fluency and accessibility, making texts engaging for English-speaking readers, though some stylistic features and subtle cultural elements were occasionally diminished [6;45–46 p.].

Table 1

Examples of Domestication of Uzbek Expressions into English

Uzbek Expression	Literal Translation	Domesticated English Translation	Notes on Adaptation
Baliq boshidan sasiydi	The fish rots from the head	A fish rots from the head down	Preserves figurative meaning
Besh qo'l barobar emas	Five hands are not equal	Not everyone is the same	Simplified idiomatic meaning
Teng tengi bilan, Tezak qopi bilan	Like with like, hornet with its sting	Birds of a feather flock together	Cultural equivalence applied
Navruz	Spring festival	Persian New Year	Provides cultural context
Chak-chak	Uzbek sweets	Uzbek honey sweets	Descriptive adaptation

Conclusion. This study examined the use of domestication in Uzbek–English translation, focusing on proverbs, idioms, metaphors, and culturally specific references. The analysis of twenty literary, media, and educational texts shows that domestication is widely applied to enhance readability, fluency, and audience comprehension. Proverbs such as "Baliq boshidan sasiydi" were rendered as "A fish rots from the head down", and cultural items like Navruz and chak-chak were translated as "spring festival" and "Uzbek honey sweets", making culturally bound concepts accessible to English-speaking readers.

While domestication improves understanding and engagement, it may reduce stylistic and cultural nuances. Translators must therefore balance clarity with cultural preservation. In literary texts, combining domestication with selective foreignization–retaining key cultural elements with brief explanations—can preserve stylistic and cultural richness, whereas media and educational texts often prioritize fluency and comprehension.

Overall, this study highlights the translator's dual role as both mediator and interpreter, demonstrating that domestication is a practical and strategic approach in cross-cultural translation. Future research could expand the corpus, examine audience reception, or explore the combined use of domestication and selective foreignization to further optimize cultural and linguistic fidelity.

References:

(1). Baker M. (2011). In other words: A coursebook on translation (2nd ed.). Routledge.
(2). Bassnett S. (2014). Translation studies (4th ed.). Routledge.
(3). Chiaro D. (2006). Translation, humour and the media. Continuum.
(4). Gambier Y., van Doorslaer L. (2016). Handbook of translation studies (Vol. 2). John Benjamins.
(5). Hatim B., Mason I. (1997). The translator as communicator. Routledge.
(6). House J. (2015). Translation quality assessment. Past and present. Routledge.
(7). Koller W. (2001). Die Skopostheorie und ihre Anwendung in der Praxis. St. Jerome Publishing.
(8). Newmark P. (1988). A textbook of translation. Prentice Hall.
(9). Nida E.A. (1964). Toward a science of translating. Brill.
(10). Shuttleworth M., Cowie M. (2014). Dictionary of translation studies (3rd ed.). Routledge.
(11). Venuti L. (1995). The translator's invisibility: A history of translation. Routledge.

Rustamova Ruxsora Rasulbek qizi (Studentin im 2. Studienjahr der Fakultät "Fremdsprachen" an der Urgantscher Staatlichen Universität nach Abu Rayhon Beruni; E-mail: r.ruhsoraa@icloud.com)

ANALYSE DER SICHTBARKEIT DER DISKRIMINIERUNG IN TRAININGSDATEN FÜR SPRACHMODELLE UND DEREN FOLGEN

Anmerkung. *Diese Arbeit analysiert die Sichtbarkeit und Formen von Diskriminierung in Trainingsdaten für Sprachmodelle sowie deren potenzielle Auswirkungen auf die Ergebnisse künstlicher Intelligenz. Der Fokus liegt auf der Identifikation von Verzerrungen (Bias) in großen Textkorpora, die soziale Ungleichheiten, stereotype Darstellungen oder strukturelle Diskriminierung widerspiegeln können. Die Studie untersucht, wie solche Verzerrungen während des Trainingsprozesses entstehen und in der Modellleistung reproduziert oder verstärkt werden. Darüber hinaus werden die gesellschaftlichen, ethischen und technologischen Folgen diskriminierender Sprachmodelle diskutiert. Abschließend werden*

Ansätze zur Erkennung, Bewertung und Reduzierung von Diskriminierung in Trainingsdaten vorgestellt, um die Entwicklung fairerer und transparenterer KI-Systeme zu fördern.

