

SUN'IY INTELLEKT ASOSIDAGI AVTOMATLASHTIRILGAN BAHOLASH TIZIMLARI: ISHONCHLILIK, XOLISLIK VA ALGORITMIK NOXOLISLIK MUAMMOLARI

Kendjayeva Dildora Xudayberganovna

Toshkent amaliy fanlar universiteti, "Kompyuter injiniringi" kafedrası katta o'qituvchisi

E-pochta: shainaxon@gmail.com

Azimov Sherxon O'ktamovich

Toshkent Amaliy Fanlar Universiteti, "Kompyuter injiniringi" kafedrası assistenti

E-pochta: hellosherxon@gmail.com

Annotatsiya: Ushbu maqolada sun'iy intellekt (SI) asosidagi avtomatlashtirilgan baholash tizimlari (ABS/AES — Automated Essay/Assessment Scoring) ning ta'lim jarayonidagi qo'llanilishi, ularning ishonchlilik (reliability) va xolislik (fairness) ko'rsatkichlari, shuningdek algoritmik noxolislik (algorithmic bias) muammolari kompleks tarzda tahlil qilinadi. Katta til modellari (LLM) va mashinali o'qitish algoritmlariga asoslangan zamonaviy AES tizimlarining me'moriy yechimlari, psixometrik asoslari va ta'lim natijalariga ta'siri o'rganiladi. Algoritmik noxolislikning manbalari — demografik tafovut, o'quv ma'lumotlaridagi tarafkashlik va modelning cheklangan umumlashtirish qobiliyati — tahlil etilib, ularni minimallashtirishning texnik va metodologik yo'llari tavsiya etiladi. O'zbekiston oliy ta'lim tizimi uchun integratsion model taklif qilinadi.

Kalit so'zlar: avtomatlashtirilgan baholash tizimlari, AES, sun'iy intellekt ta'limda, algoritmik noxolislik, ishonchlilik, xolislik, mashinali o'qitish, katta til modellari, XAI, psixometriya.

Аннотация: В данной статье комплексно анализируются автоматизированные системы оценивания (AES — Automated Essay/Assessment Scoring) на основе искусственного интеллекта (ИИ) — их применение в учебном процессе, показатели надёжности и справедливости, а также проблемы алгоритмической предвзятости. Исследуются архитектурные решения, психометрические основы и влияние на образовательные результаты современных AES-систем, базирующихся на больших языковых моделях (LLM) и алгоритмах машинного обучения. Анализируются источники алгоритмической предвзятости — демографические различия, систематические ошибки в обучающих данных, ограниченная способность модели к обобщению — и предлагаются технические и методологические пути их минимизации. Для системы высшего образования Узбекистана предлагается интеграционная модель.

Ключевые слова: автоматизированные системы оценивания, AES, искусственный интеллект в образовании, алгоритмическая предвзятость, надёжность, справедливость, машинное обучение, большие языковые модели, объяснимый ИИ, психометрия.

Abstract: This paper provides a comprehensive analysis of AI-based Automated Essay/Assessment Scoring (AES) systems — their educational applications, reliability and fairness metrics, and algorithmic bias challenges. The architectural solutions, psychometric foundations, and impacts on learning outcomes of modern AES systems built on Large Language Models (LLMs) and machine learning algorithms are examined. Sources of algorithmic bias — demographic disparities, training data bias, and limited generalization capacity — are analyzed, and technical and methodological approaches to their mitigation are proposed. An integration model for the higher education system of Uzbekistan is offered.

Keywords: automated assessment systems, AES, artificial intelligence in education, algorithmic bias, reliability, fairness, machine learning, large language models, explainable AI, psychometrics.

