Principles for Responsible AI in Medical School and Residency Selection

Artificial intelligence (AI) refers to a broad range of advanced techniques and processes that perform complex tasks, such as large-language models (LLMs), machine learning (ML), and natural language processing (NLP). Historically, simpler statistical methods have been used to analyze application data and predict performance in medical school or training.[1] AI can build upon the existing body of literature and traditional techniques by using more advanced mathematical algorithms or models. This evolution can make AI a powerful tool for identifying patterns and improving decision-making in both undergraduate and graduate medical education selection processes.

The integration of AI into selection processes offers promising advancements in streamlining operations and promoting equity. For example, ML can assist in predicting applicant performance or in prioritizing applications for review. Applications can be screened in a more standardized way by using NLP to simulate expert judgment when evaluating applicant documents such as personal statements or letters of recommendation, to promote fairness and predict valued outcomes.[2] LLMs can be used to improve upon a draft of interview protocols that capture competencies and characteristics important to the institution.

By thoughtfully applying AI, institutions can collectively advance toward more efficient, effective, fair, and informed selection processes.

Ferguson, et al.

Nevertheless, experts in medical school and residency selection are essential when making effective selection decisions. Any use of AI should be balanced with human judgment, insights, and ethical standards. What's more, significant concerns regarding privacy, fairness, transparency, and validity of AI tools remain. It is critical that AI-driven decision-making tools be subjected to the same scrutiny applied to traditional selection methods.

Each institution is unique, with its own mission, goals, and legal context. Therefore, tailoring the application of AI in selection processes to align with the specific needs and values of individual institutions is critical, as is review and approval by legal counsel. Institutions likely already address aspects of each principle in their process. As they consider incorporating AI into their process, they

must think about how to extend their existing best practices to AI-based systems and, in some instances, address entirely new issues related to AI.

As institutions consider what is best for their process, the AAMC recommends six key principles to guide the design and use of AI-based selection systems:

1. **Balance Prediction and Understanding.** Ensure that AI tools deliver insights that improve prediction and efficiency while being comprehensible and usable by the institution, aligning with its objectives and needs.

2. **Protect against Algorithmic Bias.** Rigorously assess and manage biases arising from historical data to ensure fair AI processes and outcomes.

3. **Provide Notice and Explanation.** Maintain transparency by informing applicants how AI is used and how it affects the assessment of their application.

4. **Protect Data Privacy.** Safeguard information with the utmost care, maintaining confidentiality at every step.

5. **Incorporate Human Judgment.** It is crucial to strike the appropriate balance between technology and the irreplaceable value of human judgment and ethical standards.

6. **Monitor and Evaluate.** Assess the outputs and outcomes of the AI system to ensure they remain fair, accurate, and aligned with institutional goals.

Key Principles

Balance Prediction and Understanding

AI holds the promise of measuring aspects of applicant potential at a scale and precision previously unattainable. However, navigating the development and use of AI tools in selection is a delicate balance. And it is vital that those using AI understand the processes behind how the tools work and how to interpret their outcomes. This transparency not only aids in understanding but also contributes to the legitimacy and defensibility of the tools.

As with any tool used in selection, what makes an AI tool valid, fair, and relevant is ensuring the chosen data and metrics reflect the qualities essential for success in each institution. Additionally, the data should meet high data quality standards. Failing to maintain both validity and understandability could pose risks to institutions and applicants.

From principle to practice:

- **Identify characteristics linked to success**. Start by clearly defining the desired characteristics and outcomes the institution intends to measure. Characteristics can be determined using focus groups, surveys, and questionnaires. Collaborate with faculty or assessment and evaluation staff to identify and obtain these data. Design or adapt existing AI tools to capture these qualities, ensuring they effectively target what defines an effective student or trainee.

- **Ensure understandability**. Implement an AI solution that balances accuracy with clarity. Make sure to provide clear explanations for its results and instructions for how to appropriately use that information in selection decisions. This will aid in effective communication with the team that makes the selection decisions, as well as with applicants, leadership, and regulatory bodies.

- **Maintain simplicity**. Use an AI tool that avoids unnecessary complexity in the number of variables and analytic techniques. This will make it easier for the team to understand and use the tool. It will also make it easier to explain to applicants, leadership, and regulatory bodies.

- **Establish and monitor standards for interpretability**. Create and track standards, such as feature importance, model transparency scores, and user understanding scores, to ensure the model's decisions are understandable and trustworthy.

Protect Against Algorithmic Bias

Incorporating AI into selection has the potential to enhance fairness, in part through the standardization of processes. To fully capitalize on this potential benefit, incorporate strategies to combat potential bias in the development and implementation of AI, given that the quality and fairness of AI depend on appropriate technologies, data, algorithms, and decisions relevant to the whole system.

It is essential to use high-quality, representative data to inform thoughtfully developed AI systems and avoid biases and other systematic distortions. By following the guidelines set forth by the Data and Trust Alliance,[3] you can more effectively choose the appropriate selection criteria. Moreover, while AI may improve efficiency, responsible oversight is crucial. Users must be educated about how to use AI appropriately. This balanced approach ensures that AI complements rather than supplants human judgment, thus maintaining the integrity and fairness of selection decisions.

