

Responsible AI in Assessment: A Bibliography

This bibliography is a centralized resource of selected works on the responsible use of AI in testing, including domain-agnostic and domain-specific publications on ethics, governance, validity, and fairness. The Duolingo English Test research team hopes it will be useful to researchers and practitioners in educational measurement and language assessment. Entries are grouped by theme and listed newest first. Initial references came from Duolingo English Test publications on this topic and were supplemented with additional works we thought would be useful. We expect the bibliography to grow over time, within the current categories and new ones. We welcome suggestions for revisions and updates from the community.

This first version was prepared by Jill Burstein in October 2026. Many thanks to Alina von Davier for her review and helpful suggestions. If you would like to suggest changes, please contact: jill@duolingo.com.

Declaration of Generative AI Use

This bibliography was compiled with the help of Claude Opus 5.5 (Anthropic), which extracted references from Duolingo English Test papers,formatted them, and wrote one-line summaries of the publication content. I reviewed and edited the final list.

Contents

Responsible AI programs in practice
Assessment standards and professional guidance
Ethical principles for AI
AI governance, risk management, and regulation
Auditing, accountability, and human oversight
Fairness, bias, and equity
AI in education



Responsible AI programs in practice

Standards, transparency reporting, and case studies from an operational high-stakes test, with the technical documentation behind them.

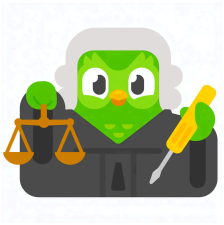
1. Burstein, J., von Davier, A. A., & Shaked, N. (in press). Custodianship of AI in education: Foundations for sustainable and responsible practice. In A. A. von Davier & D. Yan (Eds.), *Artificial intelligence in educational learning and assessment*. Routledge.
   Frames responsible AI in education as custodianship: an ongoing practice of care, guidance, accountability, and reflection, with shared authority and diverse perspectives in AI design and use.
2. Burstein, J., LaFlair, G. T., & von Davier, A. A. (2026, July 1). *A theoretical assessment ecosystem for a digital-first assessment—The Duolingo English Test* (Duolingo Research Report No. DRR-26-02). Duolingo. https://go.duolingo.com/ecosystem
   Integrates four frameworks (language assessment design, measurement, test security, and validity argument) with test-taker experience, all guided by the DET Responsible AI Standards, and maps digital and AI considerations onto each validity inference. This 2026 version replaces all earlier versions.
3. Burstein, J. (2026). *The Duolingo English Test Responsible AI Standards* (Duolingo Research Report No. DRR-26-02). Duolingo. https://go.duolingo.com/ResponsibleAI
   Four principles (validity and reliability; fairness; privacy and security; accountability and transparency) with concrete goals and practices. This 2026 version replaces all earlier versions.



4. Burstein, J. (2026). *The Duolingo English Test Responsible AI Transparency Report* (Duolingo Research Report No. DRR-26-08). Duolingo. https://go.duolingo.com/DET_RAI_Transparency_Report
   Inaugural RAI transparency report for an assessment, documenting practice from Summer 2024 to Summer 2026 for test takers, institutions, and researchers.
5. Burstein, J., & LaFlair, G. T. (2025). Where assessment validation and responsible AI meet. *Language Teaching Research Quarterly, 50*, 120–137. https://doi.org/10.32038/ltrq.2025.50.09
   Maps RAI practices onto each inference in an argument-based validity framework, treating RAI evidence as validity evidence.
6. Burstein, J., LaFlair, G. T., Yancey, K., von Davier, A. A., & Dotan, R. (2025). Responsible artificial intelligence for test equity and quality: The Duolingo English Test as a case study. In E. M. Tucker, E. Armour-Thomas, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume I: Foundations for assessment in the service of learning*. University of Massachusetts Amherst Libraries. https://doi.org/10.7275/2h95-jf35
   Case study of how a high-stakes test operationalizes RAI principles across item generation, scoring, security, and test-taker experience.
7. Johnson, M. S. (2025). *Responsible AI for measurement and learning: Principles and practices* (Research Report No. RR-25-03). ETS. https://doi.org/10.64634/4ftg8n64
   ETS framework for responsible AI in measurement and learning: fairness and bias mitigation, privacy and security, transparency, explainability, accountability, educational impact and integrity, and continuous improvement.
8. Galaczi, E., & Pastorino-Campos, C. (2025). Ethical AI for language assessment: Principles, considerations, and emerging tensions. *Annual Review of Applied Linguistics*, *45*, 294–314. https://doi.org/10.1017/S0267190525100081
   Proposes shared principles and an “ethical-by-design” approach for AI in high-stakes language assessment.



