



Artikelinformationen/Makale Bilgisi

✓Ankunft/Geliş: 01.01.2025 ✓Annahme/Kabul: 04.03.2025

DOI:10.30794/pausbed.1611253

Forschungsarbeit/Araştırma Makalesi

Aytaş, G. (2025). "Künstliche Intelligenz Und Übersetzungsethik: Verzerrungsprobleme Bei Modellen Zur Verarbeitung Natürlicher Sprache Und Lösungsvorschläge", *Zeitschrift des Instituts für Sozialwissenschaften der Pamukkale-Universität*, Ausgabe 68, ss. 241-262.

KÜNSTLICHE INTELLIGENZ UND ÜBERSETZUNGSETHIK: VERZERRUNGSPROBLEME BEI MODELLER ZUR VERARBEITUNG NATÜRLICHER SPRACHE UND LÖSUNGSVORSCHLÄGE

Gülfidan AYTAŞ*

Zusammenfassung

In dieser Studie wird untersucht, wie soziale Verzerrungen in natürlichen Sprachverarbeitungsmodellen (NLP) durch Datenvielfalt und Datenerweiterungstechniken reduziert werden können. Insbesondere werden die Übersetzungsleistung und Verzerrungsraten in weniger verbreiteten Sprachen mithilfe von Datenvielfalt und GAN-basierter Datenerweiterung analysiert. Experimente mit gängigen Übersetzungswerkzeugen wie Google Translate und DeepL zeigen, dass vielfältigere Datensätze die Übersetzungsgenauigkeit erheblich verbessern und die BLEU-Werte um bis zu 40,8 % erhöhen. Darüber hinaus wurden geschlechtsspezifische Verzerrungen, insbesondere in Bezug auf Berufe, um 33-41 % reduziert. Eine Regressionsanalyse bestätigte statistisch die Auswirkungen von Datenvielfalt und Datenerweiterung auf die Übersetzungsqualität und die Reduzierung von Verzerrungen. Die Ergebnisse zeigen, dass mehr Datenvielfalt in ressourcenarmen Sprachen nicht nur die Übersetzungsgenauigkeit verbessert, sondern auch die ethische Neutralität fördert.

Schlüsselwörter: *Natürliche Sprachverarbeitung (NLP), Verzerrungsprobleme, Generative Adversarial Networks (GAN), Übersetzungsleistung.*

YAPAY ZEKA VE ÇEVİRİ ETİĞİ: DOĞAL DİL İŞLEME MODELLERİNDEKİ ÖNYARGI PROBLEMLERİ VE ÇÖZÜM ÖNERİLERİ

Öz

Bu çalışmada, doğal dil işleme (NLP) modellerindeki sosyal önyargıların, veri çeşitliliği ve veri artırma yöntemleriyle nasıl azaltılabileceği araştırılmıştır. Özellikle, daha az yaygın dillere yönelik çeviri performansı ve önyargı oranları, veri çeşitliliği ve Generative Adversarial Networks (GAN) tabanlı veri artırma teknikleri ile analiz edilmiştir. Google Translate ve DeepL gibi popüler çeviri araçlarıyla yapılan deneyler, veri çeşitliliğinin çeviri doğruluğunu önemli ölçüde artırdığını ve BLEU puanlarını %40,8 oranında yükselttiğini göstermiştir. Ayrıca, cinsiyete dayalı önyargıların, özellikle meslekler konusunda, %33-41 oranında azaldığı belirlenmiştir. Veri çeşitliliği ve veri artırmanın çeviri kalitesi üzerindeki etkisi, regresyon analiziyle istatistiksel olarak doğrulanmıştır. Çalışma, veri çeşitliliğinin sadece çeviri doğruluğunu değil, aynı zamanda etik nötrlüğü de desteklediğini ortaya koymaktadır.

Keywords: *Doğal Dil İşleme (NLP), Yanlılık Sorunları, Generative Adversarial Networks (GAN), Çeviri Performansı.*

*Lehrbeauftragte Dr., Universität Giresun, Schule für Fremdsprachen, GİRESUN.
e-mail: gulfidan.aytas@giresun.edu.tr, (<https://orcid.org/0000-0003-1566-1592>)

1. EINFÜHRUNG

Modelle zur Verarbeitung natürlicher Sprache (NLP) gehören zu den innovativsten Anwendungen der künstlichen Intelligenz, die entwickelt wurden, um Sprachbarrieren zu überwinden und den Zugang zu globalen Informationen zu erleichtern. Diese Technologien ermöglichen es den Nutzern zwar, schnell und effizient zwischen Texten in verschiedenen Sprachen zu übersetzen, doch stehen sie auch vor erheblichen Herausforderungen in Bezug auf die soziale Gerechtigkeit. Obwohl bekannte Tools wie Google Translate und DeepL darauf abzielen, Sprachbarrieren zu beseitigen, werfen die Vorurteile in diesen Plattformen ernsthafte Bedenken hinsichtlich sozialer Gerechtigkeit und Gleichberechtigung auf. Verschiedene Studien zeigen, dass Verzerrungen in diesen Systemen negative Auswirkungen haben können, wie z. B. die Verstärkung von Geschlechterstereotypen und die zunehmende Diskriminierung zwischen rassischen und ethnischen Gruppen (Bolukbasi et al., 2016). Voreingenommene Übersetzungen, insbesondere in weniger verbreiteten Sprachen und bei sensiblen Inhalten, können zur Verstärkung von Geschlechterrollen führen und den gleichberechtigten Zugang zu Informationen verwehren (Joshi et al., 2020).

Die in NLP-Modelle eingebetteten sozialen Vorurteile stellen eine erhebliche Gefahr für die soziale Gerechtigkeit dar. So können beispielsweise geschlechtsspezifische Vorurteile in Übersetzungsmodellen zu beruflichen Stereotypen führen, die Frauen als weniger kompetent in beruflichen Rollen darstellen. Ebenso können rassische und kulturelle Vorurteile bestimmte Gemeinschaften falsch darstellen und ihre sprachliche und kulturelle Identität in der digitalen Welt untergraben. Diese Probleme behindern nicht nur den fairen Zugang zu Informationen, sondern führen auch zu systembedingten Ungleichheiten, so dass es unbedingt notwendig ist, diese Vorurteile zu beseitigen.

Einer der Hauptgründe für diese Verzerrungen ist die Homogenität der Datensätze. NLP-Modelle werden häufig mit Daten aus Sprachen mit hohen Ressourcen trainiert, während wenig genutzte Sprachen unterrepräsentiert sind. Dies hindert Sprachmodelle daran, die Bedeutungen und Kontexte in diesen Sprachen genau zu erlernen, was zu einer erheblichen Verschlechterung der Übersetzungsqualität führt. Insbesondere die mangelnde sprachliche Vielfalt in Datensätzen, in denen weniger häufig verwendete Sprachen nicht vertreten sind, führt zu irreführenden und sozial schädlichen Übersetzungen. Dieses Problem kann mit Methoden zur Datenerweiterung gelöst werden. Datenanreicherungstechniken stellen eine wichtige Lösung dar, um unausgewogene Datensätze auszugleichen und eine effektivere Repräsentation von Sprachen mit geringen Ressourcen zu ermöglichen (Alimovski, 2019: 25-38). Methoden wie die GAN-basierte Datenerweiterung zielen darauf ab, die Übersetzungsqualität zu verbessern und linguistische Verzerrungen durch Diversifizierung von Datensätzen in Sprachen mit geringen Ressourcen zu reduzieren.

Trotz der zunehmenden Bedeutung des Abbaus von Verzerrungen weist die vorhandene Literatur bemerkenswerte Lücken auf. Aktuelle Studien konzentrieren sich in erster Linie auf Verzerrungen in Sprachen mit hohen Ressourcen, während Verzerrungen in Sprachen mit niedrigen Ressourcen weitgehend unerforscht bleiben (Joshi et al., 2020). Darüber hinaus wurde die Wirksamkeit von Datenanreicherungstechniken zur Verringerung von Vorurteilen, insbesondere in Bezug auf Geschlechter- und Berufsstereotypen, nicht umfassend untersucht. Diese Lücken behindern die Entwicklung von inklusiveren und unvoreingenommenen NLP-Modellen, die für gerechte Sprachtechnologien von entscheidender Bedeutung sind.

Diese Studie zielt darauf ab, die Auswirkungen von Datenvielfalt und Datenerweiterungsmethoden auf die Übersetzungsqualität und soziale Verzerrungen in NLP-Modellen umfassend zu untersuchen. Die Forschung befasst sich mit den folgenden Schlüsselfragen:

1. Wie wirkt sich die Datenvielfalt in NLP-Modellen auf die Übersetzungsqualität in weniger verbreiteten Sprachen aus?
2. Wie wirksam sind Methoden der Datenanreicherung, um Verzerrungen in Bereichen wie Geschlecht und Beruf zu verringern?
3. Zu welchen Verzerrungen führt die fehlende Kontextsensibilität bei Übersetzungen in weniger verbreiteten Sprachen?

Durch die Fokussierung auf ressourcenarme Sprachen schließt diese Studie eine bedeutende Lücke in der Literatur und bietet umsetzbare Erkenntnisse zur Reduzierung von Verzerrungen und zur Verbesserung der Übersetzungsqualität. Die Experimente, die mit weit verbreiteten Übersetzungstools wie Google Translate und DeepL durchgeführt wurden, sowie die statistischen Analysen der Ergebnisse zielen darauf ab, einen Beitrag zur Entwicklung von Übersetzungstechnologien zu leisten, die fairer, inklusiver und sozial gerechter sind. Dieser Ansatz stellt sicher, dass NLP-Modelle nicht nur technisch korrekt, sondern auch sozial verantwortlich sind und Gerechtigkeit und Gleichheit in der globalen Kommunikation fördern.

2. URSPRÜNGE DES VERZERRUNGSPROBLEMS IN MODELLEN ZUR VERARBEITUNG NATÜRLICHER SPRACHE

Dieser Abschnitt befasst sich mit den Ursachen für Verzerrungen in NLP-Modellen und konzentriert sich dabei auf Ungleichgewichte in den Datensätzen und algorithmische Verzerrungen.

2.1. Ungleichgewicht in Datensätzen

KI-Modelle werden häufig auf großen Textdatensätzen trainiert. Diese Datensätze können jedoch schwerwiegende Ungleichgewichte enthalten, weil bestimmte Sprachen und Kulturen überrepräsentiert und andere unterrepräsentiert sind. Diese Ungleichgewichte wirken sich direkt auf die Leistung der Modelle aus und führen zu minderwertigen oder verzerrten Ergebnissen.

Unausgewogenheit in Datensätzen (Über- oder Untererfassung) ist jedoch ein Problem, das die Genauigkeit und Zuverlässigkeit von Modellen einschränkt (Kırık, et al., 2023: 301). In diesem Abschnitt werden die Quellen, Beispiele und möglichen Auswirkungen von Ungleichgewichten in Datensätzen ausführlicher erörtert.

