pdf/1606.02680.pdf>
- Kirste, M. & Schürholz, M. (2019). *Entwicklungswege zur KI*. In: Künstliche Intelligenz. Volker Wittpahl (Herausgeber). Springer Verlag. Berlin, Heidelberg.

- Koehn, P. (2017). *Neural Machine Translation*. In *Neural Machine Translation*. Center for Speech and Language Processing Department of Computer Science. Johns Hopkins University.
- Lionbridge (2022). *Maschinelle Übersetzung: Wichtige Begriffe kurz erklärt*, Zugriff 02.04.2023, URL: [Maschinelle Übersetzung: Wichtige Begriffe kurz erklärt \(lionbridge.com\)](https://www.lionbridge.com/insights/maschinelle-uebersetzung-wichtige-begriffe-kurz-erklart).
- Microsoft (O. J.). *Was ist maschinelle Übersetzung?* [Unternehmen], Zugriff 05.04.2023 URL: [Maschinelle Übersetzung-Microsoft Translator für Unternehmen](https://www.microsoft.com/de-de/translator/faq) Maschinelle Übersetzung-Microsoft Translator für.
- Schmalz, A. (2019). *Maschinelle Übersetzung*. In *Künstliche Intelligenz*. Volker Wittpahl (Herausgeber). Springer Verlag. Berlin, Heidelberg.