Schlüsselwörter: *Diskriminierung, Sprachmodelle, Trainingsdaten, künstliche Intelligenz, algorithmische Verzerrung (Bias), Datenvoreingenommenheit, Fairness in KI, ethische Herausforderungen, soziale Ungleichheit, Transparenz von Algorithmen, verantwortungsvolle KI, Modellrobustheit.*

Rustamova Ruxsora Rasulbek qizi (Abu Rayhon Beruniy nomidagi Urganch davlat universiteti "Xorijiy filologiya" fakulteti 2-kurs talabasi; e-mail: r.ruhsoraa@icloud.com)

TIL MODELLARI UCHUN O'QUV MA'LUMOTLARIDA DISKRIMINATSIYANING NAMOYON BO'LISHI VA UNING OQIBATLARINI TAHLIL QILISH

Annotatsiya. *Mazkur maqolada til modellari uchun ishlatiladigan trening ma'lumotlarida diskriminatsiyaning ko'rinishlari va ularning oqibatlari tahlil qilinadi. Tadqiqot katta hajmdagi matnlar bazasida mavjud bo'lishi mumkin bo'lgan ijtimoiy tengsizliklar, stereotiplar va tizimli kamsitishlarni aks ettiruvchi og'ishlarni aniqlashga qaratilgan. Ish jarayonida bunday og'ishlar til modellarini o'qitish bosqichida qanday shakllanishi va natijalarda qay darajada takrorlanishi yoki kuchayishi o'rganiladi. Shuningdek, diskriminatsiyaga moyil sun'iy intellekt tizimlarining axloqiy, ijtimoiy va texnologik oqibatlari yoritib beriladi. Maqola yakunida trening ma'lumotlarida diskriminatsiyani aniqlash, baholash va kamaytirish bo'yicha yondashuvlar taklif etiladi.*

Kalit so'zlar: *diskriminatsiya, til modellari, o'quv (trening) ma'lumotlari, sun'iy intellekt, algoritmik og'ish (bias), ma'lumotlardagi xolislik muammosi, adolatli sun'iy intellekt, axloqiy masalalar, ijtimoiy tengsizlik, algoritmlar shaffofligi, mas'uliyatli AI, model barqarorligi.*

Рустамова Рухсора Расулбек кизи (студентка 2 курса факультета «Иностранные языки» Ургенчского государственного университета имени Абу Райхона Беруни; e-mail: r.ruhsoraa@icloud.com)

АНАЛИЗ ПРОЯВЛЕНИЙ ДИСКРИМИНАЦИИ В ОБУЧАЮЩИХ ДАННЫХ ДЛЯ ЯЗЫКОВЫХ МОДЕЛЕЙ И ИХ ПОСЛЕДСТВИЙ

Аннотация. *В данной статье анализируется проявление дискриминации в обучающих данных языковых моделей и ее последствия для функционирования систем искусственного интеллекта. Особое внимание уделяется выявлению предвзятостей, содержащихся в крупных текстовых корпусах, которые могут отражать социальное неравенство, стереотипы и структурную дискриминацию. Исследуется, каким образом данные искажения формируются в процессе обучения и как они воспроизводятся или усиливаются в результатах работы языковых моделей. Также рассматриваются этические, социальные и технологические последствия использования дискриминационных моделей. В заключение предлагаются подходы к выявлению, оценке и снижению дискриминации в обучающих данных с целью разработки более справедливых и прозрачных систем искусственного интеллекта*

Ключевые слова: *дискриминация, языковые модели, обучающие данные, искусственный интеллект, алгоритмическая предвзятость, смещение данных, справедливость в ИИ, этические вызовы, социальное неравенство, прозрачность алгоритмов, ответственное использование ИИ, устойчивость моделей.*

Einleitung/Kirish/Введение. Sprachmodelle sind zu einem Bestandteil der modernen digitalen Landschaft geworden und bilden das Rückgrat zahlreicher Anwendungen, die unser tägliches Leben prägen. Von der Art und Weise, wie wir nach Informationen suchen und mit virtuellen Assistenten interagieren, über die Automatisierung von Übersetzungsdiensten bis hin zur Gewährleistung der Sicherheit von Online-Inhalten und der Unterstützung komplexer Entscheidungsprozesse – die Reichweite und der Einfluss dieser Modelle sind immens. Ihre beeindruckende sprachliche Kompetenz wird durch das Training auf riesigen Mengen von Textdaten erreicht, die sorgfältig aus einer Vielzahl von Quellen gesammelt wurden, darunter das unendliche Universum des Internets, die Weisheit von Büchern, die Tiefe wissenschaftlicher Publikationen und eine Fülle anderer von Menschen erzeugter Inhalte.