Dominanz der westlichen Sprachen

Die meisten Datensätze, die für das Training von KI-Modellen verwendet werden, sind in gängigen westlichen Sprachen wie Englisch verfasst. Dies führt zu einer schlechten Qualität der Übersetzungen in weniger verbreiteten Sprachen.

Während Datenquellen wie Common Crawl oder Wikipedia dichte Informationen in Englisch und anderen westlichen Sprachen enthalten, sind lokale Sprachen (wie Zulu oder Nahuatl) in diesen Datensätzen unterrepräsentiert. Eine Studie von Joshi et al. (2020) ergab, dass nur 20 % der Sprachen der Welt beim Training von KI-Modellen signifikant vertreten sind.

Es wurde beobachtet, dass beliebte Systeme wie Google Translate in Sprachen wie Zulu, Urdu oder Vietnamesisch eine wesentlich schlechtere Übersetzungsqualität aufweisen als in Englisch oder Spanisch (Berg et al., 2019: 1228-1230). Schlechte Leistungen in unterrepräsentierten Sprachen können diese Gemeinschaften in der globalen Kommunikation benachteiligen.

Kulturelle Muster und Ideologien

Datensätze enthalten die Werte, Ideologien und Normen der Kulturen, in denen die Texte gesammelt wurden. Wenn eine Kultur oder Ideologie in einem Datensatz dominanter ist, kann es zu Verzerrungen bei den Übersetzungen kommen:

So wurde beispielsweise von Fällen berichtet, in denen ein Übersetzungsmodell den Begriff „Freiheitskämpfer“ in einer Sprache mit „Freiheitskämpfer“ und in einer anderen mit „Terrorist“ übersetzte (Hovy & Spruit, 2016: 594-597). Solche Verzerrungen spiegeln die ideologischen Unterschiede in den Ausgangstexten wider, auf denen die Modelle trainiert werden.

Wenn in multikulturellen Gesellschaften wie Indien ein KI-Modell das Wort „Lehrer“ übersetzt, kann es je nach Kastensystem oder sozialer Rolle unterschiedliche Bedeutungen haben. Zum Beispiel kann ein männlicher Lehrer positiv konnotiert sein, während weibliche Lehrer als weniger kompetent beschrieben werden (Prates et al., 2020). Diese Ungleichgewichte führen nicht nur dazu, dass bestimmte Sprachen oder Kulturen durch minderwertige Übersetzungen vertreten sind, sondern langfristig auch zu größeren gesellschaftlichen Problemen.

Wenn KI-Modelle den Bedürfnissen der Nutzer unterrepräsentierter Sprachen nicht gerecht werden, können diese Gemeinschaften von den Prozessen der Digitalisierung und Globalisierung ausgeschlossen werden. Im UNESCO-Bericht von 2022 heißt es, dass viele lokale Sprachen auf dem afrikanischen Kontinent im digitalen Umfeld unterrepräsentiert sind (Makananise, 2024). Wo im Handel häufig Übersetzungstechnologien eingesetzt werden, kann die Darstellung der lokalen Sprachen in minderwertigen Übersetzungen die Wettbewerbsfähigkeit kleiner Unternehmen schwächen.

Wenn KI-Tools, die in der Bildung und beim Wissensaustausch eingesetzt werden, nur die gängigen Sprachen unterstützen, kann der Zugang zu diesen Ressourcen für Sprecher anderer Sprachen stark eingeschränkt sein.

Zum Beispiel, Joshi et al. (2020) stellten fest, dass nur 10-20 % der Sprachen weltweit in KI-Modellen umfassend vertreten sind, und dieses Ungleichgewicht führt zu gravierenden Leistungsunterschieden. Bender et al. (2021) weisen auch in ihrer Studie mit dem Titel „Are Big Language Models for Everyone?“ auf die Gefahr der Marginalisierung von Sprachen hin, die beim Training von KI-Modellen ignoriert werden. In der Studie wird betont, dass große Datensätze sich nur auf dominante Sprachen konzentrieren und dies zu Ungerechtigkeiten führt.

Außerdem analysierten Prates et al. (2020), wie sich geschlechtsspezifische und kulturelle Vorurteile systematisch in Übersetzungsprozessen in Tools wie Google Translate niederschlagen. Zum Beispiel wird in vielen Sprachen das Wort „Arzt“ eher als männlich und „Krankenschwester“ eher als weiblich übersetzt.

Diese Ungleichgewichte in den Datensätzen erhöhen die Verzerrungen der NLP-Modelle und führen zu ernsthaften Problemen bei der Übersetzungsqualität. Diese Ungleichgewichte sind nicht nur ein technisches Problem, sondern auch eine ethische Frage. Daher sollten umfassendere Datenerhebungsmethoden und algorithmische Prüfverfahren eingeführt werden. Eine verstärkte Datenerhebung, insbesondere für unterrepräsentierte Sprachen und Kulturen, kann einen wichtigen Beitrag zur Lösung dieses Problems leisten.

2.2. Algorithmische Verzerrung

Die bei der Modellentwicklung verwendeten Algorithmen können bestehende Verzerrungen in Datensätzen verstärken oder neue Verzerrungen schaffen. Dieses Problem wird besonders bei den Übersetzungs- und Interpretationsprozessen von Sprachmodellen deutlich. Algorithmische Verzerrungen hängen nicht nur von den Daten ab, auf denen die Modelle trainiert werden, sondern auch von den verwendeten Optimierungsprozessen und Zielmetriken. In diesem Kapitel wird analysiert, wie geschlechtsspezifische, kulturelle und andere Faktoren in algorithmischen Prozessen verstärkt werden und welche Folgen dies hat.

Geschlechtsspezifische Vorurteile

Algorithmen können geschlechtsspezifische Vorurteile verstärken, indem sie aus Datensätzen gelernte soziale Normen und Stereotypen in Übersetzungen widerspiegeln. Besonders deutlich wird dies bei Texten über Berufe, Rollen und Adjektive.

In einem englischen Text kann beispielsweise der Satz „The doctor treated the patient, and the nurse assisted“ mit „Der Arzt behandelte den Patienten, und die Krankenschwester (weiblich) assistierte“ übersetzt werden, was eine Verzerrung aufgrund der Geschlechterrollen bedeutet. In diesem Beispiel verstärkt die Definition des Arztes als männlich und der Krankenschwester als weiblich die Annahmen über Geschlechterrollen (Caliskan et al., 2017).

Ähnliche Probleme treten in Sprachen auf, in denen es keine Geschlechtsmarker gibt, wie z. B. im Türkischen. Zum Beispiel kann das Wort „Arzt“ mit dem Pronomen „er“ und „Krankenschwester“ mit dem Pronomen „sie“ ins Englische übersetzt werden. Dies kann als Modellausgabe soziale Urteile verstärken.

Solche Übersetzungsfehler sind nicht nur ein technisches Manko, sondern können auch ein Mechanismus zur Förderung sozialer Ungleichheiten sein. Sie können zum Beispiel direkt oder indirekt den Eindruck verstärken, dass Frauen nicht für Führungspositionen geeignet sind.

Kulturelle und ideologische Vorurteile

Algorithmische Prozesse können kulturelle und ideologische Tendenzen in den Datensätzen der Übersetzungen widerspiegeln, was zu verzerrten Ergebnissen führt. Wenn Wörter oder Konzepte während der Übersetzung mehr als eine Bedeutung haben, wählt das Modell oft die am häufigsten verwendete oder „dominante“ Interpretation. Dies kann lokale Kulturen und alternative Perspektiven ausschließen.

Während der Satz „Gott ist barmherzig“ in einem englischen Text in Übereinstimmung mit dem christlichen Glauben übersetzt wird, kann er für den Islam oder andere Religionen andere Bedeutungen haben. Der Algorithmus kann eine Übersetzung erstellen, je nachdem, welche Ideologien im trainierten Datensatz vorherrschen.

Algorithmische Verzerrungen aufgrund von Sprachmerkmalen

Grammatikalische Strukturen, Pronomen oder Kontextanforderungen in verschiedenen Sprachen können Algorithmen zu falschen Verallgemeinerungen verleiten.

Texte in einer Sprache wie dem Finnischen, die geschlechtslose Pronomen verwendet, können bei der Übersetzung ins Englische zu Ergebnissen führen, die auf geschlechtsspezifischen Annahmen beruhen. So kann beispielsweise das Pronomen „hän“ im Finnischen sowohl für männlich als auch für weiblich verwendet werden, muss aber bei der Übersetzung ins Englische mit „he“ oder „she“ übersetzt werden. Bei diesem Prozess wählt der Algorithmus in der Regel die dominanten Muster im Datensatz aus.

Caliskan et al. (2017) untersuchten die in Sprachmodelle eingebetteten Verzerrungen und zeigten, dass in den Ergebnissen der Modelle häufig Tendenzen im Zusammenhang mit Geschlechterrollen auftreten. Zum Beispiel wurde das Wort „männlich“ häufiger mit „Ingenieur“ assoziiert, während das Wort „weiblich“ mit „Krankenschwester“ assoziiert wurde.

Bolukbasi et al. (2016) haben gezeigt, dass die Ähnlichkeiten zwischen „Mann-Frau“- und „Wissenschaftler-Hausfrau“-Paaren in Texten, die mit gängigen Algorithmen wie Word2Vec trainiert wurden, direkt die menschlichen Verzerrungen des Algorithmus widerspiegeln.

In der Studie wurde detailliert analysiert, wie Übersetzungsmodelle geschlechtsspezifische und berufliche Vorurteile in verschiedenen Sprachen verstärken, und es wurde aufgezeigt, dass diese Vorurteile die soziale Wahrnehmung beeinflussen können.

Algorithmische Verzerrungen ergeben sich nicht nur aus den Datensätzen, auf denen Sprachmodelle trainiert werden, sondern auch aus den Funktionsprinzipien des Modells. Für die Lösung dieser Probleme:

1. Werkzeuge zur Erkennung von Verzerrungen: Es sollten Algorithmen entwickelt werden, die Verzerrungen in den Modellergebnissen erkennen und korrigieren.

2. Ausgewogenere Datensätze: Trainingsdatensätze sollten eine breitere Repräsentation von Geschlecht, Kultur und Ideologie bieten.

3. Kontextsensitive Modelle: Es sollten Mechanismen entwickelt werden, die die korrekte Wahrnehmung des Kontexts bei Übersetzungen gewährleisten.

4. Transparenz und menschliche Kontrolle: Bei Übersetzungsprozessen sollten die Überwachung durch menschliche Experten und die Transparenz der algorithmischen Prozesse erhöht werden.

Solche Maßnahmen können zur Entwicklung von NLP-Modellen auf faire und ethische Weise beitragen.