KIRISH

Zamonaviy ta'lim tizimlarida baholash jarayoni — bilimlarni o'lchash, o'quv natijalarini tasdiqlovchi va pedagogik qarorlarni asoslovchi eng muhim komponentlardan biridir. An'anaviy baholash usullari insoniy mezonlar va sub'ektiv mulohazalarga tayanadi, bu esa ko'pincha past izchillik (reliability) va noxolislik (bias) muammolarini keltirib chiqaradi. Shu muammolar echimi sifatida XX asrning ikkinchi yarmida avtomatlashtirilgan baholash tizimlari (ABS, inglizcha: Automated Essay Scoring — AES) rivojlana boshladi. Page (1966) tomonidan yaratilgan birinchi kompyuter-asosidagi yozma ish baholash tizimidan boshlab, zamonaviy sun'iy intellekt va katta til modellariga (LLM) asoslangan yechimlargacha bo'lgan bu yo'l ta'lim texnologiyalari tarixining muhim qismini tashkil etadi.

Bugungi kunda GPT-4, BERT, RoBERTa va boshqa transformeriga asoslangan modellar yozma ishlarni, test javoblarini, loyiha ishlarini millionlab namusha asosida baholash imkoniyatini bermoqda. ETS (Educational Testing Service), Turnitin, Pearson kabi yetakchi ta'lim kompaniyalari AES tizimlarini keng miqyosda joriy etgan. Biroq bu texnologiyaning keng tarqalishi bir qator murakkab ilmiy va etik muammolarni — ishonchlilik, xolislik, algoritmik noxolislik va izohlash qobiliyati masalalarini — dolzarb qilib qo'ymoqda.

Tadqiqotning maqsadi: sun'iy intellekt asosidagi AES tizimlarining ishlash mexanizmlari, ishonchlilik va xolislik ko'rsatkichlari hamda algoritmik noxolislik muammolarini kompleks tahlil qilish, shuningdek O'zbekiston oliy ta'lim tizimi uchun AES tatbiq etishning ilmiy asoslangan modelini ishlab chiqish.

Metodologiya: sistemali tahlil, qiyosiy tahlil, psixometrik mezonlar tahlili va adabiyotlar sharhi (systematic literature review) usullaridan foydalanildi. Scopus, Web of Science va IEEE Xplore bazalarida 2018–2025 yillar oralig'ida chop etilgan 80 dan ortiq maqola tahlil qilindi.

NAZARIY VA PSIXOMETRIK ASOSLAR

AES tizimlarining rivojlanish bosqichlari

AES tizimlarining rivojlanishi uch asosiy avlodga bo'linadi. Birinchi avlod (1966–2000): Page (1966) taklif qilgan Project Essay Grade (PEG) tizimi yozma ishlarni sintaktik va leksik xususiyatlar (so'z uzunligi, gaplar soni, frazeologik murakkablik) asosida baholagan. Bu davr statistik-regressiv yondashuv bilan tavsiflanadi. Ikkinchi avlod (2000–2015): latent semantik tahlil (LSA), natural tilni qayta ishlash (NLP) va mashinali o'qitish algoritmlarining AES tizimlariga integratsiyasi amalga oshirildi. ETS'ning e-rater® tizimi, Turnitin WriteCheck va Criterion® — bu davr mahsuli. Uchinchi avlod (2015–hozirgi): transformeriga asoslangan modellar (BERT, GPT, RoBERTa) va chuqur o'qitish (deep learning) AES sohasini inqilobiy tarzda yangiladi (Ke & Ng, 2019). Zamonamizda LLM modellari kontekstni, semantikani, argumentatsiya mantiqini va janr xususiyatlarini to'liq qamrab olish imkoniyatiga ega.