Diese Skalierung des Trainingsdatensatzes ist zwar die Quelle ihrer bemerkenswerten sprachlichen Fähigkeiten, birgt aber auch eine tiefgreifende und kritische Schwachstelle. Die riesigen Texte, auf denen diese Modelle trainiert werden, sind keine neutralen Behälter von Wissen. Vielmehr sind sie ein Spiegelbild der komplexen und oft widersprüchlichen Gesellschaft, aus der sie stammen. Sie spiegeln unweigerlich die historischen und gesellschaftlichen Ungleichheiten wider, die tief in unseren Kulturen verwurzelt sind. Infolgedessen können Vorurteile wie Diskriminierung, die unsere Gesellschaft durchdringen, unbeabsichtigt in das Verhalten und die Ausgaben dieser Sprachmodelle eingebettet werden. Dies bedeutet,

dass die Modelle nicht nur Sprache lernen, sondern auch die subtilen und nicht so subtilen Muster von Vorurteilen und Diskriminierung, die in den Daten vorhanden sind.

Dieser Aufsatz untersucht, wie in Trainingsdaten kodierte diskriminierende Muster in Sprachmodellen sichtbar werden, was das für uns bedeutet und was man dagegen tun kann. Es wird noch in Kauf genommen, dass das Ganze nicht nur ein technisches Artefakt ist, sondern auch betrifft die ganze Gesellschaft und braucht deshalb verschiedene Lösungen.

Literaturübersicht/Adabiyotlarning tahlili/Обзор литературы. Die genannten Literaturverzeichnisse und Arbeiten zeigen gemeinsam, dass KI und Algorythmensysteme bestehende gesellschaftliche Vorurteile aufnehmen und verstärken können. Aus Caliskan und Bolukbasi lässt sich lernen, dass selbst "neutrale" Sprachmodelle Geschlechter- und Rassismus-Stereotype in ihren Repräsentationen tragen und dass Debiasing nur begrenzt Abhilfe schafft.

Noble sowie Buolamwini und Gebru machen deutlich, dass solche Verzerrungen reale Folgen haben, etwa diskriminierende Suchergebnisse oder unzuverlässige Gesichtserkennung bei marginalisierten Gruppen. Bender et al. und Mitchell et al. betonen schließlich, dass große Modelle transparent dokumentiert werden müssen, um Risiken, Einsatzgrenzen und Fairness-Probleme überhaupt erkennbar und verantwortbar zu machen.

Forschungsmethodik/Tadqiqot metodologiyasi/Методология исследования. Sprachmodelle erlernen Sprache durch das Erkennen statistischer Muster in Texten. Anstatt die Bedeutung von Wörtern wie ein Mensch zu erfassen, prägen sie sich Assoziationen zwischen Wörtern, Phrasen und Kontexten ein. Da die menschliche Sprache soziale Normen und Machtverhältnisse widerspiegelt, kodieren diese Assoziationen häufig Stereotypen. Zum Beispiel verbinden große Textkorpora oft Frauen mit häuslichen Rollen und Männer mit technischen oder Führungspositionen. Wenn Sprachmodelle diese Muster aufnehmen, können sie diese so wiedergeben, dass sie für Nutzende natürlich oder autoritativ wirken. Studien zeigen, dass Worteinbettungen, die mit großen Datenmengen trainiert wurden, geschlechts-, rassische- und kulturspezifische Vorurteile reproduzieren, welche durch Assoziationstests messbar sind. Das zeigt, dass diskriminierende Strukturen von Entwicklern gelernt werden, anstatt absichtlich eingeführt zu werden.