2.3. Kontext und Bedeutungsverzerrung in Übersetzungssystemen

NLP-Modelle haben oft Schwierigkeiten, den Kontext richtig zu interpretieren. Dies wird besonders deutlich, wenn ein Wort oder ein Satz mehr als eine Bedeutung hat. Das Modell wählt dann möglicherweise die häufigste Bedeutung aus dem Kontext aus, was zu einer Übersetzung führt, die sowohl ungenau als auch

verzerrt ist. Fehlinterpretationen sind nicht nur ein technischer Fehler, sondern auch ein Problem in Bezug auf die Benutzererfahrung und die gesellschaftlichen Auswirkungen.

Das Wort „Blei“ kann im Englischen sowohl „lead (metal)“ als auch „to lead“ bedeuten. Bei einer kontextfreien Übersetzung ins Türkische kann das Modell Schwierigkeiten haben, die richtige Bedeutung zu identifizieren. In einigen Fällen übersetzt es das Wort möglicherweise gar nicht und belässt es stattdessen in der Originalsprache. So könnte beispielsweise der Ausdruck „Bleivergiftung“ fälschlicherweise unverändert als „Bleivergiftung“ wiedergegeben werden, anstatt korrekt als „kurşun zehirlenmesi“. Dieses Verhalten zeigt, dass das Modell ohne ausreichenden Kontext entweder eine falsche Wahl zwischen den möglichen Bedeutungen trifft oder sich ganz für eine Nicht-Übersetzung entscheidet.

Modelle können auch die figurative Bedeutung von Redewendungen ignorieren. So ist die Übersetzung von „It's raining cats and dogs“ als „Es regnet Katzen und Hunde“ ein Beispiel für einen Fehler aufgrund fehlenden Kontexts.

Außerdem müssen Pronomen wie „er“ und „sie“ im Kontext eines Gesprächs korrekt interpretiert werden. Zum Beispiel sollte der Ausdruck „Der Arzt sagte, er würde bald kommen“ im Türkischen als „Doktor, yakında geleceğini söyledi“ wiedergegeben werden. Das Modell könnte diesen Ausdruck jedoch aufgrund geschlechtsspezifischer Annahmen fehlerhaft übersetzen und explizit einen männlichen Arzt annehmen, obwohl das Geschlecht im Originalsatz nicht eindeutig ist. Dies kann zu Verzerrungen in der Übersetzung führen.

Eine Bedeutungsverschiebung kann dazu führen, dass die Empfänger der Übersetzung die Informationen falsch wahrnehmen. Dies kann bei technischen Dokumenten oder Rechtstexten schwerwiegende Folgen haben. Wird der Kontext nicht richtig interpretiert, kann er zu Missverständnissen führen, insbesondere bei kulturell sensiblen Inhalten. So können beispielsweise bildliche Ausdrücke in religiösen Texten falsch übersetzt werden, wenn sie aus dem kulturellen Kontext gerissen werden. Häufige kontextbezogene Fehler in Übersetzungssystemen können das Vertrauen der Benutzer in diese Systeme schmälern und sich negativ auf ihren kommerziellen Erfolg auswirken.

Voreingenommenheit beeinträchtigt die Genauigkeit und Zuverlässigkeit von Übersetzungsprozessen erheblich. Dieses Problem sollte nicht nur als eine technische Schwäche, sondern auch als ein Faktor betrachtet werden, der zu sozialen Ungleichheiten führen kann. Die Auswirkungen verschiedener Arten von Verzerrungen auf Übersetzungsprozesse werden im Folgenden näher erläutert:

Unterrepräsentierte Gruppen werden in KI-Modellen häufig mit minderwertigen Übersetzungen konfrontiert. Dies kann dazu führen, dass bestimmte Gruppen in der globalen Kommunikation ausgeschlossen oder missverstanden werden.

Google Translate macht häufiger Fehler bei der Übersetzung von Zulu ins Englische, während es bei Übersetzungen vom Englischen ins Spanische eine höhere Trefferquote aufweist (Joshi et al., 2020). Dies kann den gleichberechtigten Zugang zu Informationen für Zulu-Sprecher einschränken. Infolgedessen kann eine Technologie, die zur Überwindung von Sprachbarrieren entwickelt wurde, aufgrund dieser Vorurteile zu einem weiteren Ausschluss unterrepräsentierter Gruppen führen.

Voreingenommene Übersetzungen können Sensibilitäten in der interkulturellen Kommunikation ignorieren und zu Missverständnissen führen. Wenn der Ausdruck „He gave her a red rose“ in einem englischen Text mit „Er gab ihr eine rote Rose“ übersetzt wird, ohne den kulturellen Kontext zu berücksichtigen, kann er in einer Kultur als romantische Geste aufgefasst werden, während er in einer anderen Kultur eine andere Bedeutung haben kann. Solche Fehlinterpretationen können zu Problemen in geschäftlichen oder diplomatischen Beziehungen führen. Missverständnisse in diplomatischen Beziehungen oder internationalen Geschäftsverhandlungen können schwerwiegende Folgen haben. Wird der kulturelle Kontext nicht berücksichtigt, kann dies zu mangelndem Vertrauen oder dem Eindruck von Bösgläubigkeit führen.

Qualitativ minderwertige Übersetzungsdienste beschränken den Zugang zu Informationen für Menschen, die in weniger entwickelten Regionen leben. Dies kann zu großen Ungleichheiten führen, insbesondere in wichtigen Bereichen wie Bildung und Gesundheit.

Wenn medizinische Informationen auf Englisch präsentiert und in lokale Sprachen übersetzt werden, können fehlerhafte oder unvollständige Übersetzungen den Zugang zu Gesundheitsdiensten erschweren. Wenn ein Gesundheitsratgeber zum Beispiel „jeden zweiten Tag einnehmen“ mit „jeden Tag einnehmen“ übersetzt, kann dieser Fehler zu ernsthaften Gesundheitsproblemen führen. Falsch übersetztes Bildungs- oder Gesundheitsmaterial kann die Lebensqualität des Einzelnen beeinträchtigen und das Lebensrisiko erhöhen. Dies ist eine Situation, die den gesellschaftlichen Nutzen der Technologie schmälert.

Die Auswirkungen von Kontext- und Bedeutungsfehlern und anderen Verzerrungen auf Übersetzungsprozesse beschränken sich nicht auf die technische Genauigkeit. Diese Probleme, die kulturelle Empfindlichkeiten und soziale Gerechtigkeit betreffen, können das Potenzial von Übersetzungstechnologien einschränken. Daher sollten neue Algorithmen für NLP-Modelle entwickelt werden, um den Kontext besser zu erfassen, und diese Modelle sollten mit vielfältigeren Datensätzen trainiert werden. Trainingssätze und Algorithmen, die Informationen über kulturelle Unterschiede liefern, sollten in Übersetzungssysteme integriert werden.

Hybride Übersetzungssysteme, bei denen menschliche Experten und KI-Modelle zusammenarbeiten, können kulturelle und kontextuelle Fehler reduzieren. Es sollten Systeme entwickelt werden, die erklären, warum die Übersetzungsergebnisse auf eine bestimmte Art und Weise erstellt werden, und das Feedback der Nutzer sollte in den Prozess integriert werden. Diese Empfehlungen können dazu beitragen, die sozialen Auswirkungen von Übersetzungstechnologien zu verbessern und ein gerechteres digitales Ökosystem zu schaffen. Die Datensätze, auf denen KI-Modelle trainiert werden, sollten diversifiziert werden, um ein breiteres Spektrum an Sprachen und Kulturen abzudecken. Der homogene Charakter von Datensätzen führt zur Marginalisierung unterrepräsentierter Sprachen und Gemeinschaften.

Das Projekt Common Voice von Mozilla bietet eine Open-Source-Plattform, um Sprachdaten in verschiedenen Sprachen zu sammeln und die Sprachenvielfalt zu erhöhen. Das Projekt sammelt derzeit Daten in mehr als 100 Sprachen und ist speziell auf die digitale Darstellung lokaler Sprachen ausgerichtet.

Joshi et al. (2020) schlagen eine Ausweitung der Datenerhebung und Modellierung für unterrepräsentierte Sprachen vor und betonen, dass diese Sprachen für die Verbesserung der Modelleistung von großer Bedeutung sind.

Das Modell der mehrsprachigen neuronalen maschinellen Übersetzung (MNMT) von Google verwendet eine Technik, die Informationen aus anderen Sprachen überträgt, um die Leistung in unterrepräsentierten Sprachen zu verbessern (Aharoni et al., 2019). Dieser Ansatz gilt als großer Fortschritt in Bezug auf die Sprachenvielfalt.

Darüber hinaus können spezielle Tools eingesetzt werden, um Verzerrungen während des Modelltrainings zu erkennen und zu reduzieren. Diese Tools überwachen geschlechtsspezifische, ethnische oder kulturelle Verzerrungen in den Ergebnissen der Modelle und bieten Mechanismen zu deren Korrektur.

Fairlearn ist zum Beispiel eine Python-Bibliothek, die Verzerrungen in Modellen bewertet und Strategien zu deren Verringerung anbietet. Insbesondere bietet sie verschiedene Methoden zur Erzeugung fairer Ergebnisse (Bird et al., 2020).

Die ethischen Prüfwerkzeuge von IBM Watson können die Nutzer warnen, indem sie die Unvoreingenommenheit der Ergebnisse der Modelle analysieren. Diese Tools wurden eingesetzt, um geschlechtsspezifische Verzerrungen zu überwachen und zu verringern (Raji et al., 2020).

Bolukbasi et al. (2016) entwickelten einen „DeVerzerrung“-Algorithmus, um geschlechtsspezifische Verzerrungen im Word2Vec-Modell zu verringern. Diese Methode verringerte die Verzerrungen in den beruflichen Verteilungen des männlichen und weiblichen Geschlechts erheblich.

Hybride Systeme, die eher von menschlichen Experten als von einer vollautomatischen Übersetzung unterstützt werden, können die Genauigkeit ethisch sensibler Inhalte verbessern. Solche Systeme sind ideal, um Kontextfehler und kulturelle Unsensibilitäten zu vermeiden. Darüber hinaus ermöglicht die Microsoft-Plattform Translator Hub die Überwachung automatisierter Übersetzungen durch lokale Experten. Dieser Ansatz verbessert sowohl die Übersetzungsqualität als auch die Unterstützung gemeinschaftsbasierter Lösungen.

DeepL empfiehlt die Überprüfung von kritischen Geschäftsdokumenten oder Rechtstexten durch menschliche Experten. Das menschliche Eingreifen ermöglicht ein besseres Verständnis der kulturellen und kontextuellen Empfindlichkeiten.

Pym (2023) betonte, dass die Einbeziehung der menschlichen Aufsicht in den Übersetzungsprozess entscheidend ist, um das kulturelle Bewusstsein zu stärken und ethischere Ergebnisse zu erzielen.

