



UDK: 81'322.2:004.8

Muzaffar T. TADJIBAYEV,

PhD, dotsent, Toshkent Kimyo xalqaro universiteti Namangan filiali, Namangan, O'zbekiston

E-mail: abdugaffor16@gmail.com, ORCID: 0009-0002-2690-2208

Dotsent, PhD D.Kuliyeva taqrizi asosida

PROBLEMS AND LIMITATIONS OF ARTIFICIAL INTELLIGENCE INTEGRATION IN MODERN LINGUISTICS

Annotation

This article analyzes the problems and limitations arising from the integration of artificial intelligence technologies into modern linguistics. Based on recent scholarly studies, the research examines the application of natural language processing, large language models, generative artificial intelligence, and machine translation in linguistic research and practice. Particular attention is paid to the quality and sufficiency of linguistic data, the underrepresentation of low-resource languages in digital environments, the modeling of semantic and pragmatic meaning, multilingualism, and the consideration of linguocultural characteristics.

Key words: Artificial intelligence, modern linguistics, natural language processing, large language models, low-resource languages, machine translation, linguistic data, generative artificial intelligence.

ПРОБЛЕМЫ И ОГРАНИЧЕНИЯ ИНТЕГРАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОВРЕМЕННУЮ ЛИНГВИСТИКУ

Аннотация

В статье анализируются проблемы и ограничения, возникающие в процессе интеграции технологий искусственного интеллекта в современную лингвистику. На основе научных источников рассматриваются возможности и проблемные аспекты применения технологий обработки естественного языка, больших языковых моделей, генеративного искусственного интеллекта и машинного перевода. Особое внимание уделяется вопросам достаточности и качества лингвистических данных, недостаточной представленности языков с ограниченными ресурсами в цифровой среде, моделирования семантического и прагматического значения, многоязычия и учёта лингвокультурных особенностей.

Ключевые слова: Искусственный интеллект, современная лингвистика, обработка естественного языка, большие языковые модели, языки с ограниченными ресурсами, машинный перевод, лингвистические данные, генеративный искусственный интеллект.

ZAMONAVIY TILSHUNOSLIKDA SUN'IY INTELEKT INTEGRATSIYASIDAGI MUAMMOLAR VA CHEKLOVLAR

Annotatsiya

Mazkur maqolada zamonaviy tilshunoslikda sun'iy intellekt texnologiyalarining integratsiyasi jarayonida yuzaga kelayotgan muammolar va cheklovlar tahlil qilinadi. Tadqiqotda tabiiy tilni qayta ishlash, katta til modellari, generativ sun'iy intellekt va mashinaviy tarjima texnologiyalarining lingvistik imkoniyatlari hamda ularni qo'llash jarayonida kuzatiladigan muammolar ilmiy manbalar asosida o'rganiladi. Xususan, lingvistik ma'lumotlarning yetarililigi va sifati, kam resursli tillarning raqamli muhitda yetarlicha ifodalanmasligi, semantik va pragmatik ma'nolarni modellashtirish, ko'p tillilik hamda lingvomadaniy xususiyatlarni hisobga olish masalalariga alohida e'tibor qaratiladi.

Kalit so'zlar: Sun'iy intellekt, zamonaviy tilshunoslik, tabiiy tilni qayta ishlash, katta til modellari, kam resursli tillar, mashinaviy tarjima, lingvistik ma'lumotlar, generativ sun'iy intellekt.

Kirish. XXI asrda sun'iy intellekt (SI) texnologiyalarining jadal rivojlanishi tilshunoslikning nazariy va amaliy yo'nalishlariga sezilarli ta'sir ko'rsatmoqda. Tabiiy tilni qayta ishlash (Natural Language Processing – NLP), mashinaviy tarjima, avtomatik nutqni tanish, matn generatsiyasi, korpus tahlili hamda katta til modellari (Large Language Models – LLM)ning rivojlanishi lingvistik ma'lumotlarni yig'ish, qayta ishlash va tahlil qilish jarayonlarini yangi bosqichga olib chiqdi. Biroq sun'iy intellektning tilshunoslikka chuqur integratsiyalashuvi bilan bir qatorda lingvistik, metodologik, texnik va ma'lumotlar bilan bog'liq qator muammolarni ham yuzaga keltirmoqda.