9. Johnson, M. S. (2024). How do we demonstrate AI responsibility: The devil is in the details. *Journal of Educational and Behavioral Statistics, 49*(5), 723–729. https://doi.org/10.3102/10769986241257963
   Commentary on the DET Responsible AI Standards, proposing specific statistical ways to evaluate AI-based automated scoring for fairness and bias.
10. Briggs, D. C. (2024). Strive for measurement, set new standards, and try not to be evil. *Journal of Educational and Behavioral Statistics, 49*(5), 694–701. https://doi.org/10.3102/10769986241238479
   Argues that AI standards for testing should complement the *Standards for Educational and Psychological Testing*, focus on fairness and social responsibility, and acknowledge that marketplace competition may limit transparency.
11. Chang, H.-H. (2024). Harnessing AI for educational measurement: Standards and emerging frontiers. *Journal of Educational and Behavioral Statistics, 49*(5), 702–708. https://doi.org/10.3102/10769986241264033
   Contends that long-standing psychometric standards, such as sample invariance and avoiding item bias, should guide AI-based assessment, with a focus on AI for item writing by instructors.
12. Dixon-Román, E. (2024). AI and psychometrics: Epistemology, process, and politics. *Journal of Educational and Behavioral Statistics, 49*(5), 709–714. https://doi.org/10.3102/10769986241280623
   Responds to the DET Responsible AI Standards by contrasting psychometrics with cybernetic theories of AI and raising ethical and political questions about "human-in-the-loop" approaches.
13. Ho, A. D. (2024). Artificial intelligence and educational measurement: Opportunities and threats. *Journal of Educational and Behavioral Statistics, 49*(5), 715–722. https://doi.org/10.3102/10769986241248771
   Reviews the opportunities and threats of AI for item generation, scoring, security, and reporting, and calls for standards that address the context of test use and ongoing monitoring of bias and scale drift.



Assessment standards and professional guidance

Testing-industry and professional-association guidance on technology, AI, and ethics in assessment.

1. Naismith, B., von Davier, A. A., Burstein, J., & LaFlair, G. T. (Eds.). (in press). *Routledge handbook of digital language assessment: Innovations and insights from the Duolingo English Test*. Routledge.
   Edited volume covering the design, AI methods, security, and validation of a digital-first language test.
2. Evanini, K., & Burstein, J. (in press). Ethical considerations for AI in language learning and assessment. *Journal of Applied Linguistics* [Special issue: AI and applications in applied linguistics: Interdisciplinary perspectives].
   Ethical issues specific to AI in language learning and assessment.
3. Association of Test Publishers. (2026, August). *Assessment in 2040: Voices from the field* [Special publication]. Responsible AI in Assessment Subcommittee, ATP AI and Technology in Assessment Committee. https://www.testpublishers.org/publication-resources
   Short perspectives from assessment practitioners (and a few LLMs) on how AI will change assessment over the next 10 to 15 years, including oversight, validity, privacy, and governance.