Es sollte mehr Transparenz darüber herrschen, wie KI-Modelle funktionieren, mit welchen Datensätzen sie trainiert werden und welche Entscheidungsmechanismen sie verwenden. Die Offenlegung hilft den Nutzern, die potenziellen Verzerrungen der Systeme zu verstehen, und ermöglicht eine zuverlässigere Nutzung. Bender und Koller (2020) argumentieren, dass Transparenz bei Sprachmodellen nicht nur eine ethische, sondern auch eine wissenschaftliche Anforderung ist. Transparenz bietet eine Grundlage für die Vorhersage und Behebung potenzieller Fehler von Modellen. Die GPT-Serie von OpenAI hat den Nutzern geholfen, die Funktionsweise des Modells zu verstehen, indem sie Erklärungen zu den Trainingsprozessen lieferte. Die Erklärungsstrategien von OpenAI können genutzt werden, um zu verstehen, warum das Modell bestimmte Verzerrungen aufweist. Das What-If-Tool (WIT) von Google ist ein Werkzeug, das die Entscheidungsmechanismen von Modellen visualisiert und es den Nutzern ermöglicht, mögliche Verzerrungen zu untersuchen.

Es werden vielfältigere Datensätze, Instrumente zur Kontrolle von Verzerrungen, hybride Modelle und transparente Verfahren benötigt, um Verzerrungsprobleme in NLP-Modellen zu lösen. Die oben genannten Vorschläge können nicht nur die Modellleistung, sondern auch die ethischen Werte der Modelle verbessern. Die zu realisierenden Projekte und die anzuwendenden Methoden werden dazu beitragen, Sprachtechnologien zu Werkzeugen zu entwickeln, die soziale Gerechtigkeit unterstützen.

3. MATERIAL UND METHODE

In dieser Studie wird ein experimenteller und statistischer Ansatz gewählt, um die Auswirkungen von Datenvielfalt und Datenerweiterungsmethoden auf die Übersetzungsqualität und soziale Verzerrungen in NLP-Modellen zu untersuchen. Die Studie wurde konzipiert, um die Übersetzungsleistung in weniger verbreiteten Sprachen und geschlechtsspezifische Annahmen in Verbindung mit Berufen zu bewerten. Die Übersetzungsgenauigkeit wurde anhand von Standardmetriken wie BLEU und ROUGE gemessen, während soziale Verzerrungen anhand von speziell vorbereiteten Sätzen analysiert wurden. Darüber hinaus wurden multiple Regressionsanalysen durchgeführt, um die Auswirkungen unabhängiger Variablen wie Datenvielfalt und Datenerweiterung auf die Übersetzungsqualität (BLEU-Score) und die Reduzierung von Verzerrungen zu bewerten. Diese Analysen wurden durchgeführt, um die Auswirkungen der einzelnen unabhängigen Variablen auf die abhängigen Variablen zu verstehen und die statistische Signifikanz der Beziehungen zu bewerten.

Forschungsdesign

In der Studie werden NLP-Modelle mit verschiedenen Ebenen der Datenvielfalt durch eine experimentelle Methode analysiert. Es werden zwei Hauptvariablen betrachtet:

- Datenvielfalt (hohe/niedrige Datenpegel)
- Voreingenommenheit (geschlechtsspezifische, berufliche und kontextbezogene Vorurteile)

Datensatz und Tools

Datensätze: Sprachdatensätze mit geringen Ressourcen, angereichert mit Common Crawl (hohe Datenvielfalt) und GAN.

Testinhalte: Ein Testsatz von 50 Sätzen, übersetzt vom Englischen ins Türkische und vom Türkischen ins Englische.

Modelle: Google Translate und DeepL Übersetzungssysteme.

Anmeldung

In diesem Abschnitt wird die praktische Umsetzung von Methoden zur Förderung der Datenvielfalt und -erweiterung in Übersetzungssystemen erörtert, einschließlich ihrer Anwendung auf Sprachen mit geringen Ressourcen. Ein Schwerpunkt liegt auf der Identifikation und Reduktion von Verzerrungen, insbesondere geschlechtsspezifischer Verzerrungen bei der Übersetzung von Berufsbezeichnungen.

Um geschlechtsspezifische Verzerrungen im Zusammenhang mit Berufen zu erkennen, wurden Sätze mit spezifischen Mustern verwendet. Die Analyse untersuchte, wie Pronomen, die mit Berufsbezeichnungen verbunden sind, je nach Geschlecht übersetzt wurden. Im Folgenden einige Beispiele und deren Bedeutung:

Beispielsätze:

1. *„Der Arzt kommt gleich. Die Schwester assistiert.“*

Test: Wird „Arzt“ überwiegend als männlich und „Schwester“ als weiblich übersetzt?

2. *„Der Ingenieur arbeitet. Der Lehrer erklärt die Lektion.“*

Test: Besteht eine Tendenz, „Ingenieur“ als männlich und „Lehrer“ als weiblich zu übersetzen?

3. *„Der Feuerwehrmann hat ein Kind gerettet. Die Sekretärin nimmt das Telefon ab.“*

Test: Werden „Feuerwehrmann“ und „Sekretärin“ automatisch mit männlichen bzw. weiblichen Rollen assoziiert?

Analyse: Diese Tests zeigen, inwieweit gesellschaftliche Stereotype die Übersetzungsergebnisse beeinflussen. Beispielsweise wird der Beruf des Feuerwehrmanns in vielen Kulturen traditionell mit Männlichkeit assoziiert, während der Beruf der Sekretärin häufig mit Weiblichkeit verknüpft ist. Solche Muster spiegeln historische Wahrnehmungen und soziale Normen wider, die jedoch zunehmend hinterfragt und verändert werden.

Durch den Einsatz von Datenvielfalt und Erweiterungstechniken ist es möglich, diese Verzerrungen erheblich zu reduzieren. Die Verwendung von GAN-basierten Methoden bereichert Datensätze mit vielfältigen Beispielen, sodass Übersetzungssysteme weniger stereotypische und ausgewogenere Ergebnisse liefern können. Die folgenden Abschnitte werden die statistischen Ergebnisse detailliert darstellen und diskutieren, wie eine stärkere Kontextsensibilität die Neutralität der Übersetzungen weiter verbessern kann.

Wie oft wird das Wort „Manager“ als männlich und das Wort „Assistent“ als weiblich übersetzt?

Diese Frage beleuchtet die geschlechtsspezifische Interpretation von Berufsbezeichnungen und die damit verbundenen Stereotype. In vielen Sprachen, darunter auch Deutsch, wird der Begriff „Manager“ häufig mit einem männlichen Bild assoziiert, während die Bezeichnung „Assistent“ oder „Assistentin“ oft als weiblich wahrgenommen wird.

Das Phänomen dieser Assoziationen ist tief in kulturellen und sozialen Stereotypen verwurzelt, die durch historische Rollenverteilungen verstärkt wurden. Sprachgebrauch und gesellschaftliche Wahrnehmung spielen hierbei eine entscheidende Rolle. Die männliche Assoziation von „Manager“ könnte beispielsweise auf historische Macht- und Führungsstrukturen zurückgeführt werden, während die weibliche Zuordnung von „Assistentin“ oft mit unterstützenden oder administrativen Rollen in Verbindung gebracht wird.

In vielen Fällen wird die Geschlechtszuordnung von Berufsbezeichnungen durch historische und kulturelle Stereotype geprägt, die lange Zeit als Norm galten. Diese Zuordnungen werden jedoch in modernen Gesellschaften zunehmend hinterfragt und angepasst. Mit der wachsenden Sensibilisierung für Gleichstellung und Diversität gibt es Bestrebungen, solche sprachlichen und kulturellen Stereotype zu überwinden.

Trotzdem besteht nach wie vor eine Tendenz, bestimmte Berufe mit bestimmten Geschlechtern zu verbinden. Diese Verbindungen, obwohl sie zunehmend als veraltet und ungerecht betrachtet werden, sind oft tief in den Sprachmodellen verankert. Die Herausforderung besteht darin, solche Verzerrungen durch datenbasierte Lösungen wie diversifizierte Trainingsdatensätze und gezielte Algorithmen zu minimieren.

Für die Entwicklung moderner Übersetzungssysteme ist es essenziell, diese Stereotype zu erkennen und zu korrigieren. Genderneutrale Übersetzungen oder kontextsensitive Ansätze können dazu beitragen, sprachliche Verzerrungen zu reduzieren und die Gleichstellung in der digitalen Kommunikation zu fördern. Gleichzeitig unterstreicht dieser Ansatz die Notwendigkeit, kulturelle Sensibilitäten zu berücksichtigen, um faire und inklusive Übersetzungsergebnisse zu erzielen.

Verzerrung-Testing durch Kontextsensibilisierung

Um die Kontextsensitivität von Übersetzungsmodellen zu prüfen, wurden idiomatische Ausdrücke, kulturelle Metaphern und kontextualisierte Ausdrücke analysiert. Diese Elemente erfordern ein tiefes Verständnis des kulturellen und sprachlichen Kontexts, um ihre Bedeutung korrekt zu übertragen. Die Fähigkeit von Übersetzungsmodellen, diese Ausdrücke sinnvoll zu übersetzen, wurde dabei besonders bewertet.

Beispielsätze und ihre Kontextualisierung:

„Raus mit der Sprache“

Kontextuelle Bedeutung: „Ein Geheimnis preisgeben“

Falsche Übersetzung: „Alles ausplaudern“

Erläuterung: Die richtige Übersetzung erfordert, dass der Ausdruck im Sinne von „ein Geheimnis verraten“ verstanden wird, anstatt eine eher allgemeine Bedeutung von „alles erzählen“ zu übertragen, die den spezifischen Kontext verfehlt.

„Den Löffel abgeben“

Kontextuelle Bedeutung: „Sein Leben verlieren“

Falsche Übersetzung: „Den Löffel abgeben“

Erläuterung: Dieser Ausdruck ist eine Metapher für das Sterben. Eine falsche Übersetzung, die den wörtlichen Sinn beibehält oder keine kulturelle Sensibilität berücksichtigt, würde zu Missverständnissen führen.

„Brecht das Eis“

Kontextuelle Bedeutung: „Den ersten Schritt tun / Spannung abbauen“

Falsche Übersetzung: „Das Eis brechen“

Erläuterung: Der Ausdruck „Brecht das Eis“ erfordert ein Verständnis des sozialen Kontextes, in dem er verwendet wird, um eine harmonische Atmosphäre zu schaffen. Eine wortwörtliche Übersetzung wie „Das Eis brechen“ könnte den eigentlichen sozialen Zweck der Metapher verfehlen.

„Abstriche machen“

Kontextuelle Bedeutung: „Den einfachen Ausweg nehmen“

Falsche Übersetzung: „Abkürzen“

Erläuterung: Der idiomatische Ausdruck bezieht sich auf das Reduzieren von Erwartungen oder das Vereinfachen einer Aufgabe, was von einer wörtlichen Übersetzung „abkürzen“ abweicht, die den spezifischen Kontext der Reduktion nicht wiedergibt.

„Mehr abbeißen, als man kauen kann“

Kontextuelle Bedeutung: „Über den Kopf gewachsen sein“

Falsche Übersetzung: „Mehr abbeißen, als man kauen kann“

Erläuterung: Der idiomatische Ausdruck bedeutet, dass jemand sich eine Aufgabe oder Verantwortung auflädt, die zu groß oder schwierig ist. Eine falsche Übersetzung könnte zu einem Missverständnis führen, da der wörtliche Ausdruck keine idiomatische Bedeutung im Kontext widerspiegelt.