Diese erlernten Vorurteile werden sowohl in generativen als auch in analytischen Aufgaben sichtbar. Bei der Textgenerierung kann ein Sprachmodell stereotypische Fortsetzungen hervorrufen, wie z.B. bestimmte Berufe unverhältnismäßig mit Männern zu assoziieren oder Frauen häufiger zu beschreiben, die eine auf Aussehen ausgerichtete Sprache verwenden als Männer. Bei analytischen Aufgaben wie dem Sortieren oder Einordnen von Dingen können Vorurteile zu Problemen führen. Beispielsweise können Programme, die Sprachmodelle nutzen, um Bewerbungen zu prüfen oder Meinungen zu analysieren, Menschen aus bestimmten Gruppen benachteiligen. Das liegt daran, dass die Sprache, die diese Menschen verwenden, mit Geschlecht, Herkunft oder sozialem Status zusammenhängen kann.

Analyse und Ergebnisse/Tahlil va natijalar/Анализ и результаты. Studien haben gezeigt, dass solche Unterschiede auch dann bestehen können, wenn geschützte Attribute nicht explizit im Modell enthalten sind, da Proxy-Variablen, die in Sprache eingebettet sind, als indirekte Signale dienen.

Die Voreingenommenheit wird weiter verstärkt, wenn Sprachmodelle in großem Maßstab eingesetzt werden. Eine einzelne voreingenommene Ausgabe mag trivial erscheinen, aber wenn ähnliche Ausgaben millionenfach über Plattformen hinweg generiert werden, können sie Normen und Erwartungen prägen. Dieser Verstärkungseffekt ist besonders besorgniserregend in pädagogischen, beruflichen und informativen Kontexten, in denen Benutzer Modellausgaben als neutral oder autoritativ behandeln können. Weil die Modelle oft mit Daten trainiert werden, die von älteren, fehlerhaften Systemen stammen, kann eine Rückkopplungsschleife entstehen. Die voreingenommenen Ergebnisse beeinflussen dann die zukünftigen Trainingsdaten, wodurch die Diskriminierung mit der Zeit schlimmer und fester wird.

Die Folgen für die Nutzer sind vielfältig. Auf individueller Ebene kann diskriminierendes Modellverhalten zu falscher Darstellung, Ausgrenzung oder Verweigerung von Möglichkeiten führen. Beispielsweise können voreingenommene automatisierte Screening-Tools den Zugang zu Beschäftigung oder Finanzdienstleistungen für marginalisierte Gruppen verringern. In Konversationssystemen können Benutzer auf beleidigende oder abweisende Reaktionen stoßen, die ihre Würde untergraben oder von Engagement abhalten. Untersuchungen zu demografischen Leistungslücken in automatisierten Systemen, wie z. B. Gender-Klassifizierungstools, veranschaulichen, wie technische Fehler bereits marginalisierte Bevölkerungsgruppen unverhältnismäßig stark beeinträchtigen und bestehende Ungleichheiten verschärfen.

Wenn Sprachmodelle Vorurteile haben, kann das in der Gesellschaft gefährlich sein, weil Stereotypen normalisiert und Diskriminierung verstärkt werden. Wenn Technik bestimmte Gruppen immer wieder als kompetent und intelligent darstellt, andere aber ausgrenzt, werden soziale Unterschiede verstärkt. Das Problem ist, dass diese Vorurteile oft unbemerkt in scheinbar neutrale Interaktionen einfließen und so den öffentlichen Diskurs, kulturelle Ansichten und sogar politische Entscheidungen beeinflussen können. Die Auswirkungen gehen also weit über den einzelnen Nutzer hinaus.

Das Vorhandensein von Diskriminierung in Sprachmodellen hat auch Auswirkungen auf Vertrauen. Benutzer, die voreingenommen behandelt werden, können das Vertrauen in digitale Systeme und die Institutionen, die sie einsetzen, verlieren. Für Unternehmen geht diese Erosion des Vertrauens mit rechtlichen und Reputationsrisiken einher. In vielen Gerichtsbarkeiten können diskriminierende Ergebnisse in Bereichen wie Einstellung, Kreditvergabe oder Zugang zu Dienstleistungen gegen Antidiskriminierungsgesetze verstoßen, selbst wenn die Voreingenommenheit indirekt durch automatisierte Systeme entsteht. Folglich ist die Voreingenommenheit des Sprachmodells nicht nur ein ethisches Anliegen, sondern auch eine Frage der rechtlichen und organisatorischen Verantwortung.