Sun'iy intellekt tizimlarining samaradorligi, avvalo, ularni o'qitishda foydalaniladigan ma'lumotlarning hajmi, sifati va xilmaxilligiga bog'liq. Turli tillar va madaniyatlarning SI korpuslarida teng darajada ifodalanmasligi, ayniqsa, kam resursli tillar uchun jiddiy muammo hisoblanadi. Elektron korpuslar, parallel matnlar, lingvistik bazalar va nutq resurslarining cheklanganligi bunday tillarda SI tizimlarining samaradorligiga ta'sir ko'rsatadi.

Tilning pragmatik va madaniy xususiyatlarini avtomatik modellashtirish ham murakkab masalalardan biridir. Kinoya, metafora, presuppozitsiya, implikatura, kontekstual ma'no va

milliy-madaniy konnotatsiyalarni faqat statistik bog'lanishlar asosida to'liq anglash qiyin.

Mavzuga oid adabiyotlar tahlili. Sun'iy intellekt va tilshunoslikning o'zaro integratsiyasi so'nggi yillarda xalqaro ilmiy tadqiqotlarning muhim yo'nalishlaridan biriga aylandi. Mazkur yo'nalishdagi tadqiqotlarda sun'iy intellektning imkoniyatlari bilan bir qatorda uning tilning mazmuniy, kontekstual va pragmatik xususiyatlarini modellashtirishdagi cheklovlariga ham alohida e'tibor qaratilmoqda. Bender va Ko'ller tilni avtomatik qayta ishlash masalasini tahlil qilar ekan, til shakli bilan uning mazmuni o'rtasidagi farqqa e'tibor qaratadi [2]. Mazkur yondashuv sun'iy intellektning tilshunoslikdagi eng muhim nazariy muammolaridan biri - "tilni qayta ishlash" va "tilni tushunish" o'rtasidagi farqni ko'rsatib beradi.

Bender, Gebru, Makmillan-Major va Shmitchell kabi tadqiqotchilar katta til modellari o'rganib, ularning o'quv ma'lumotlari, model hajmi va natijalarining ijtimoiy hamda lingvistik oqibatlariga e'tibor qaratgan [4]. Katta til modellari imkoniyatlarini tadqiq qilgan Braun va boshqalar GPT-3 modelining turli lingvistik vazifalarni bajarish qobiliyatini ko'rsatgan. Tadqiqot natijalari modelning hech qanday maxsus

qayta o'qitishsiz, ya'ni few-shot learning (kam namunali o'rganish) mexanizmi orqali turli vazifalarni bajarishi mumkinligini ko'rsatadi.

Jurafskiy va Martin tomonidan ishlab chiqilgan "Speech and Language Processing" asarida tabiiy tilni qayta ishlashning zamonaviy nazariyasi va texnologik asoslari keng yoritilgan. Mualliflar zamonaviy NLP tizimlarida neyron modellar va Transformer arxitekturasining markaziy o'rin egallashini ko'rsatadi [1]. Manningning inson tilini tushunish va mantiqiy xulosa chiqarish bilan bog'liq tadqiqotlari ham mazkur masalani chuqurroq anglash imkonini beradi. Uning ishlarida tabiiy tilni tushunish faqat grammatik shakllarni aniqlash bilan cheklanmasligi, balki bilim, kontekst va mantiqiy xulosalash jarayonlarini ham qamrab olishi ta'kidlanadi [6]. Dvivedi va boshqalar generativ sun'iy intellektning turli sohalaridagi ta'sirini kompleks ravishda o'rganib, katta til modellarining ilmiy va amaliy faoliyatga olib kirayotgan o'zgarishlarini tahlil qilgan [7].

Kaznechi va boshqalar katta til modellarining ta'limdagi qo'llanilishini tahlil qilib, ularning til o'rganish, matn yaratish va individual ta'lim jarayonlarini qo'llab-quvvatlash imkoniyatlarini ko'rsatib o'tishgan [8]. Vasvani va boshqalar tomonidan taklif etilgan Transformer arxitekturasini zamonaviy katta til modellarining texnologik asoslaridan birini yaratdi [9].