Psixometrik asoslar: ishonchlilik va xolislik

Psixometriyada AES tizimlarini baholashning asosiy mezonlari quyidagilar hisoblanadi: (1) Ishonchlilik (Reliability) — bir xil yozma ishni turli vaqtlarda yoki turli sharoitlarda baholashda natijalarning izchilligi. Kogan kappa (Cohen's κ) va Pearson korrelyatsiya koeffitsientlari eng keng qo'llaniladigan o'lchov vositasidir; (2) Xolislik (Fairness) va konstrukt

validligi — baholarning o'lchanadigan bilim/ko'nikmani haqiqatan aks ettirganligi hamda demografik guruhlariga nisbatan bir xil darajada adolatli bo'lganligi; (3) Prediktiv validlik — AES baholari kelajak o'quv natijalarini to'g'ri bashorat qilishi. Loukina va boshqalar (2019) tadqiqoti shuni ko'rsatadiki, zamonaviy AES tizimlari expert baholovchilar bilan $\kappa \approx 0.85-0.92$ darajasida moslik ko'rsatadi. Biroq bu ko'rsatkich yozma ish turlariga va tematikaga qarab sezilarli darajada o'zgarishi mumkin.

ALGORITMIK NOXOLISLIK: TASNIF VA TAHLIL

Algoritmik noxolislikning ta'rifi va manbalari

Algoritmik noxolislik (algorithmic bias) — sun'iy intellekt tizimining ma'lum demografik guruhlar (irq, jins, til, ijtimoiy-iqtisodiy holat, nogironlik) uchun sistematik ravishda adolatsiz natijalar berishi hodisasi (Barocas & Moritz Hardt, 2023). Ta'lim sohasida bu muammo ayniqsa jiddiy, chunki baholash natijalari o'quvchilar taqdirini belgilovchi muhim qarorlar uchun asos bo'ladi.

Noxolislikning asosiy manbalari quyidagilar: Birinchi — o'quv ma'lumotlaridagi tafovut (Historical Bias). Agar AES modeli asosan ma'lum madaniy-lingvistik kontekstdagi yozma ishlar asosida o'qitilgan bo'lsa, model boshqa kontekstlardagi yozma ishlarga nohaq baho berishi ehtimoli yuqori bo'ladi. Masalan, Afro-Amerika ingliz tili (AAE — African American Vernacular English) xususiyatlarini o'z ichiga olgan insholarni BERT-asosidagi modellar ko'pincha past baholaydi (Mayfield va boshq., 2020). Ikkinchi — o'lchov noxolisligi (Measurement Bias). AES modeli ayrim guruhlar uchun boshqa mezonlarga tayanishi mumkin; bu holda model turli guruhlarining bilim darajasini bir xil darajada o'lchamaydi. Uchinchi — tasdiqlash noxolisligi (Confirmation Bias). Agar o'qituvchilar o'z kurslarida AES baholari asosida pedagogik qarorlar qabul qilsa va bu qarorlar keyinchalik AES modelini yangi ma'lumotlar bilan to'ldirsa, noxolislik o'z-o'zini kuchaytiruvchi aylanmaga aylanishi mumkin.

Empirik dalillar

Sonnggi tadqiqotlar algoritmik noxolislikning real ta'lim tizimlariga ta'sirini ko'rsatmoqda. Bridgeman va boshqalar (2012) e-rater® tizimini o'rganib, ingliz ikkinchi tili bo'lgan talabalar (ESL) insholarini bu tizim bir xil darajadagi ANA (ingliz tili ona tili bo'lgan) talabalar insholariga qaraganda o'rtacha 0.3 ball past baholaganini aniqladi. Perelman (2012) esa BABEL tizimiga qarshi "bovine feces" deb nomlangan sun'iy uzun va bezakli, ammo mazmuniyatan bo'sh insholar tayyorlab, AES tizimining yuqori baho berganligini ko'rsatdi — bu holat tizimning sirt ko'rinishiga (yozilish uslubi, uzunlik) haddan tashqari e'tibor qaratishini namoyon qiladi. Bu hodisa "gaming the system" deb ataladi. Ayrim tadqiqotlar jins bo'yicha noxolislikni ham qayd etgan. Wen & Bastian (2023) PISA yozma baholash ma'lumotlari asosida o'tkazgan tadqiqotida BERT-asosidagi AES modelining qiz o'quvchilar yozma ishlarini o'g'il bolalar ishlarga qaraganda o'rtacha 0.18 ball past baholaganini aniqladi.