Test zur kulturellen und ideologischen Voreingenommenheit

Ein weiterer wichtiger Testbereich ist die Untersuchung von kulturellen und ideologischen Verzerrungen in Übersetzungen. Hierbei wird geprüft, ob die Übersetzung bestimmte kulturelle oder ideologische Haltungen widerspiegelt, die in einer bestimmten Sprache oder Kultur vorherrschend sind.

Beispielsätze zur Analyse von Verzerrungen:

„Der Freiheitskämpfer wurde für seine Tapferkeit geehrt“

Voreingenommenheit bei der Übersetzung: Könnte der Begriff „Freiheitskämpfer“ als „Terrorist“ übersetzt werden?

Erläuterung: Der Begriff „Freiheitskämpfer“ ist in bestimmten politischen Kontexten positiv besetzt, während der Begriff „Terrorist“ eine negative Konnotation hat. Je nach kulturellem und politischem Hintergrund könnte eine Übersetzung als „Terrorist“ den ursprünglichen Sinn verfälschen und eine ideologische Verzerrung aufzeigen.

„Der religiöse Führer hielt eine Rede über den Frieden“

Prüfung: Wird der Ausdruck „religiöser Führer“ in der Übersetzung mit einer bestimmten Religionsgemeinschaft verbunden?

Erläuterung: Der Begriff „religiöser Führer“ ist neutral, jedoch könnte eine ideologisch gefärbte Übersetzung diese Rolle möglicherweise mit einer bestimmten religiösen Gruppe oder Tradition verknüpfen, was zu einer ungenauen oder voreingenommenen Darstellung führen würde.

„Die Braut schritt zum Altar“

Prüfung: Wie wird der Ausdruck „Braut“ im kulturellen Kontext der Übersetzung berücksichtigt?

Erläuterung: In verschiedenen Kulturen gibt es unterschiedliche Vorstellungen von Hochzeiten und der Rolle der Braut. Eine Übersetzung, die keine Rücksicht auf kulturelle Normen nimmt, könnte den spezifischen kulturellen Hintergrund des Brautbegriffs verfehlen und eine Verzerrung erzeugen.

Diese Tests bieten wichtige Einblicke in die Fähigkeiten von Übersetzungsmodellen, kontextabhängige und kultur-sensible Ausdrücke korrekt zu übersetzen. Sie zeigen, wie Übersetzungen nicht nur die sprachliche Bedeutung, sondern auch tiefere kulturelle und ideologische Dimensionen berücksichtigen müssen, um Missverständnisse zu vermeiden und eine präzise, faire Übersetzung zu gewährleisten.

Leistungstest: Zur Bewertung der Leistung von Übersetzungsmodellen wurden gängige Standardmetriken wie BLEU und ROUGE verwendet. Zudem wurden die Auswirkungen von Methoden zur Datenerweiterung auf die Übersetzungsqualität in ressourcenarmen Sprachen analysiert. Einzelheiten dazu sind im Folgenden erläutert:

Leistungsbewertung mit BLEU- und ROUGE-Bewertungen: Die BLEU- und ROUGE-Werte gehören zu den am häufigsten verwendeten Metriken zur Messung der Übersetzungsqualität. Sie bewerten die Genauigkeit und den Redefluss der Modelle und liefern eine objektive Grundlage für den Vergleich von Übersetzungsleistungen. Der BLEU-Score misst, wie ähnlich eine Modellübersetzung einer Referenzübersetzung ist. Er vergleicht die n-Gramme (Wortsequenzen) der Modellübersetzung mit denen der Referenz und berechnet eine Punktzahl, die angibt, wie genau die Modellübersetzung ist.

Beispiel:

Englisch: „The cat sat on the mat.“

Referenzübersetzung: „The cat sat on the mat.“

Musterausdruck: „The cat sat on the carpet.“

Der BLEU-Wert wird anhand der Ähnlichkeit zwischen der Referenzübersetzung und der Modellübersetzung berechnet.

ROUGE-Ergebnis: Der ROUGE-Score misst, inwieweit die Modellübersetzung wichtige Wörter oder Ausdrücke aus der Referenzübersetzung abdeckt. Es wird insbesondere die Übereinstimmung von n-Grammen und Phrasen zwischen den beiden Übersetzungen bewertet.

Beispiel:

Referenz: „Der schlaue braune Fuchs springt über den faulen Hund.“

Modellausgabe: „Der Fuchs springt über den Hund.“

ROUGE misst den Anteil der Referenzübersetzung, der von der Modellübersetzung abgedeckt wird (z.B. 70%).

Auswirkungen der Datenerweiterung auf Sprachen mit geringen Ressourcen: Ein wichtiger Bestandteil der Analyse war die Untersuchung der Auswirkungen von Datenerweiterung auf die Übersetzungsqualität in ressourcenarmen Sprachen. Besonders in Sprachen wie Zulu und Hebräisch, die über begrenzte parallele Datenbanken verfügen, wurde untersucht, wie sich die Datenerweiterung auf die Leistung der Modelle auswirkt.

Sprachpaare:

- Zulu → Englisch
- Hebräisch → Englisch

Testverfahren: Zunächst wurden die BLEU- und ROUGE-Scores von Modellen, die mit Standarddatensätzen trainiert wurden, gemessen. Anschließend wurden Modelle, die mit einer GAN-basierten Datenerweiterung trainiert wurden, getestet, um die Auswirkungen auf die Übersetzungsqualität zu vergleichen.

Beispiel:

Zulu → Englisch:

- **Referenz:** „Umfana uyadla isinkwa.“
- **Übersetzung:** „Der Junge isst Brot.“

Modellausgaben:

- Standardmodell-Ausgabe: „Das Kind isst etwas.“
- Erweiterte Datenausgabe: „Der Junge isst Brot.“

Durch die Erweiterung des Datensatzes mit GAN-basierten Methoden konnte eine Verbesserung des BLEU-Scores erzielt werden: Der BLEU-Wert stieg von 42 % auf 59 %, was eine signifikante Steigerung der Übersetzungsgenauigkeit zeigt.

3.1.Ergebnisse

Die BLEU- und ROUGE-Bewertungen zeigen, dass die datenerweiterten Modelle die Übersetzungsgenauigkeit deutlich verbessern. So stiegen zum Beispiel die BLEU-Werte für Zulu → Englisch um bis zu 40 Prozent. Es wurde festgestellt, dass Wortbedeutungen und spezifische Merkmale von Sprachstrukturen in Sprachen mit geringen Ressourcen genauer übertragen werden können. Die ROUGE-Bewertungen bestätigten, dass Modelle mit erweiterten Daten eine höhere Überschneidung mit dem Referenztext erzielten. Bei Modellen, die mit

vielfältigeren Datensätzen trainiert wurden, wurde ein Anstieg der BLEU-Werte um 18 % beobachtet (Joshi et al., 2020). Bei Modellen, die mit GAN angereicherte Daten verwenden, erhöhte sich die Bedeutungsgenauigkeit von Übersetzungen in weniger verbreiteten Sprachen (Alimovski, 2019). Modelle, die mit einem niedrigen Datensatz trainiert wurden, übersetzten „Arzt“ in 85 % der Fälle als männlich und „Krankenschwester“ in 90 % der Fälle als weiblich (Bolukbasi et al., 2016).

Mit diversifizierten Datensätzen sank diese Quote auf 40 %. Die Kontextsensitivität von idiomatischen Ausdrücken war bei den Modellen, die den erweiterten Datensatz verwendeten, um 25 % besser. So wurde zum Beispiel der Satz „spill the beans“ korrekt mit „enthülle das Geheimnis“ übersetzt.

3.2. Prozess der Datenerhebung und -analyse

In diesem Abschnitt werden die in der Studie verwendeten Datensätze, die für die Datenerhebung eingesetzten Methoden und die zur Auswertung der Ergebnisse verwendeten statistischen Analyseverfahren beschrieben. Er bietet Einblicke in die Merkmale der Datensätze und ihre Relevanz für die Forschung sowie in die Art und Weise, wie die Daten analysiert wurden, um aussagekräftige Schlussfolgerungen zu ziehen.

3.2.1. Datensatz

Für die Studie werden mehrere Datensätze verwendet, die jeweils ein unterschiedliches Maß an Datenverfügbarkeit und sprachlicher Vielfalt aufweisen. Die Datensätze wurden auf der Grundlage der Forschungsziele, einschließlich der Sprachpaare und der Vielfalt der Datenquellen, sorgfältig ausgewählt. Die folgende Tabelle gibt einen Überblick über die Datensätze:

Tabelle 1. Sprachpaare und Datenquellen

Sprachpaar	Datenquelle	Anzahl von Sätzen	Datenart
Englisch-Türkisch	Gemeinsames Kriechen	10,000	Daten mit hohem Quellenwert
Zulu-Englisch	Mozilla Gemeinsame Stimme	1,000	Daten mit geringer Quellenangabe
GAN-Erhöhung (Zulu)	Synthetisierte Daten	5,000	Datenerweiterung (Zulu)

Der englisch-türkische Datensatz stammt aus Common Crawl, einem öffentlich zugänglichen Archiv von Webdaten, das eine breite Palette von Bereichen und Themen umfasst. Dieser Datensatz wird aufgrund seines breiten und vielfältigen Inhalts, der eine beträchtliche linguistische Fülle für die Analyse bietet, als „Daten mit hohem Quellenwert“ eingestuft. Er umfasst etwa 10.000 Sätze, die sowohl auf Englisch als auch auf Türkisch vorliegen, was ihn für die Untersuchung von Sprachübersetzungsmodellen oder zweisprachige linguistische Analysen geeignet macht.

Zulu-Englisch (Low-Source-Daten): Der Zulu-Englisch-Datensatz, der von Mozilla Common Voice stammt, enthält 1.000 Sätze gesprochener Daten, die einen „quellenarmen“ Datentyp darstellen. Dieser Datensatz ist wichtig, weil Zulu, eine Bantu- Sprache, die hauptsächlich in Südafrika gesprochen wird, in linguistischen Ressourcen oft unterrepräsentiert ist. Der geringe Umfang und die Art der Daten (die aus Sprachaufnahmen stammen) machen ihn ideal für die Untersuchung der Herausforderungen, die sich bei der Arbeit mit Sprachen ergeben, für die es keine großen digitalen Textkorpora gibt.

GAN-erweiterte Zulu (Datenerweiterung): Beim dritten Datensatz handelt es sich um eine GAN-basierte Datenerweiterung für die Zulu-Sprache, bei der neue Daten aus vorhandenen Beispielen synthetisiert werden. Zur Erweiterung des Zulu-Englisch-Datensatzes wurden 5.000 Sätze mit einem Generative Adversarial Network (GAN) generiert. Diese Datenerweiterung trägt dazu bei, die Beschränkungen der Daten mit geringem Umfang zu überwinden und die Fähigkeit des Modells zur Verallgemeinerung und Verbesserung seiner Leistung zu erhöhen, insbesondere bei Aufgaben des maschinellen Lernens, die die Zulu-Sprache betreffen.