Tadqiqot metodologiyasi. Mazkur tadqiqotda zamonaviy tilshunoslikda sun'iy intellekt texnologiyalarining integratsiyasi natijasida yuzaga kelayotgan muammolar va cheklovlarini aniqlash hamda tizimli tahlil qilish maqsadida kompleks metodologik, ilmiy adabiyotlarni tahlil qilish va umumlashtirish metodlari orqali sun'iy intellektning tilshunoslikdagi qo'llanilishi, katta til modellarining ishlash xususiyatlari, lingvistik ma'lumotlarni qayta ishlash imkoniyatlari hamda mavjud muammolar haqidagi ilmiy tadqiqotlar o'rganildi. Shuningdek, qiyosiy-tahliliy metod yordamida an'anaviy lingvistik tahlil va sun'iy intellekt asosidagi avtomatlashtirilgan tahlil imkoniyatlari qiyoslandi. Maqolada tizimli yondashuv ham muhim metod sifatida qo'llanildi. Deskriptiv va tasniflash metodlari yordamida sun'iy intellekt texnologiyalarining tilshunoslikda qo'llanilishida kuzatiladigan asosiy cheklovlar guruhlashtirildi.

Tahlil va natijalar. O'rganilgan ilmiy adabiyotlar va zamonaviy tadqiqotlar tahlili shuni ko'rsatadiki, sun'iy intellektning tilshunoslikka integratsiyasi bir qator muhim muammolarni yuzaga keltirmoqda. Ushbu muammolarni shartli ravishda quyidagicha tasniflash mumkin:

Lingvistik ma'lumotlarning yetariligi va sifati. Sun'iy intellekt modellarining ishlash samaradorligi, avvalo, ularni o'qitish uchun foydalaniladigan ma'lumotlarga bog'liq. Katta til modellarini juda katta hajmdagi matnlar asosida o'qitilishda, matn hajmining kattaligi har doim ham ularning lingvistik jihatdan sifatli va reprezentativ ekanini anglatmaydi. Model o'qitiladigan korpusda ma'lum janrlar, mavzular, tillar yoki kommunikativ guruhlarning haddan tashqari ko'p yoki kam ifodalanishi keyingi natijalarga ta'sir ko'rsatishi mumkin.

2024-yilda Nigatu va boshqalar tomonidan amalga oshirilgan tadqiqotda "kam resursli til" tushunchasining o'zi bir nechta o'lchovga ega ekanligi ko'rsatildi. Mualliflar ACL Anthology va nutq texnologiyalariga oid konferensiyalardagi 150 ta tadqiqotni tahlil qilib, tilning kam resursli yoki yuqori resursli ekanini faqat mavjud matnlar hajmi bilan belgilash yetarli emasligini aniqlagan. Ma'lumotlar miqdori, annotatsiya sifati, texnologik resurslar, tilning raqamli muhitdagi mavjudligi va boshqa omillar ham ushbu holatni belgilaydi [10].

Kam resursli tillar muammosi. Sun'iy intellektning zamonaviy tilshunoslikdagi eng muhim muammolaridan biri yuqori resursli va kam resursli tillar o'rtasidagi texnologik tafovutdir. Ingliz, xitoy, ispan, fransuz kabi tillar uchun katta hajmdagi korpuslar, parallel ma'lumotlar, lug'atlar, nutq bazalari va annotatsiyalangan resurslar mavjud. Biroq ko'plab milliy va mintaqaviy tillarda bunday resurslar yetarli darajada shakllanmagan.