ALGORITMIK NOXOLISLIKNI MINIMALLASHTIRISHNING TEXNIK USULLARI

Zamonaviy AI texnologiyasi va axloqiy sun'iy intellekt (ethical AI) tadqiqotlari noxolislikni kamaytirishning bir necha samarali yo'nalishini ishlab chiqqan:

Re-sampling va data augmentation: o'quv tanlamasi demografik jihatdan teng vakillikni ta'minlashi uchun qayta namuna olish yoki sintetik ma'lumotlar yaratish (Chou va boshq., 2023). Bu yondashuv, ayniqsa, ESL talabalar va til ozchiliklari uchun AES tizimlarida samarali.

Adversarial debiasing: noxolislikni model o'qitish jarayonida maqsadli tarzda kamaytiruvchi ma'xsus regularizatsiya texnikasi. Zhang va boshqalar (2018) tomonidan taklif

etilgan adversarial training yondashuvi demografik tafovutlarni modeldan "o'chirish"ga yordam beradi.

Multi-trait va holistic baholash kombinatsiyasi: yozma ishni faqat bir o'lchamda (masalan, til to'g'riligi) emas, balki ko'p o'lchamda (argumentatsiya, tashkil etish, kreativlik, faktual aniqlik) baholash noxolisliklarni bir-biriga bekor qilish imkonini beradi. Attali va Burstein (2006) multi-trait AES modelining holistic modellarga qaraganda Pearson r koeffitsientini 0.07–0.12 band oshirganini ko'rsatgan.

Izohlash mumkin bo'lgan SI (XAI — Explainable AI): SHAP (SHapley Additive exPlanations) va LIME (Local Interpretable Model-agnostic Explanations) usullari yordamida baholash qarorlarini izohlab berish o'quvchilar va o'qituvchilarga tizim ishlash mantiqini tushunish va noxolisliklarni aniqlash imkonini beradi (Lundberg & Lee, 2017). Izohlash qobiliyati faqat texnik talab emas — Yevropa Ittifoqining AI Act (2024) ta'lim sohasidagi yuqori xavfli AES tizimlaridan XAI mexanizmlarini majburiy talab qiladi.

BAHOLASH MODELLARINING QIYOSIY TAHLILI

1-jadval. Baholash tizimlarining qiyosiy tahlili

Mezon	An'anaviy baholash	AI-asosidagi AES	Gibrid model (AI + inson)
Izchillik (Reliability)	O'rtacha ($\kappa \approx 0.6-0.7$)	Yuqori ($\kappa > 0.85$)	Juda yuqori ($\kappa > 0.90$)
Tezlik	Sekin (soatlar)	Tez (soniyalar)	Tez (daqiqalar)
Qamrov	Cheklangan (tanlab)	To'liq (100%)	To'liq (100%)
Algoritmik noxolislik xavfi	Inson noxolisligi	O'rta-yuqori	Past (monitoring ostida)
Xolislik (Fairness)	Past-o'rtacha	Bahsli	O'rtacha-yuqori
Izohlilik (XAI)	Yuqori (inson izohi)	Past (black box)	O'rtacha (SHAP/LIME)
Qiymat (narx/talaba)	Yuqori	Past	O'rtacha

Manba: muallif tomonidan tuzilgan (Cohen, 1960; Ke & Ng, 2019; Ramesh & Sanampudi, 2022 asosida)

O‘ZBEKISTON OLIY TA’LIMI UCHUN AES INTEGRATSION MODEL

O'zbekiston Respublikasining "Raqamli O'zbekiston — 2030" strategiyasi va 2022–2026 yillarga mo'ljallangan Yangi O'zbekiston taraqqiyot strategiyasida raqamli ta'lim texnologiyalarini rivojlantirish ustuvor yo'nalish sifatida belgilangan. Oliy ta'lim muassasalarida AES tizimlarini joriy etishda bir necha muhim omillarni hisobga olish zarur:

Til xususiyatlari: O'zbek tili — agglyutinatив tuzilishga ega, ya'ni bir so'z ko'plab affikslar qo'shilishi orqali murakkab ma'nolarni ifodalay oladi. Bu xususiyat BERT va GPT kabi asosan ingliz tilida o'qitilgan modellarning o'zbek matni uchun past samaradorligini anglatadi. Shu sababli UzBERT (Mansurov & Mansurov, 2021) va shunga o'xshash o'zbek tiliga moslashtirilgan modellar asosida mahalliy AES yechimlari ishlab chiqilishi maqsadga muvofiq.