4. Association of Test Publishers. (2026, June). *Human oversight of AI in assessment: Considerations for responsible use* (Version 1.1) [Special publication]. Responsible AI in Assessment Subcommittee, ATP AI and Technology in Assessment Committee. https://www.testpublishers.org/publication-resources
   Defines human oversight for AI in item generation, essay scoring, and proctoring; offers a framework of oversight approaches and governance considerations, with anonymized case studies added in v1.1. First released January 2026.
5. International Language Testing Association. (2026, March). *ILTA code of ethics*. https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/code_of_ethics/code_of_ethics_260306.pdf
   The language testing profession's ethics code, approved March 2026.
6. Association of Test Publishers. (2025). *AI laws, regulations and governance frameworks*. https://www.testpublishers.org/publication-resources
   ATP overview of AI laws and governance frameworks relevant to testing organizations.
7. Association of Test Publishers. (2025, May). *Considerations for the use of generative AI in test development*. https://www.testpublishers.org/publication-resources
   Key considerations and risks for GenAI in test development, including human review at critical error points.
8. Association of Test Publishers. (2025). *A primer on generative AI in assessment*. https://www.testpublishers.org/publication-resources
   Introductory guide to generative AI concepts and uses in assessment.
9. Association of Test Publishers. (2025). *AI in assessment glossary*. https://www.testpublishers.org/publication-resources
   Shared terminology for AI in assessment from the ATP Responsible AI in Assessment Subcommittee.

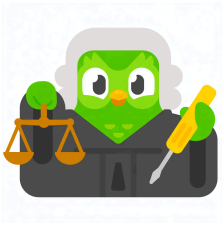


10. International Test Commission & Association of Test Publishers. (2025). *Guidelines for technology-based assessment*. https://www.intestcom.org/page/16
    Joint ITC/ATP guidelines covering design, delivery, scoring, security, and AI in technology-based testing.
11. National Commission for Certifying Agencies. (2025). *NCCA guidance document: Use of artificial intelligence in certification programs* (Version 1.1). Institute for Credentialing Excellence. https://www.credentialingexcellence.org/Portals/0/NCCA%20AI%20Guidance%20Document_v1_1%20FINAL.pdf
    States that AI cannot independently decide certification status and requires human oversight and audit of AI-driven scoring and proctoring.
12. von Davier, A. A., & Burstein, J. (2024). AI in the assessment ecosystem: A human–centered AI perspective. In *Artificial intelligence in education: The intersection of technology and pedagogy* (pp. 93-109). Cham: Springer Nature Switzerland.
13. Association of Test Publishers. (2024). *Creating responsible and ethical AI policies for assessment organizations* [White paper]. https://www.testpublishers.org/white-papers
    Practical guidance for testing organizations writing their own AI policies.
14. Society for Industrial and Organizational Psychology. (2023). *Considerations and recommendations for the validation and use of AI-based assessments for employee selection*. https://www.siop.org/post/siop-releases-recommendations-for-ai-based-assessments/
    Validation, bias review, and psychometric practice for AI-based hiring assessments.
15. International Privacy Subcommittee of the ATP Security Committee. (2021, July 6). *Artificial intelligence and the testing industry: A primer*. Association of Test Publishers.



https://www.testpublishers.org/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_6July2021_Final%20R1%20.pdf
Introductory primer on AI uses, risks, and privacy considerations for testing organizations.

16. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
The foundational standards for validity, fairness, and test use.
17. Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2–13. https://doi.org/10.1111/j.1745-3992.2011.00223.x
Widely used framework of criteria for evaluating automated scoring before and during operational use.
18. Shermis, M. D., & Burstein, J. C. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Routledge.
First edited volume on automated essay scoring that described the inner-workings of AES systems..
19. Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research, 26*(4), 407–425. https://doi.org/10.2190/CX92-7WKV-N7WC-JL0A
Compares how automated and human essay scores relate to external measures of writing ability.



20. Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103–134. https://doi.org/10.1016/S0747-5632(01)00052-8
Invites experts to fool a scoring engine, revealing vulnerabilities to construct-irrelevant strategies.

Ethical principles for AI

Domain-agnostic syntheses of AI ethics principles and human-centered design that assessment-specific standards can leverage.

1. Karamolegkou, A., Borah, A., Cho, E., Ray Choudhury, S., Galletti, M., Gupta, P., Ignat, O., Kargupta, P., Kotonya, N., Lamba, H., Lee, S.-J., Mangla, A., Mondal, I., Moudakir, F. Z., Nazarova, D., Nemkova, P., Pisarevskaya, D., Rizwan, N., Sabri, N., . . . Dementieva, D. (2026). *NLP for social good: A survey and outlook of challenges, opportunities, and responsible deployment* [Preprint]. arXiv. https://arxiv.org/abs/2505.22327
Survey of NLP applied to social good, with attention to responsible deployment.
2. Dignum, V. (2025, September 5). *From responsible AI to custodianship of AI* [LinkedIn article]. https://www.linkedin.com/pulse/from-responsible-ai-custodianship-virginia-dignum-gqmge/
Proposes custodianship: responsibility for AI as ongoing care with shared authority across diverse perspectives.



3. Mihalcea, R., Ignat, O., Bai, L., Borah, A., Chiruzzo, L., Jin, Z., Kwizera, C., Nwatu, J., Poria, S., & Solorio, T. (2025). Why AI is WEIRD and shouldn't be this way: Towards AI for everyone, with everyone, by everyone. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(27), 28657–28670. https://ojs.aaai.org/index.php/AAAI/article/view/35092
   Argues that AI reflects a narrow cultural slice of users and developers, and calls for broader participation.
4. OECD. (2024). *Recommendation of the Council on artificial intelligence* (OECD/LEGAL/0449; OECD AI Principles, adopted 2019, amended 2024). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
   The first intergovernmental AI principles, including human agency and oversight.
5. World Economic Forum. (2024). *AI value alignment: Guiding artificial intelligence towards shared human goals* [White paper]. https://www3.weforum.org/docs/WEF_AI_Value_Alignment_2024.pdf
   Discusses how to align AI systems with human values in practice, including oversight and auditing.
6. Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence, 5*. https://doi.org/10.1016/j.caeai.2023.100152
   Systematic review of how FATE concepts are treated in higher-education AI research.
7. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). John Wiley & Sons. https://doi.org/10.1002/9781119815075.ch45
   Proposes beneficence, non-maleficence, autonomy, justice, and explicability as a unifying set of AI principles.



8. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623).
   Influential critique of large language models covering bias, documentation, and environmental costs.
9. Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature, 595*, 197–204. https://doi.org/10.1038/s41586-021-03666-1
   Argues that measurement in social science must account for how algorithms shape the behavior being measured.
10. Fiesler, C., & Garrett, N. (2020, September 16). Ethical tech starts with addressing ethical debt. *Wired*. https://www.wired.com/story/opinion-ethical-tech-starts-with-addressing-ethical-debt/
    Introduces "ethical debt": harms that accumulate when ethical review is deferred during development.
11. Auernhammer, J. (2020). Human-centered AI: The role of human-centered design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), *Synergy—DRS International Conference 2020* (pp. 1315–1333). https://doi.org/10.21606/drs.2020.282
    Argues for human-centered design methods as a basis for developing AI.
12. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* [Research report]. Berkman Klein Center for Internet & Society at Harvard University. https://doi.org/10.2139/ssrn.3518482
    Maps 36 AI principles documents and identifies eight recurring themes, including fairness, accountability, and transparency.



13. Wang, G. (2019, October 20). *Humans in the loop: The design of interactive AI systems*. Stanford HAI. https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems
Makes the case for designing AI systems around human involvement rather than full automation.

14. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
Reviews 84 AI ethics guidelines; finds convergence on five principles but divergence on how to implement them.

15. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems* (Version 2). IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
Engineering-community framework for embedding human well-being in AI system design.

16. Anderson, R. E., Johnson, D. G., Gotterbarn, D., & Perrolle, J. (1993). Using the new ACM code of ethics in decision making. *Communications of the ACM, 36*(2), 98–107.
Early professional-ethics code for computing, with worked cases on applying it to decisions.

AI governance, risk management, and regulation

Frameworks, regulation, and policy that set expectations for AI systems regardless of domain.

1. Lee, H.-P., He, J., Piorkowski, D., von Davier, T. S., Forlizzi, J., & Das, S. (2026). *The perils of agency: How developers perceive, prioritize, and address risks in*

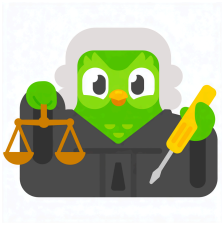


agentic AI products (arXiv:2606.15485). arXiv. https://doi.org/10.48550/arXiv.2606.15485

2. United Nations. (2024, September). *Governing AI for humanity* [Final report]. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
   UN advisory body's recommendations for international AI governance.
3. Dotan, R., Blili-Hamelin, B., Madhavan, R., Matthews, J., Scarpino, J., & Anderson, C. (2024). *A flexible maturity model for AI governance based on the NIST AI risk management framework* [Technical report]. IEEE. https://doi.org/10.48550/arXiv.2401.15229
   Turns the NIST AI RMF into a maturity model organizations can use to assess their own governance practices.
4. European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
   The enacted EU AI Act, which classes many education and assessment uses of AI as high-risk and requires human oversight.
5. PricewaterhouseCoopers. (2024). *PwC's 2024 US responsible AI survey*. https://www.pwc.com/us/en/tech-effect/ai-analytics/responsible-ai-survey.html
   Industry survey on how US companies are implementing responsible AI practices.
6. Council of the European Union. (2023, December 9). *Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world* [Press release]. https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/
   Announces the EU AI Act's risk-based approach, under which AI used in education and assessment is classed as high-risk.



7. International Organization for Standardization & International Electrotechnical Commission. (2023). *Information technology—Artificial intelligence—Management system* (ISO/IEC Standard No. 42001:2023). https://www.iso.org/standard/42001
   Certifiable management-system standard for organizations that develop or use AI.
8. National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1
   Voluntary framework organized around Govern, Map, Measure, and Manage.
9. OECD. (2023). *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI* (OECD Digital Economy Papers No. 349). OECD Publishing. https://doi.org/10.1787/2448f04b-en
   Lifecycle approach to AI accountability linking OECD AI Principles to risk-management steps.
10. Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of responsible AI: From ethical guidelines to cooperative policies. *Frontiers in Computer Science, 4*. https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2022.873437/full
    Examines the gap between AI ethics guidelines and governance, arguing for cooperative policy approaches.
11. International Organization for Standardization & International Electrotechnical Commission. (2022). *Information technology—Artificial intelligence—Artificial intelligence concepts and terminology* (ISO/IEC Standard No. 22989:2022). https://www.iso.org/standard/74296.html
    International reference vocabulary for AI concepts.



12. European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. https://eur-lex.europa.eu/eli/reg/2016/679/oj
EU data protection law, including provisions on automated decision-making relevant to test-taker data.

Auditing, accountability, and human oversight

Methods for auditing AI systems, making their behavior open to scrutiny, and keeping humans meaningfully in charge of AI decisions.

1. von Davier, A. A. (in press). From item generation to agentic measurement systems: A conceptual framework for the validity of AI-driven assessments. In *Proceedings of the LLM Psychometrics Workshop at the 16th International Conference on Learning Analytics and Knowledge (LAK 2026)*.
2. Boston Consulting Group. (2025). *You won't get GenAI right if you get human oversight wrong*. https://www.bcg.com/publications/2025/wont-get-gen-ai-right-if-human-oversight-wrong
Argues that oversight works when reviewers have clear guidelines and incentives, and warns that efficiency pressure leads to cursory review.
3. Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885
Literature review of responsible AI governance, including human oversight, with a framework for future research.