Die Bewältigung dieser Herausforderungen erfordert eine Kombination von Gegenmaßnahmen. Ein wichtiger Ansatz konzentriert sich auf Trainingsdaten. Die Verbesserung der Datenkuration, -dokumentation und -herkunft kann Entwicklern helfen zu verstehen, wessen Stimmen dargestellt werden und wessen fehlt. Techniken wie Datensatzdokumentation und Transparenzberichte machen Vorurteile sichtbarer und ermöglichen gezielte Interventionen. Datenzentrierte Ansätze allein reichen jedoch nicht aus, da sie strukturelle Ungleichheiten, die in der Sprache selbst eingebettet sind, nicht vollständig beseitigen können.

Diese Methoden können zwar einige Vorurteile abschwächen, aber oft geht das auf Kosten der Leistung des Modells. Außerdem können sie subtilere oder situationsabhängige Diskriminierung nicht immer erkennen. Daher sollten sie als Teillösungen und nicht als endgültige Lösungen verstanden werden.

Ebenso wichtig sind Maßnahmen. Die disaggregierte Bewertung über demografische Gruppen hinweg ermöglicht es Entwicklern und Bereitstellungen, unterschiedliche Auswirkungen zu identifizieren, die aggregierte Metriken verbergen. Transparenzmechanismen, wie z. B.Modellkarten, kommunizieren beabsichtigte Verwendungen, Einschränkungen und bekannte Vorurteile an nachgelagerte Benutzer. Menschliche Aufsicht und zugängliche Beschwerdemechanismen reduzieren den Schaden weiter, indem sichergestellt wird, dass automatisierte Ergebnisse, die Einzelpersonen betreffen, überprüft und korrigiert werden.

Schließlich sind breitere organisatorische und regulatorische Rahmenbedingungen unerlässlich. Vorurteile in Sprachmodellen spiegeln gesellschaftliche Machtstrukturen wider und können nicht allein durch technische Mittel gelöst werden. Integrative Entwicklungsteams, interdisziplinäre Überprüfungsprozesse und die Zusammenarbeit mit betroffenen Gemeinschaften tragen dazu bei, Schäden zu antizipieren und gerechtere Systeme zu entwickeln. Regulatorische Maßnahmen, einschließlich Anforderungen an algorithmische Folgenabschätzungen und Rechenschaftspflichtstandards, können Anreize für einen verantwortungsvollen Einsatz ausrichten und Rückgriff gewähren, wenn ein Schaden auftritt.

Fazit und Ratschläge/Xulosa va takliflar/Заключение и рекомендации. Diskriminierungsformen in den Trainingsdaten von Sprachmodellen zeigen sich durch erlernte statistische Zusammenhänge, die die Ergebnisse der Modelle auf feine und offensichtliche Weise beeinflussen. Diese Vorurteile äußern sich, indem Stereotypen in der Sprache erzeugt werden, bestimmte Bevölkerungsgruppen ungleich behandelt werden und vorherrschende kulturelle Ansichten verstärkt werden. Für die Nutzer hat das negative Folgen, wie Ausgrenzung und falsche Darstellung, aber auch größere gesellschaftliche Schäden, wie die Normalisierung von Ungleichheit und der Verlust des Vertrauens in digitale Systeme. Um diese Probleme anzugehen, braucht es einen umfassenden Ansatz, der verbesserte

Datenpraktiken, Techniken zur Abschwächung von Vorurteilen auf Modellebene, strenge Bewertungen, Transparenz, menschliche Aufsicht und rechtliche Regelungen kombiniert.

Auch wenn keine Lösung Vorurteile in Systemen, die auf menschlicher Sprache basieren, vollständig beseitigen kann, können langfristige und koordinierte Anstrengungen Schäden erheblich reduzieren und gerechtere und vertrauenswürdigere Sprachtechnologien fördern.

Literaturverzeichnis/Foydalanilgan adabiyotlar ro‘yxati:

[1]. Caliskan A., Bryson J.J., Narayanan A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), p. 183–186.

[2]. Bolukbasi T., Chang, K.-W., Zou J., Saligrama,V., Kalai A.T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Advances in Neural Information Processing Systems.

[3]. Noble S.U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

[4]. Buolamwini J., Gebru T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research.

[5]. 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 ACM Conference on Fairness, Accountability, and Transparency.

[6]. Mitchell M., Wu S., Zaldivar A., et al. (2019). Model cards for model reporting. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.