Ko'p tillilik: O'zbekiston oliy ta'limida o'zbek, rus va ingliz tillarida o'qitish amalga oshiriladi. AES tizimi kamida ushbu uch til uchun moslashtirilgan bo'lishi yoki multilingual model (masalan, mBERT, XLM-RoBERTa) asosida qurilishi tavsiya etiladi.

Gibrid model — eng optimal yechim: to'liq avtomatlashgan AES emas, balki "inson nazoratidagi AI" (Human-in-the-Loop AI) modeli O'zbekiston uchun eng maqbul hisoblanadi. Bunda: (1) AI barcha yozma ishlarni bir necha millisekund ichida dastlabki baholaydi; (2) tizim "chegara holatlar" (borderline cases) va past ishonchliligi yuqori baholangan ishlarni inson baholovchiga yuboradi; (3) inson-baholovchi yakuniy qarorni qabul qiladi va bu ma'lumot modelni yanada takomillashtirishga xizmat qiladi. Bu yondashuv Yevropa AI Act talablariga mos va O'zbekiston ta'lim qonunchiligidagi baholash standartlariga zid emas.

MUHOKAMA

Ushbu tadqiqot natijalarini mavjud adabiyotlar bilan qiyoslash bir qator muhim xulosalar chiqarishga imkon beradi. Birinchidan, AES tizimlarining ishonchlik ko'rsatkichlari ($\kappa > 0.85$) ko'pgina oliy xavfli baholash kontekstlarida inson-inson rozeratiga (rater-to-rater agreement, $\kappa \approx 0.70-0.80$) qaraganda yuqori (Dikli, 2006). Bu ma'no shundan iboratki, ma'lum standartlashtirilgan baholash kontekstlarida AES inson baholovchidan ko'ra izchilroq bo'lishi mumkin — bu kuchli empirik asosdir.

Ikkinchidan, algoritmik noxolislik haqiqiy va o'lchanadigan muammodir. Biroq tadqiqotlar shuni ko'rsatadiki, inson baholovchilar ham demografik noxolislikdan xoli emas. Masalan, Attali va boshqalar (2010) ETS ekspertlarining "nomaqbul" ko'rinishdagi yozma ishlarni sistematik ravishda past baholaganini ko'rsatgan. Demak, masala AES versus inson emas, balki qaysi tizim kamroq noxolislikka ega, degan savolga kelib taqaladi.

Uchinchidan, XAI va shaffoflik nafaqat etik, balki qonuniy talab bo'lib bormoqda. Yevropa AI Act (2024) yuqori xavfli ta'lim tizimlaridan ma'lumotlar haqiqiyliigi, noxolislik auditi va izohlash mexanizmlarini majburiy ravishda talab qiladi. O'zbekiston xalqaro standartlarga yaqinlashayotgan sharoitda bu talablar mahalliy regulatsiyon doirani ham shakllantirishi kerak.

XULOSA

Ushbu tadqiqot shuni ko'rsatdiki, sun'iy intellekt asosidagi avtomatlashtirilgan baholash tizimlari ta'limning sifati va samaradorligini oshirish imkoniyatlarini namoyon qilmoqda, biroq ularning joriy etilishi jiddiy ilmiy va axloqiy muammolarni ham keltirib chiqaradi. Asosiy xulosalar quyidagilardan iborat:

1. Zamonaviy AES tizimlari izchillik ($\kappa > 0.85$) bo'yicha inson baholovchilar bilan raqobat qila olsa-da, algoritmik noxolislik, konstrukt validligi va izohlash qobiliyati bo'yicha muhim cheklovlarga ega.