4. Archer, W., & Thomas, T. (2025, June 11). The carbon cost of your university's English testing policy—and how to cut it by 98%. *Duolingo English Test Blog*. https://blog.englishtest.duolingo.com/the-carbon-cost-of-your-universitys-english-testing-policy-and-how-to-cut-it-by-98/
5. Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., & Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3630106.3659051
   Synthesizes psychological, legal, philosophical, and technical views on what makes human oversight of AI effective.
6. Mökander, J., & Axente, M. (2023). Ethics-based auditing of automated decision-making systems: Intervention points and policy implications. *AI & Society, 38*, 153–171. https://doi.org/10.1007/s00146-021-01286-x
   Identifies where in an organization's processes ethics-based audits can intervene, and policy implications.
7. Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines, 31*(2), 323–327. https://doi.org/10.1007/s11023-021-09557-8
   Introduces ethics-based auditing as a governance mechanism for trustworthy AI.
8. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873
   Proposes SMACTR, an end-to-end framework for internal algorithmic audits across the development lifecycle.
9. Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products.



Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435. https://dl.acm.org/doi/10.1145/3306618.3314244
Shows that public disclosure of audit results prompted vendors to reduce bias, supporting the value of transparency.

Fairness, bias, and equity

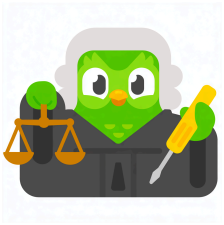
Fairness frameworks for testing and methods for detecting bias in items, raters, and AI systems.

1. Church, J., Park, Y., & Burstein, J. (2025). *Guidelines for fair test content: The Duolingo English Test example* (Duolingo Research Report No. DRR-25-02). Duolingo. https://go.duolingo.com/DETFairnessGuidelines
   Fairness and bias guidelines applied to human- and AI-generated test content.
2. Suk, Y., & Han, K. T. (2024). A psychometric framework for evaluating fairness in algorithmic decision making: Differential algorithmic functioning. *Journal of Educational and Behavioral Statistics, 49*(2), 151–172. https://doi.org/10.3102/10769986231171711
   Adapts DIF logic to algorithmic decisions, defining differential algorithmic functioning.
3. von Davier, A. A., & Burstein, J. (2024). AI in the assessment ecosystem: Implications for fairness, bias, and equity. In P. Ilic, I. Casebourne, & R. Wegerif (Eds.), *Artificial intelligence in education: The intersection of technology and pedagogy* (pp. 93–109). Springer. https://doi.org/10.1007/978-3-031-71232-6_6
   Traces fairness, bias, and equity risks across each part of an AI-enabled assessment ecosystem.
4. Belzak, W. C. M. (2023). The multidimensionality of measurement bias in high-stakes testing: Using machine learning to evaluate complex sources of



differential item functioning. *Educational Measurement: Issues and Practice, 42*(1), 24–33. https://doi.org/10.1111/emip.12486
Uses machine learning to detect DIF across many interacting test-taker characteristics.