Kahyavijaya, Loveniya va Funngning 2024-yildagi tadqiqotida katta til modellarining 25 ta kam resursli va 7 ta nisbatan yuqori resursli tildagi faoliyati o'rganilgan. Tadqiqotchilar few-shot (kam namunali) va cross-lingual learning (tillararo o'rganish) yondashuvlari kam resursli tillarda model natijalarini yaxshilashi mumkinligini ko'rsatgan bo'lsa-da, bunday tillar bo'yicha tadqiqotlar hali ham yuqori resursli tillarga nisbatan ancha kamligini qayd etadi [11]. Bu holat o'zbek tili uchun ham dolzarbdir. O'zbek tilining katta hajmdagi sifatli korpuslari, morfologik va sintaktik annotatsiyalari, parallel korpuslari, ovozli ma'lumotlar bazalari hamda maxsus til modellarini rivojlantirish zarur.

Ko'p tillilik va "multilinguality" muammosi. Sun'iy intellekt tizimlarining ko'p tillilik xususiyati dastlab katta afzallik sifatida qaralsa-da, tadqiqotlar ko'p tilli modellar ham barcha tillarga bir xil darajada xizmat qila olmasligini ko'rsatmoqda. Bir model tarkibiga ko'plab tillarning kiritilishi ayrim tillar uchun ijobiy natija berishi mumkin, biroq modelning cheklangan resurslari sababli boshqa tillarning samaradorligi pasayishi ehtimoli mavjud. Chang va boshqalar 2024-yilda 250 dan ortiq yuqori va kam resursli tillar bo'yicha minglab monolingval va multilingual modellarni o'rganib, ko'p tilli o'qitish ma'lum sharoitlarda kam resursli tillarning natijalarini yaxshilashini, biroq modelga tillar haddan tashqari ko'p qo'shilganda "curse of multilinguality" deb ataluvchi holat yuzaga kelishi mumkinligini aniqlagan [12]. Semantik va pragmatik ma'noni modellashtirish muammosi. Tilning eng murakkab xususiyatlaridan biri uning ma'nosi faqat so'z va gaplarning shakliyl tuzilishidan iborat emasligidir. Inson nutqida ma'no ko'pincha kontekst, vaziyat, kommunikativ maqsad, umumiy bilim, madaniy tajriba va suhbatdoshlar o'rtasidagi munosabat orqali shakllanadi. Shu sababli SI tizimining grammatik jihatdan to'g'ri gap yaratishi uning mazkur gapning barcha pragmatik jihatlarini anglatmaydi. Masalan, kinoya, metafora, yashirin ma'no, presuppozitsiya, implikatura yoki madaniy ishoralarni avtomatik aniqlash faqat statistik korrelyatsiyalar asosida murakkab kechadi. Bender va Ko'llerning tadqiqotida ham shakl va ma'no o'rtasidagi farq markaziy masala sifatida ko'rsatilgan [2]. Shunga ko'ra, zamonaviy LLMlarning yuqori darajadagi matn generatsiyasi imkoniyatlarini "insoniy til tushunchasi" bilan aynanlashtirish metodologik jihatdan to'g'ri emas. Bu muammo ayniqsa diskurs tahlili va stilistikada muhimdir. SI bunday tahlilda yordamchi vosita bo'lishi mumkin, biroq natijani lingvistik talqin qilish tadqiqotchining ilmiy kompetensiyasi aralashuvini talab etadi.

Lingvistik va madaniy nomutanosiblik. Til madaniyatdan ajralgan holda mavjud bo'lmaydi. Shu sababli sun'iy intellektning tilshunoslikdagi samaradorligini baholashda faqat grammatik va statistik aniqlik emas, balki lingvomadaniy moslik ham hisobga olinishi kerak. Naus va boshqalarning 2024-yildagi tadqiqotida 16 ta til modeli arab tilidagi turli vazifalar, jumladan, hikoya yaratish, nomlangan obyektlarni aniqlash va sentiment tahlili bo'yicha tekshirilgan. Tadqiqot natijalari ayrim modellarda G'arb madaniyatiga oid tushunchalarga nisbatan ustunlik va madaniy stereotiplar mavjudligini ko'rsatgan [13]. Mazkur natijalar sun'iy intellektning tilshunoslikka integratsiyasida "lingvistik to'g'rilik" va "madaniy to'g'rilik" tushunchalarini farqlash zarurligini ko'rsatadi. Bu muammo o'zbek tili uchun ham alohida ahamiyatga ega. O'zbek tilidagi murojaat shakllari, qarindoshlik terminlari, hurmat kategoriyasi, maqol va iboralar, milliy realiyalar hamda diniy-madaniy konnotatsiyalarning katta qismi bevosita boshqa tillarga ekvivalent bo'lmashligi mumkin. Shuning uchun o'zbek tilidagi SI tizimlarini baholashda faqat tarjima yoki grammatik aniqlik emas, balki lingvomadaniy adekvatlik ham mezon sifatida qo'llanishi zarur.