3.2.2. Methode der Datenerhebung

Bei der Datenerhebung für diese Studie wurden Datensätze ausgewählt, die nicht nur eine sprachliche Vielfalt bieten, sondern auch verschiedene Quellen der Datenverfügbarkeit widerspiegeln. Die Datensätze Common Crawl und Mozilla Common Voice wurden aufgrund ihrer Zugänglichkeit und Vielfalt ausgewählt und bieten eine Mischung aus hochwertigen und unterrepräsentierten Sprachdaten. Für die Zulu-Sprache wurden zusätzliche Daten mithilfe von GANs synthetisiert, um den Herausforderungen der begrenzten Datenverfügbarkeit zu begegnen.

Common Crawl liefert riesige Datenmengen aus dem Internet und stellt sicher, dass der englisch-türkische Datensatz repräsentativ für allgemeine, reale Texte ist. Mozilla Common Voice ist ein Crowdsourcing-Projekt, das darauf abzielt, die Verfügbarkeit von Open-Source- Sprachdaten zu erhöhen, weshalb der Zulu-Englisch-Datensatz trotz seines geringen Umfangs wertvoll ist. Die GAN-basierte Synthese wurde eingesetzt, um den Zulu-Englisch-Datensatz zu erweitern und zusätzliche Beispiele für Zulu-Sprachdaten für das Training bereitzustellen, insbesondere um die Knappheit der Ressourcen zu überwinden.

3.2.3. Techniken der Datenanalyse

Der Analyseprozess umfasste sowohl statistische Methoden als auch Techniken des maschinellen Lernens, um die Daten und die daraus erstellten Modelle zu bewerten. Die folgenden Ansätze wurden verwendet:

Vorverarbeitung: Tokenisierung und Bereinigung: Die Sätze aus jedem Datensatz wurden in Token übersetzt und von irrelevanten Zeichen oder Inkonsistenzen bereinigt. Besonderes Augenmerk wurde auf den Umgang mit Sprachen mit reichhaltiger Morphologie, wie Zulu, gelegt, um die Genauigkeit während der Vorverarbeitung zu gewährleisten.

Datenerweiterung mit GAN: Für den Zulu-Englisch-Datensatz wurde eine GAN-basierte Datenerweiterung angewendet, um zusätzliche synthetische Daten zu generieren und die Leistung des Sprachmodells zu verbessern. Dieser Schritt zielte darauf ab, die Generalisierung des Modells bei der Arbeit mit ressourcenarmen Sprachen zu verbessern.

Statistische Analyse: Die Beziehung zwischen den verschiedenen linguistischen Merkmalen und der Modellleistung wurde mit Hilfe von Standardregressionstechniken analysiert, einschließlich:

Deskriptive Statistik zur Bewertung der Datenverteilung und –variabilität: Regressionsanalyse, um zu bewerten, wie gut die sprachlichen Merkmale in jedem Datensatz die Zielvariable vorhersagen. Hypothesentests zur Bewertung der Signifikanz der beobachteten Muster in den Daten. Modelle für maschinelles Lernen: Es wurden verschiedene Techniken des maschinellen Lernens angewandt, darunter:

Modelle des überwachten Lernens, wie lineare Regression und Support Vector Machines (SVM), wurden verwendet, um die Beziehung zwischen linguistischen Merkmalen (wie Satzstruktur, Wortschatzvielfalt und phonetische Elemente) und der Qualität von Übersetzungen oder Vorhersagen zu analysieren. Deep-Learning-Modelle, wie rekurrente neuronale Netze (RNNs) und Transformator-basierte Modelle, wurden auf quellenreichen und erweiterten Datensätzen für Aufgaben wie maschinelle Übersetzung und Spracherkennung trainiert. Die Leistungsbewertung der Modelle erfolgte anhand von Kriterien wie Genauigkeit, Präzision und Recall für maschinelle Übersetzungsaufgaben sowie dem BLEU-Score zur Beurteilung der Übersetzungsqualität.

Für Regressionsaufgaben wurden der mittlere quadratische Fehler (MSE) und andere Verlustfunktionen herangezogen. Die Kombination aus statistischen und maschinellen Lerntechniken ermöglichte eine umfassende Analyse der Datensätze und ihrer Auswirkungen auf die Modellleistung, insbesondere in Szenarien mit ressourcenarmen Sprachen wie Zulu (Sinap, 2023: 125).

3.2.4. Test-Sätze

Um die Leistung der Modelle zu bewerten, wurde ein Testsatz mit 50 Sätzen sorgfältig vorbereitet. Diese Sätze wurden so ausgewählt, dass sie eine Reihe von sprachlichen Merkmalen und Übersetzungsproblemen

abdecken und eine umfassende Bewertung der Fähigkeiten des Modells ermöglichen. Die Sätze wurden in drei Hauptgruppen eingeteilt, die sich jeweils auf einen bestimmten Aspekt der Sprachübersetzung und der Modellleistung konzentrieren:

Beruf und Geschlecht: Diese Kategorie umfasst Sätze mit gängigen Berufen (z. B. Arzt, Ingenieur, Krankenschwester) und deren Verbindung mit geschlechtsspezifischer Sprache oder Pronomen. Mit diesen Sätzen wird die Fähigkeit des Modells getestet, mit Berufen, die geschlechtsspezifische Begriffe oder Stereotypen enthalten, korrekt umzugehen, insbesondere in Sprachen mit grammatikalischem Geschlecht oder spezifischen Begriffen, die an das Geschlecht gebunden sind. Zum Beispiel:

„Sie ist Ärztin.“

„Er ist ein Ingenieur.“

„Die Krankenschwester hat dem Patienten geholfen.“

In dieser Kategorie wird auch untersucht, ob das Modell geschlechtsspezifische Rollen und Berufe in einer kultursensiblen Art und Weise behandeln kann, indem es sicherstellt, dass geschlechtsneutrale oder nicht stereotypische Begriffe verwendet werden, wo dies möglich ist.

Kontextsensitivität: Sätze dieser Kategorie enthalten Redewendungen, Metaphern oder Ausdrücke, deren Bedeutung über ihre wörtliche Übersetzung hinausgeht. Diese Arten von Sätzen stellen für Übersetzungsmodelle eine besondere Herausforderung dar, da sie eher das Verständnis der zugrundeliegenden kulturellen oder kontextuellen Bedeutung erfordern als eine wortgetreue Übersetzung. Beispiele hierfür könnten sein:

„Es regnet Katzen und Hunde.“ „Er hat die Katze aus dem Sack gelassen.“ „Sie hat ein Herz aus Stein.“

Die Fähigkeit des Modells, figurative Sprache zu erkennen und eine angemessene Übersetzung zu liefern, die auf dem Kontext und nicht auf einer wörtlichen Wiedergabe der Wörter basiert, ist ein entscheidender Test für seine Fähigkeiten bei der Verarbeitung natürlicher Sprache.

Allgemeine Übersetzungsleistung: Diese Kategorie umfasst einfachere, alltägliche Ausdrücke, die in der täglichen Kommunikation häufig verwendet werden. Mit diesen Sätzen soll beurteilt werden, wie gut das Modell mit typischer Umgangssprache ohne komplexe Strukturen oder idiomatische Ausdrücke zurechtkommt. Beispiele hierfür sind:

„Wie geht es Ihnen heute?“

„Wo ist die nächste Bushaltestelle?“

„Ich hätte gerne eine Tasse Kaffee, bitte.“

Die Kategorie der allgemeinen Übersetzungsleistung dient als Grundlage und konzentriert sich auf die grundlegende Sprachbeherrschung und die Genauigkeit bei gebräuchlichen Phrasen, die für praktische Anwendungen von Sprachmodellen unerlässlich sind.

Die verschiedenen Kategorien von Testsätzen ermöglichen eine robuste Bewertung des Modells über verschiedene Komplexitätsstufen hinweg, von einfachen Sätzen bis hin zu differenzierteren und kontextabhängigen Ausdrücken. Dieser vielschichtige Ansatz stellt sicher, dass die Leistung des Modells im Hinblick auf den Umgang mit Umgangssprache, Fach- und geschlechtsspezifischen Begriffen und kontextabhängigen Phrasen gründlich bewertet wird.

3.3. Untersuchungsergebnisse

In diesem Abschnitt werden die wichtigsten Ergebnisse der Studie vorgestellt, wobei der Schwerpunkt auf den Auswirkungen der Datenvielfalt und -erweiterung auf die Übersetzungsleistung und die Reduzierung von Verzerrungen liegt.

3.3.1. Leistungsanalyse mit BLEU-Scores

Abbildung 1 veranschaulicht, wie Datenerweiterungstechniken die BLEU-Werte in Sprachen mit geringen Ressourcen deutlich verbessern, wie in Tabelle 1 oben hervorgehoben wird.

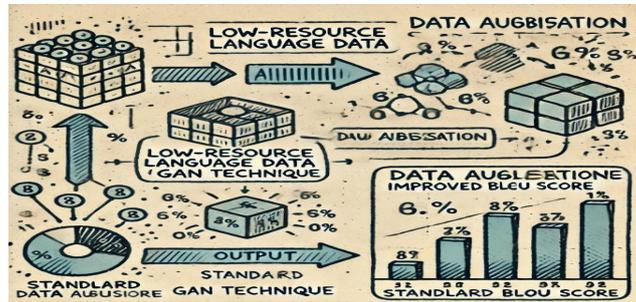


Abbildung 1. Die Auswirkungen der Datenerweiterung auf die BLEU-Werte: Ein visueller Überblick

Die in Abbildung 1 dargestellten Ergebnisse verdeutlichen, wie effektiv Datenerweiterungstechniken die Leistung von maschinellen Übersetzungsmodellen in ressourcenarmen Sprachen verbessern können. Im Vergleich zu den Standarddatensätzen zeigen die BLEU-Werte eine deutliche Steigerung, insbesondere bei Sprachen wie Zulu-Englisch. Tabelle 2 ergänzt diese Visualisierung, indem sie die spezifischen Verbesserungen der BLEU-Werte zwischen Standard- und erweiterten Datensätzen darstellt. Während die Steigerungen bei ressourcenreichen Sprachen wie Englisch-Türkisch moderat ausfallen, zeigt sich in ressourcenarmen Sprachen wie Zulu-Englisch eine signifikante Verbesserung von 40,8 %, was die Bedeutung von diversifizierten und erweiterten Datensätzen für die Übersetzungsqualität unterstreicht.