5. Belzak, W. C. M., Naismith, B., & Burstein, J. (2023). Ensuring fairness of human- and AI-generated test items. In N. Wang et al. (Eds.), *Artificial intelligence in education: Posters and late breaking results* (pp. 701–707). Springer. https://doi.org/10.1007/978-3-031-36336-8_108
Compares DIF in human-written and AI-generated items on a high-stakes test.
6. Johnson, M. S., & McCaffrey, D. F. (2023). Evaluating fairness of automated scoring in educational measurement. In V. Yaneva & M. von Davier (Eds.), *Advancing natural language processing in educational assessment* (pp. 142–164). Routledge. https://doi.org/10.4324/9781003278658
Methods for checking whether automated scores are fair across test-taker groups.
7. Huggins-Manley, A. C., Booth, B. M., & D'Mello, S. K. (2022). Toward argument-based fairness with an application to AI-enhanced educational assessments. *Journal of Educational Measurement, 59*(3), 362–388. https://onlinelibrary.wiley.com/doi/abs/10.1111/jedm.12334
Extends argument-based validation to fairness claims in AI-enhanced assessment.
8. Johnson, M. S., Liu, X., & McCaffrey, D. F. (2022). Psychometric methods to evaluate measurement and algorithmic bias in automated scoring. *Journal of Educational Measurement, 59*(3), 338–361. https://doi.org/10.1111/jedm.12335
Separates measurement bias from algorithmic bias in automated scores and shows how to test for each.



9. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91.
Landmark audit showing intersectional accuracy gaps in commercial facial analysis; relevant to AI-based proctoring.

AI in education

Guidance and research on responsible AI in education more broadly.

1. von Davier, A. A., & Yan, D. (Eds.). (in press). *Artificial intelligence in educational learning and assessment*. Routledge.
2. Ho, A. (in press). Artificial intelligence and the role of human agency in education, assessment, and measurement. In A. A. von Davier & D. Yan (Eds.), *Artificial intelligence in educational learning and assessment*. Routledge.
3. OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. https://doi.org/10.1787/062a7394-en
International review of how generative AI is being used in education.
4. Khan Academy. (2025, April 3). *Khan Academy's framework for responsible AI in education*. https://blog.khanacademy.org/khan-academys-framework-for-responsible-ai-in-education
An EdTech provider's published principles for responsible AI.



5. Bentley, C., Aicardi, C., Poveda, S., Magela Cunha, L., Kohan Marzagao, D., Glover, R., Rigley, E., Walker, S., Compton, M., & Acar, O. (2023). *A framework for responsible AI education*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4544010
   Framework for teaching responsible AI.
6. Foltýnek, T., Bjelobaba, S., Glendinning, I., Khan, Z. R., Santos, R., Pavletic, P., & Kravjar, J. (2023). ENAI recommendations on the ethical use of artificial intelligence in education. *International Journal for Educational Integrity, 19*, Article 12. https://doi.org/10.1007/s40979-023-00133-4
   European Network for Academic Integrity recommendations on ethical AI use.
7. OECD & Education International. (2023). *Opportunities, guidelines and guardrails for effective and equitable use of AI in education*. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/smart-data-and-digital-technology-in-education/Opportunities.%20guidelines%20and%20guardrails%20for%20effective%20and%20equitable%20use%20of%20AI%20in%20education.pdf
   Joint OECD and teachers' union guidelines and guardrails for AI in education.
8. U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. https://www2.ed.gov/documents/ai-report/ai-report.pdf
   Federal policy report on AI in teaching, learning, and assessment.
9. Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education, 19*(1), 1–11.
   Argues that responsible AI requires education for developers, users, and the public.



10. Institute for Ethical AI in Education. (2021). *The ethical framework for AI in education*. University of Buckingham. https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education

    Framework of objectives and criteria for ethical AI procurement and use in education.

11. Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. UNESCO. https://doi.org/10.54675/PCSP7350

    UNESCO guidance on AI policy in education, with an emphasis on equity and inclusion.

12. Cope, B., Kalantzis, M., & Searsmith, D. (2020). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*. https://doi.org/10.1080/00131857.2020.1728732

    Examines what AI can and cannot do for learning and assessment.

13. Aiken, R. M., & Epstein, R. G. (2000). Ethical guidelines for AI in education: Starting a conversation. *International Journal of Artificial Intelligence in Education, 11*, 163–176.

    One of the first sets of proposed ethical guidelines for AI in education.