Mashinaviy tarjimada yuzaga keladigan muammolar. Sun'iy intellektning rivojlanishi mashinaviy tarjima sifatini sezilarli darajada oshirgan bo'lsa-da, tarjimada hali ham muayyan muammolar mavjud. Ayniqsa, kam resursli tillar, murakkab terminologiya, madaniy birliklar va kontekstga bog'liq iboralarini

tarjima qilishda modelning xatolari kuzatilishi mumkin. Benkireyn va boshqalar 2024-yilda mashinaviy tarjimada "hallucination" hodisasini, ya'ni asl matnda mavjud bo'lmagan yoki unga mos kelmaydigan mazmunning tarjima natijasida paydo bo'lishini tadqiq qilgan. Tadqiqot 16 ta til yo'nalishini qamrab olgan va mavjud aniqlash usullari yuqori resursli tillarda nisbatan yaxshi ishlasa-da, kam resursli tillarda ularning samaradorligi sezilarli darajada pasayishini ko'rsatgan [14]. Bu natija mashinaviy tarjima tizimlarida faqat tarjima sifatini emas, balki xatolarni aniqlash imkoniyatini ham rivojlantirish zarurligini ko'rsatadi.

Sun'iy intellekt tomonidan yaratilgan matnlarning ishonchliligi. Generativ SI bilan bog'liq muhim muammolardan yana biri — model tomonidan ishonchli ko'rinadigan, biroq faktik jihatdan noto'g'ri yoki mavjud bo'lmagan ma'lumotlarning yaratilishidir. Bunday holat odatda "hallucination" (uydirma) yoki "confabulation" (shaklan izchil, mazmunan uydirma) atamali bilan ifodalanadi. Lingvistik tadqiqotlarda bu muammo ayniqsa muhim, chunki SI tomonidan yaratilgan misollar, iqtiboslar, terminlar yoki bibliografik ma'lumotlar tadqiqotchi tomonidan tekshirilmagan ilmiy matnga kiritilsa, tadqiqotning ishonchliligiga zarar yetishi mumkin. 2024-yilda Sui va boshqalar LLM "hallucination"larini alohida tahlil qilib, ularni faqat xato sifatida emas, balki modelning matn yaratish xususiyatlari bilan bog'liq murakkab hodisa sifatida ko'rib chiqdi [15]. Lingvist tadqiqotchi SI tomonidan yaratilgan har qanday ilmiy ma'lumotni mustaqil manbalar asosida tekshirilishi lozim.

Tadqiqot natijalarining takrorlanuvchanligi va metodologik muammo. Sun'iy intellektdan lingvistik tadqiqotlarda foydalanish tadqiqot metodologiyasiga ham yangi talablarni qo'ymoqda. An'anaviy lingvistik tadqiqotda ma'lumotlar bazasi, tahlil mezonlari va tadqiqot bosqichlarini aniq ko'rsatish orqali tajribani takrorlash imkoniyati nisbatan yuqori. Generativ SI vositalaridan foydalanilganda esa natija model versiyasi, prompt (so'rov) shakli, kontekst hajmi va boshqa parametrlar bilan bog'liq

bo'lishi mumkin. Shu sababli SI yordamida olib borilgan lingvistik tadqiqotlarda ishlatilgan model, topshiriq shakli, ma'lumotlar to'plami va tahlil mezonlarini imkon qadar aniq ko'rsatish talab etiladi.