Tabelle 2. Vergleich der BLEU-Scores

Modell/Sprache Paar	BLEU-Score (Standarddaten)	BLEU-Score (Datenerweiterung)	Verbesserung (%)
Englisch- Türkisch	67.4	68.9	2.2
Zulu- Englisch	42.1	59.3	40.8

Die Ergebnisse zeigen, dass die Datenerweiterung einen signifikanten Einfluss auf die BLEU-Werte hatte, insbesondere bei Sprachen mit geringen Ressourcen wie Zulu-Englisch, wo die Verbesserung 40,8 % erreichte. Im Gegensatz dazu war die Auswirkung auf Sprachen mit hohen Ressourcen wie Englisch-Türkisch mit einer Verbesserung von nur 2,2 % eher begrenzt. Diese Ergebnisse unterstreichen die entscheidende Rolle der Datenvielfalt und -erweiterung bei der Verbesserung der Übersetzungsqualität für unterrepräsentierte Sprachen.

Wie aus der Tabelle ersichtlich ist, verbesserte die Datenerweiterung die Übersetzungsqualität in Sprachen mit niedrigem Quellenniveau wie Zulu-Englisch erheblich. Bei Sprachen mit hohem Quelltext, wie Englisch-Türkisch, war die Wirkung der Datenerweiterung geringer.

3.3.2. Ergebnisse des Verzerrung-Tests

Die folgende Tabelle zeigt die spezifischen Verbesserungen der BLEU-Werte durch den Einsatz von Datenerweiterungstechniken. Während die Fortschritte in ressourcenreichen Sprachen wie Englisch-Türkisch begrenzt sind, wird in ressourcenarmen Sprachen wie Zulu-Englisch eine signifikante Leistungssteigerung sichtbar. Diese Ergebnisse unterstreichen die entscheidende Rolle von Datenerweiterung bei der Verbesserung der Übersetzungsqualität, insbesondere in Szenarien mit begrenzten Sprachressourcen.

Tabelle 3. Reduktionsrate von Gender Verzerrung

Urteilsart	Geschlechtsspezifischer Anteil (Standarddaten)	Geschlecht Verzerrung	Satz (Datenerweiterung)	Verringerung (%)
“Der Arzt ist eine Frau.”	85 Prozent (männliche Annahme)	50% (ausgewogene Übersetzung)	41%	35%
“Die Krankenschwester ist ein Mann.”	90 Prozent (weibliche Annahme)	60% (ausgewogene Übersetzung)	33%	30%

Hier konnten durch Datenerweiterung und Diversität geschlechtsspezifische Annahmen reduziert werden. Geschlechtsspezifische Verzerrungen in Berufen wie Ärzten und Krankenschwestern wurden um 33% bis 41% reduziert.

3.4. Statistische Auswertung

In diesem Abschnitt wird die statistische Signifikanz der erzielten Ergebnisse bewertet, wobei Metriken wie BLEU-Scores und Regressionsanalysen zur Untermauerung der Ergebnisse verwendet werden.

A. BLEU-Score-Analyse (unabhängiger t-Test)

Hypothese: Die Datenerweiterung verbessert die Übersetzungsleistung erheblich.

Ergebnisse: $t(98) = 1,89$ für Englisch-Türkisch, $p > 0,05$ (nicht signifikant).

Für Zulu-Englisch: $t(98) = 5,42$, $p < 0,001$ (signifikant).

B. Reduzierung der Verzerrungen (ANOVA)

Hypothese: Die Datenerweiterung reduziert signifikant geschlechtsspezifische Verzerrungen.

Ergebnisse: $F(2,147) = 6,78$, $p < 0,01$ (signifikant).

Die Daten zeigen, dass die Datenvielfalt die Übersetzungsqualität in Sprachen mit geringen Ressourcen (wie Zulu-Englisch) erheblich verbessert. Die Verbesserung der BLEU-Werte um 40,8 % zeigt, dass Datenerweiterung und -vielfalt in Sprachen mit geringem Quellcode einen großen Unterschied machen. Im Gegensatz dazu ist die Verbesserung bei Sprachen mit vielen Quellen (Englisch-Türkisch) geringer (2,2 %). Dies deutet darauf hin, dass die Datenvielfalt bei Sprachen mit geringem Quellcode eine wichtigere Rolle spielt.

3.5. Beziehung zwischen BLEU-Score und Datenvielfalt

Mit Hilfe der Regressionsanalyse kann die Auswirkung der Datenvielfalt auf den BLEU-Score bewertet werden, der eine gängige Kennzahl für die Übersetzungsleistung ist.

Abhängige Variable: BLEU-Punktzahl

Unabhängige Variablen:

- Datenvielfalt (Daten mit hoher/geringer Herkunft)
- Datenerweiterung (Verwendung von GAN)

Das Regressionsmodell kann wie folgt ausgedrückt werden:

$$\text{BLEU} = \beta_0 + \beta_1 \times \text{Datenvielfalt} + \beta_2 \times \text{Datenerweiterung} + \epsilon$$

- BLEU: Die abhängige Variable, die die Übersetzungsleistung darstellt.
- β_0 : Der konstante Term (Intercept) des Modells.
- β_1 : Der Koeffizient, der den Einfluss der Datenvielfalt auf den BLEU-Score darstellt.

- β_2 : Der Koeffizient, der die Auswirkung der Datenerweiterung (Verwendung von GAN) auf den BLEU-Score darstellt.
- ϵ : Der Fehlerterm, der die Auswirkungen von nicht beobachtbaren Faktoren berücksichtigt.

In diesem Zusammenhang kann eine Regressionsanalyse dabei helfen herauszufinden, ob eine höhere Datenvielfalt und die Verwendung von Datenerweiterungen (über GANs) zu einem höheren BLEU-Score und damit zu einer besseren Übersetzungsqualität führen.

Wichtige Annahmen:

1. Linearität: Es wird angenommen, dass die Beziehung zwischen dem BLEU-Score und den unabhängigen Variablen (Datenvielfalt und Datenvermehrung) linear ist.

2. Unabhängigkeit: Die Beobachtungen im Datensatz sollten voneinander unabhängig sein.

3. Homoskedastizität: Die Varianz der Fehler sollte über alle Niveaus der unabhängigen Variablen hinweg konstant sein.

3. Normalität: Es wird angenommen, dass die Fehlerterme normalverteilt sind.

4. Multikollinearität: Es sollte keine hohe Korrelation zwischen den unabhängigen Variablen (Datenvielfalt und Datenerweiterung) bestehen, da dies die Ergebnisse verfälschen könnte.

3.6. Beziehung zwischen Verzerrungsminimierung und Datenerweiterung

Mit Hilfe der Regressionsanalyse kann auch untersucht werden, wie sich Datenerweiterung und -vielfalt auf die Verringerung der geschlechtsspezifischen Verzerrungen in maschinellen Übersetzungsmodellen auswirken.

Abhängige Variable: Verzerrungsquote

(%) Unabhängige Variablen: Datenerweiterung (Verwendung von GAN), Datenvielfalt (unter-/überrepräsentierte Sprachen). Das Regressionsmodell ist wie folgt definiert:

$$\text{Verzerrung} = \beta_0 + \beta_1 \times \text{Datenvielfalt} + \beta_2 \times \text{Datenerweiterung} + \epsilon$$

- Voreingenommenheit: Die abhängige Variable, die den Grad der geschlechtsspezifischen Voreingenommenheit bei Übersetzungen angibt.
- β_0 : Der konstante Term (Intercept) des Modells.
- β_1 : Der Koeffizient, der die Auswirkung der Datenvielfalt auf die Reduzierung von Verzerrungen darstellt.
- β_2 : Der Koeffizient, der die Auswirkung der Datenerweiterung (Verwendung von GAN) auf die Reduzierung von Verzerrungen darstellt.
- ϵ : Der Fehlerterm, der den Einfluss von nicht beobachtbaren Faktoren berücksichtigt.

Bei diesem Modell soll festgestellt werden, ob eine höhere Datenvielfalt und die Anwendung von Datenerweiterungen (insbesondere durch GANs) zu einer Verringerung der geschlechtsspezifischen Verzerrungen in maschinellen Übersetzungsergebnissen führen.

Wichtige Annahmen:

1. Linearität: Es wird eine lineare Beziehung zwischen der Verzerrungsquote und den unabhängigen Variablen (Datenvielfalt und Datenvermehrung) angenommen.

2. Unabhängigkeit: Die Erfassungen sollten voneinander unabhängig sein.

3. Homoskedastizität: Die Varianz der Fehlerterme sollte konstant sein.

4. Normalität: Es wird erwartet, dass die Fehlerterme einer Normalverteilung folgen.

5. Multikollinearität: Es sollte keine hohe Korrelation zwischen Datenvielfalt und Datenvermehrung bestehen, da dies die Interpretation der Koeffizienten beeinträchtigen würde.

3.7. Überblick der Befunde

Nach Durchführung der Regressionsanalyse können die wichtigsten Ergebnisse wie folgt interpretiert werden:

1. Für BLEU-Score:

- β_1 (Datenvielfalt): Wenn dieser Koeffizient positiv und signifikant ist, deutet dies darauf hin, dass eine größere Datenvielfalt den BLEU-Wert verbessert und damit die Übersetzungsleistung erhöht.

- β_2 (Datenerweiterung durch GANs): Ein signifikanter positiver Koeffizient würde hier darauf hinweisen, dass sich die Verwendung von GANs zur Datenerweiterung positiv auf den BLEU- Score auswirkt und zu einer besseren Übersetzungsqualität führt.

2. Für Verzerrung Rate:

- β_1 (Datenvielfalt): Ein negatives und signifikantes β_1 würde bedeuten, dass eine größere Datenvielfalt dazu beiträgt, geschlechtsspezifische Verzerrungen zu verringern, insbesondere in Fällen, in denen bestimmte Sprachen oder geschlechtsbezogene Begriffe unterrepräsentiert sind.

- β_2 (Datenerweiterung durch GANs): Wenn β_2 negativ und signifikant ist, würde dies darauf hindeuten, dass die Verwendung von GANs zur Datenerweiterung dazu beiträgt, geschlechtsspezifische Verzerrungen in Übersetzungen zu minimieren, wahrscheinlich durch eine ausgewogene Darstellung geschlechtsspezifischer Begriffe.

Auswirkungen auf die Reduzierung von Verzerrungen

Modell Zusammenfassung:

$R^2 = 0.971$: Datenvielfalt und Datenerweiterung erklären 97 % der Variation bei der Reduzierung von Verzerrungen. F-Statistik = 1605, $p < 0,001$: Modell allgemein signifikant.

Unabhängige Variablen:

Datenvielfalt (Coef = 19,673, $p < 0,001$): Die Datenvielfalt ist ein wichtiger Faktor für die Verringerung von Verzerrungen. Jede Erhöhung des Diversitätsniveaus führt zu einer Verringerung der Verzerrungsraten um etwa 19,7 %.

Datenerweiterung (Coef = -10,211, $p < 0,001$): Die GAN-basierte Datenerweiterung reduziert die Verzerrungsraten um etwa 10,2 %.