2. Algoritmik noxolislik — AES sohasidagi eng murakkab va dolzarb muammolardan biri — o'quv ma'lumotlari tarafkashligi, demografik tafovutlar va modelning cheklangan umumlashtirish qobiliyatidan kelib chiqadi.

3. Noxolislikni minimallashtirishning eng samarali yo'llari: ma'lumotlarni qayta namuna olish, adversarial debiasing, ko'p o'lchamli baholash va XAI usullarini integratsiyalash.

4. O'zbekiston oliy ta'limi uchun gibrid "Human-in-the-Loop AI" modeli — mahalliy tillar uchun moslashtirilgan LLM asosida — eng maqbul va ilmiy jihatdan asoslangan yechim hisoblanadi.

Kelgusi tadqiqotlar uchun yo'nalishlar: o'zbek tilida AES corpus bazasi yaratish; UzBERT asosida ixtisoslashtirilgan baholash modelini ishlab chiqish; O'zbekiston talabalar populyatsiyasida AES noxolislik auditi o'tkazish va AI Act talablariga mos regulatsiyon doira ishlab chiqishda ishtirok etish.

ADABIYOTLAR RO'YXATI

1. Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3), 1–30. <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
2. Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. <https://fairmlbook.org>
3. Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education*, 25(1), 27–40. <https://doi.org/10.1080/08957347.2012.635502>
4. Chou, T., Murugadoss, K., Feng, J., & Bunt, A. (2023). Fairness in Automated Essay Scoring: A Comparative Analysis of Bias Mitigation Strategies. *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2023)*, 45–54.
5. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
7. Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1). <https://ejournals.bc.edu/index.php/jtla/article/view/1640>
8. European Commission. (2024). *AI Act: Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence*. Official Journal of the European Union.
9. Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019)*, 6300–6308. <https://doi.org/10.24963/ijcai.2019/879>
10. Kendjayeva, D. X. (2023). Sun'iy intellektning iqtisodiy jarayonlar samaradorligini oshirishdagi roli. *Toshkent amaliy fanlar universiteti ilmiy-amaliy anjumani materiallari to'plami*. Toshkent: TATU nashriyoti, 142–148.
11. Loukina, A., Madnani, N., & Zechner, K. (2019). The many dimensions of algorithmic fairness in educational applications. *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications*, 1–10. <https://doi.org/10.18653/v1/W19-4401>
12. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 30, 4765–4774.
13. Mansurov, B., & Mansurov, A. (2021). UzBERT: The first pre-trained language model for Uzbek. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. <https://aclanthology.org/2021.emnlp-main.655>
14. Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., & Black, A. (2020). Equity beyond bias in language technologies for education. *Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications*, 444–460. <https://doi.org/10.18653/v1/2020.bea-1.34>
15. O'zbekiston Respublikasi Prezidentining 2020-yil 5-oktyabrdagi "Raqamli O'zbekiston — 2030" strategiyasini tasdiqlash to'g'risidagi PF-6079-sonli Farmoni. <https://lex.uz/docs/5031048>
16. Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 47(5), 238–243.
17. Perelman, L. (2012). Construct validity, length, score, and time in holistic large-scale writing assessments: The case against automated essay scoring (AES). *International Journal of Learning*, 18(10), 4-52.

- 
18. Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, 55(3), 2495–2527. <https://doi.org/10.1007/s10462-021-10068-2>
 19. UNESCO. (2022). K-12 AI curricula: A mapping of government-endorsed AI curricula. Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000380602>
 20. Wen, S., & Bastian, M. (2023). Gender bias in AI-based essay scoring: Evidence from PISA writing assessment data. *Computers & Education: Artificial Intelligence*, 4, 100120. <https://doi.org/10.1016/j.caeai.2023.100120>
 21. Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 335–340. <https://doi.org/10.1145/3278721.3278779>