Tadqiqot natijalarining alohida jihati shundaki, yuqorida ko'rsatilgan muammolar o'zbek tili sharoitida yanada murakkab ko'rinishga ega bo'lishi mumkin. O'zbek tili agglutinativ til bo'lib, grammatik ma'nolar ko'pincha qo'shimchalar ketma-ketligi orqali ifodalanadi. Bu holat morfologik tahlil, so'z shakllarini generatsiya qilish va lemmatizatsiya jarayonlariga o'ziga xos talablar qo'yadi.

Xulosa va takliflar. O'tkazilgan tahlil natijasida zamonaviy tilshunoslikda sun'iy intellekt integratsiyasining asosiy muammolari o'zaro bog'liq ekani aniqlandi. Ma'lumotlar yetishmasligi model sifatiga, model sifati esa lingvistik va pragmatik natijalarga ta'sir qiladi. O'z navbatida, kam resursli tillarda ma'lumotlar yetishmasligi ko'p tilli modellar samaradorligini cheklaydi. Madaniy va lingvistik nomutanosibliklar esa model natijalarining faqat texnik emas, balki kommunikativ va ilmiy jihatdan ham baholanishini talab qiladi. Ko'p hollarda muammo modelning arxitekturasidan ko'ra, uning qanday ma'lumotlarda o'qitilgani, qaysi til va domen uchun moslashtirilgani hamda natijalar qanday baholanganiga bog'liq.

Shunday qilib, zamonaviy tilshunoslikda sun'iy intellekt integratsiyasining asosiy cheklovi texnologiyaning mavjud emasligi emas, balki inson tilining murakkabligi, tillar o'rtasidagi resurslar nomutanosibligi, lingvomadaniy xilma-xillik va SI tomonidan yaratilgan natijalarni ilmiy jihatdan to'g'ri baholash zarurati bilan belgilanadi. Ayniqsa, o'zbek tili kabi kam resursli tillar uchun milliy korpuslarni kengaytirish, sifatli lingvistik ma'lumotlar bazalarini yaratish, morfologik va sintaktik annotatsiyani takomillashtirish hamda o'zbek tiliga mos SI modellarini ishlab chiqish mazkur muammolarni kamaytirishning muhim shartlaridan biridir.

ADABIYOTLAR

1. Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed., draft). Pearson.
2. 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*, 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
3. OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., et al. (2023). GPT-4 Technical Report. arXiv:2303.08774. (Texnik hisobot; sahifa oraliq'i bibliografik nashr sahifalaridek ko'rsatilmaydi.)
4. 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. <https://doi.org/10.1145/3442188.3445922>
5. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
6. Manning, C. D. (2022). Human language understanding & reasoning. *Daedalus*, 151(2), 127–138. https://doi.org/10.1162/daed_a_01905
7. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
8. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
10. Nigatu, H. H., Tonja, A. L., Rosman, B., Solorio, T., & Choudhury, M. (2024). The Zeno's paradox of "low-resource" languages. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 17753–17774.
11. Cahyawijaya, S., Lovenia, H., & Fung, P. (2024). LLMs are few-shot in-context low-resource language learners. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 405–433. <https://doi.org/10.18653/v1/2024.naacl-long.24>
12. Chang, T. A., Arnett, C., Tu, Z., & Bergen, B. K. (2024). When is multilinguality a curse? Language modeling for 250 high- and low-resource languages. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 4074–4096. <https://doi.org/10.18653/v1/2024.emnlp-main.236>
13. Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024). Having beer after prayer? Measuring cultural bias in large language models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 16366–16393. <https://doi.org/10.18653/v1/2024.acl-long.862>
14. Benkirane, K., Gongas, L., Pelles, S., Fuchs, N., Darmon, J., Stenortorp, P., Adelani, D. I., & Sánchez, E. (2024). Machine translation hallucination detection for low and high resource languages using large language models. *Findings of the Association for Computational Linguistics: EMNLP 2024*, 9647–9665. <https://doi.org/10.18653/v1/2024.findings-emnlp.564>
15. Sui, P., Duode, E., Wu, S., & So, R. J. (2024). Confabulation: The surprising value of large language model hallucinations. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 14274–14284. <https://doi.org/10.18653/v1/2024.acl-long.770>