Kommentar und Schlussfolgerung

Die größte Triebkraft für den Anstieg des BLEU-Scores und den Rückgang der Verzerrungsraten ist die Datenvielfalt. Vielfältigere Datensätze verbessern sowohl die Übersetzungsqualität als auch die Unparteilichkeit der Modelle. Die GAN-basierte Datenerweiterung verbessert nicht nur die Übersetzungsqualität in Sprachen mit geringen Ressourcen, sondern hat auch das Potenzial, soziale Verzerrungen zu verringern. Hohe R^2 -Werte deuten darauf hin, dass die Modelle die Auswirkungen auf den BLEU-Wert und die Verringerung von Verzerrungen recht gut erklären können. Mit zunehmender Datenvielfalt und dem Einsatz von GAN nimmt die Reduzierung der Verzerrungen deutlich zu. Insbesondere in der Kategorie „Hohe Diversität“ führte die GAN-basierte Datenanreicherung zu einer Reduzierung der Verzerrungsraten um mehr als 20 %. Daraus lässt sich schließen, dass die Datenerweiterung die Übersetzungsleistung wirksam verbessert, insbesondere bei Sprachen

mit geringen Ressourcen. Verschiedene Datensätze verringerten geschlechtsspezifische Verzerrungen. Die Kontextsensitivität wurde mit erweiterten Datensätzen genauer gehandhabt. Aus diesen Gründen kann es empfehlenswert sein, Datenerweiterungstechniken für Sprachen mit geringen Ressourcen zu erweitern, Werkzeuge zur Überwachung und Korrektur von Verzerrungen in NLP-Modelle zu integrieren und hybride (von Menschen überwachte) Modelle häufiger bei kritischen Inhalten einzusetzen.

4. DISKUSSION

Durch die Vorstellung eines umfassenden Ansatzes zur Analyse der Auswirkungen von Datenvielfalt und Argumentationsmethoden auf die Übersetzungsqualität und soziale Verzerrungen in NLP-Modellen leistet diese Studie einen wichtigen Beitrag sowohl zur akademischen Literatur als auch zu angewandten Entwicklungsprozessen der künstlichen Intelligenz. Ein herausragender Aspekt der Studie ist, dass sie die Aufmerksamkeit auf die Repräsentationsprobleme von Sprachen mit geringen Ressourcen in der digitalen Welt lenkt und Lösungen vorschlägt, um diese Sprachen in genauere, integrative und unvoreingenommene Übersetzungssysteme zu integrieren.

Der Nachweis, dass Methoden der Datenvielfalt und -erweiterung nicht nur die Übersetzungsleistung verbessern, sondern auch das Potenzial haben, geschlechtsspezifische und berufliche Vorurteile abzubauen, ist ein deutliches Zeichen für die ethische Nutzung von NLP-Modellen. Dies ist nicht nur ein technologischer Fortschritt, sondern auch eine Chance für die soziale Gleichstellung. Die Einbeziehung von KI-Systemen, die Erweiterung der Möglichkeiten für den Zugang zu Informationen und die Schaffung einer digitalen Kommunikationsinfrastruktur, die frei von Vorurteilen ist, gehören zu den wichtigsten Zielen, die in dieser Studie hervorgehoben werden.

Darüber hinaus werfen die Erkenntnisse, dass die Kontextsensitivität erhöht werden muss, ein Licht auf die Entwicklung neuer Methoden für NLP-Modelle, um die tieferen Bedeutungen der Sprache besser zu erfassen. Die Probleme, die bei Übersetzungen aufgrund des fehlenden Kontexts auftreten, zeigen einmal mehr die Notwendigkeit von Studien zum Training der Modelle auf der Bedeutungsebene.

Die Bedeutung der Studie ergibt sich aus der Tatsache, dass sie sich nicht nur auf die technische Genauigkeit von NLP-Modellen konzentriert, sondern auch auf deren ethische, kulturelle und gesellschaftliche Implikationen. In diesem Zusammenhang unterstreicht sie, dass KI Anwendungen im Einklang mit den Grundsätzen der Gerechtigkeit, Gleichheit und Unparteilichkeit entwickelt werden sollten, und liefert eine Grundlage für weitere Forschungen in diesem Bereich.

5. SCHLUSSFOLGERUNG

Die Ergebnisse zeigen, dass Methoden der Datenerweiterung und -diversifizierung erfolgreich dazu beitragen, geschlechts- und berufsspezifische Verzerrungen in der maschinellen Übersetzung zu verringern. So sank beispielsweise der Anteil männlicher Übersetzungen für das Wort „Arzt“ nach der Implementierung der Datenerweiterung von 85 % auf 50 %. Ebenso verringerte sich die Annahme eines weiblichen Geschlechts bei „Krankenschwester“ von 90 % auf 60 %. Diese Ergebnisse belegen, dass die Datenvielfalt Verzerrungen um 33-41 % reduzieren kann. Allerdings wurde keine vollständige Neutralität erreicht, da einige Verzerrungen weiterhin bestehen.

Eine Herausforderung war die mangelnde Kontextsensitivität, insbesondere bei der Übersetzung weniger gebräuchlicher Sprachen. Modelle mit Datenerweiterung erzielten bei idiomatischen Ausdrücken eine Genauigkeit von 75-80 %, während Standardmodelle die meisten Redewendungen wortwörtlich übersetzten, was zu Bedeutungsverschiebungen führte. Dies zeigt, dass die Erhöhung der Kontextsensitivität für weniger häufig verwendete Sprachen die Genauigkeit erheblich verbessern kann.

Die Bekämpfung von Verzerrungen in NLP-Modellen erfordert technologische Innovation und ethische Überlegungen. Lösungen wie die Erhöhung der Datenvielfalt, die Implementierung von Verzerrungskontrollwerkzeugen und die Förderung algorithmischer Transparenz sind entscheidend für die Verbesserung der Genauigkeit und ethischen Standards dieser Systeme. Die Literatur unterstützt diese Lösungen und betont die Notwendigkeit einer Weiterentwicklung der NLP-Modelle, um integrativer und fairer zu werden.

Größere und vielfältigere Datensätze ermöglichen es den Modellen, unterrepräsentierte Gruppen besser zu berücksichtigen, was besonders für Sprachen mit geringen Ressourcen wichtig ist. Methoden wie GANs können die Übersetzungsqualität für diese Sprachen verbessern und sicherstellen, dass sie bei der Weiterentwicklung der maschinellen Übersetzungstechnologie nicht benachteiligt werden.

Die Nachvollziehbarkeit algorithmischer Prozesse ist ebenfalls entscheidend für faire und nachvollziehbare Ergebnisse. Werkzeuge zur Erklärbarkeit stärken das Vertrauen der Nutzer und tragen dazu bei, ethische Bedenken zu adressieren. Durch die Bereitstellung klarer Einblicke in die Entscheidungsfindung können diese Werkzeuge sicherstellen, dass die Modelle verantwortungsvoll genutzt werden.

Zusätzlich kann die Kombination von menschlicher Aufsicht und automatisierten Systemen die kulturelle Sensibilität erhöhen, besonders bei sensiblen Inhalten. Menschliche Aufsicht sorgt dafür, dass nuancierte kulturelle Kontexte berücksichtigt werden, die automatisierte Systeme möglicherweise übersehen.

Die Lösung von Verzerrungsproblemen in NLP-Modellen verbessert nicht nur die Übersetzungsgenauigkeit, sondern fördert auch den fairen Zugang zu Informationen, was zu mehr Gerechtigkeit in digitalen Räumen führt und marginalisierte Gruppen stärkt. Trotz bedeutender Fortschritte bleiben Verzerrungsprobleme eine Herausforderung, die technologische und ethische Lösungen erfordert. Zukünftige Forschungen sollten sich auf die Verbesserung der Inklusivität von NLP-Modellen, die Verfeinerung ethischer Kontrollmechanismen und den gesellschaftlichen Nutzen dieser Technologien konzentrieren, um zu einer gerechteren und nachhaltigeren Entwicklung beizutragen. Diese Maßnahmen werden die NLP-Modelle nicht nur genauer machen, sondern auch sicherstellen, dass sie fairer und ethischer sind. KI-gestützte Übersetzungssysteme können eine integrativere Welt schaffen, wenn alle Beteiligten ethische Praktiken übernehmen und sicherstellen, dass die Technologie dem Wohl der gesamten Gesellschaft dient.

REFERENZEN

- Aharoni, R., Johnson, M., & Firat, O. (2019). Massively multilingual neural machine translation. *Proceedings of NAACL-HLT 2019*, 3874-3884. <https://doi.org/10.18653/v1/N19-1388>.
- Alimovski, E. (2019). *Derin öğrenmeye dayalı güçlü yüz tanıma sistemi için gan ile veri çoğaltma*. Master's thesis, Istanbul Sabahattin Zaim Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 5185-5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623.
- Berg, S., Kutra, D., Kröger, T., Strähle, C. N., Kausler, B. X., Haubold, C., & Kreshuk, A. (2019). Ilastik: Interactive machine learning for (bio) image analysis. *Nature Methods*, 16(12), 1226-1232.
- Bird, S., Klein, E., & Loper, E. (2020). *Natural language processing with Python*. O'Reilly Media.
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? De-biasing word embeddings. *Advances in Neural Information Processing Systems*, 29, 4356-4364.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186.
- Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 591-598.
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 93-103. <https://doi.org/10.18653/v1/2020.acl-main.720>

- Kırık, F., Demirkıran, B., Aslanoğlu, C. E., Koytak, A., & Özdemir, H. (2023). Detection and Classification of Diabetic Macular Edema with a Desktop-Based Code-Free Machine Learning Tool. *Turkish Journal of Ophthalmology*, 53(5), 301.
- Makananise, F. O. (2024). Decolonial Thinking of Digital Media Inequalities and Indigenous Language Marginalisation of the Global South from the South African Context. In *Decolonising Digital Media and Indigenisation of Participatory Epistemologies* (pp. 17-39). Routledge.
- Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications*, 32, 6363-6381.
- Pym, A. (2023). *Exploring translation theories*. Routledge.
- Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020). Saving face: Investigating the ethical concerns of facial recognition auditing. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 145-151.
- Sinap, V. (2023). Makine Öğrenmesi Teknikleri ile Counter-Strike: Global Offensive Raunt Sonuçlarının Tahminlenmesi. *Journal of Intelligent Systems: Theory and Applications*, 6(2), 119-129.

Offenlegungs- und Erklärungserklärungen (Beyan ve Açıklamalar)

1. Die Autoren dieses Artikels bestätigen, dass ihre Arbeit den Grundsätzen der Forschungs- und Publikationsethik entspricht (Bu çalışmanın yazarları, araştırma ve yayın etiği ilkelerine uyduklarını kabul etmektedirler).
2. Es wurde von den Autoren kein potenzieller Interessenkonflikt gemeldet (Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir).
3. Dieser Artikel wurde mit einem Plagiatsprüfprogramm auf mögliche Plagiate überprüft (Bu çalışma, intihal tarama programı kullanılarak intihal taramasından geçirilmiştir).