From principles to practice:

- **Form a diverse oversight committee**. Assemble a multidisciplinary team comprising individuals from diverse backgrounds and areas of expertise, including members of the community and AI experts. This group will ensure that the use of AI in the selection process is scrutinized for fairness and representativeness.

- **Consider potential bias in the data**. Evaluate the data for potential biases, such as lacking representation from applicant groups and relying on outcomes with known biases. Just because a particular variable or source of data can be used doesn't mean it should be used. Interrogating the data will help manage the risk of the AI tool inadvertently perpetuating biases.

- **Conduct pilot tests**. Before full implementation, pilot the AI tool in a low- or no-stakes setting to ensure that it runs smoothly within the context of the overall system. This preliminary testing phase is critical for refining the tool before it is implemented in a given selection system to ensure it operates as intended for different student groups. Additionally, include specific use case testing to identify and address any unintended consequences early on. This expanded testing will help validate the metrics of success for the desired AI models, ensuring they align with institutional goals and standards.

- **Do not change the process mid-cycle**. Use the same data collection, analysis, and evaluation process for an entire application cycle. This consistency is vital for fairness and transparency. Track all changes in these processes when they occur.

- **Audit AI systems regularly**. Schedule and conduct an annual audit of the AI system and its output to identify AI-related biases and other problems in the selection process. Collaborate with a dedicated team of experts to analyze the findings and develop strategies for continuous improvement to be implemented for the next cycle. Consult recent[4] and relevant journal articles[5] and technical reports that have used AI in selection processes, explore tools used to examine the potential for bias like Admissible ML[6] or AI Fairness 360,[7] and consult legal counsel when appropriate.

Provide Notice and Explanation

Just as clear communication of selection criteria is important, it is equally critical to share openly that AI is being used in the evaluation process. Aligning with the *Blueprint for an AI Bill of Rights*,[8] applicants have the right to be aware of the use of automated systems and to understand how these systems might affect their chances and their access to opportunities.

It is an individual institution's responsibility to ensure that explanations about AI's role in evaluations are clear and understandable, while maintaining the integrity of the selection process. Generally keeping applicants well-informed about the methods and means through which their applications are assessed will build applicant trust and may reduce concerns about AI use. Additionally, establish a means for the impacted population to provide input throughout the development process, which will build trust and better ensure the system is responsive to their needs and concerns.

From principle to practice:

- **Disclose data sharing practices**. Ensure that applicants are fully informed about how their data might be shared through a privacy notice. Clarify the safeguards in place to comply with legal and ethical standards, providing insight into the process and building trust. This disclosure should align with institutional policies to ensure that potential data sharing is conducted responsibly and with full accountability.

- **Disclose AI usage**. Consider indicating on the website that AI is being used and at what stage of the decision-making process. Clarify the roles of AI and human decision-makers at each step. Consult legal counsel to ensure that all disclosures adhere to legal requirements.

Protect Data Privacy

Update the existing privacy policy as needed to cover the use of AI within the selection process. The updated policy should specify what data are collected, the methods of data collection, storage, analysis, protection, and the protocols for eventual deletion when appropriate. Ensure that only the data essential for fulfilling selection goals are collected, limiting the potential risk of data exposure.

From principle to practice:

- **Formulate a detailed AI privacy policy**. Create and display an AI-based data privacy policy on the institution website. This policy should comprehensively cover how data used in the selection process is collected, stored, analyzed, protected, and appropriately deleted. Ensure that it complies with all relevant privacy regulations and meets the unique needs of the selection process.

- **Provide contact information for inquiries**. Ensure that contact information is easy to find on the website so applicants can ask questions or raise concerns about how their data is being handled. Include provisions for the "right to be forgotten," allowing individuals to request the removal of their data from the organization's systems and models.

- **Limit data risk**. Collect only the data that is necessary to achieve the goals of the selection process. Adhere strictly to the principle of data minimization, restricting the collection of personal data to what is essential for the evaluation process.

- **Ensure secure use of AI tools**. Ensure applicant data is only used in settings with enterprise-grade security (e.g., Anthropic; ChatGPT Enterprise; Llama 3). Avoid sharing data while using personal versions of web-based AI tools. Consult with technology and legal experts to ensure that third-party vendors are contractually required to protect all data in accordance with institutional policies.

Incorporate Human Judgment

Incorporating AI into the selection process requires a thoughtful balance between technological advancements and the irreplaceable value of human judgment, insights, and ethical standards. Emphasizing human oversight as a foundational practice allows invested parties to harness AI's capabilities effectively, while ensuring that decisions are informed and aligned with the core values and objectives of the institution.

From principle to practice:

- **Involve humans from end-to-end**. To ensure effective implementation, integrate nontechnical subject matter experts at every stage, including problem formulation, data curation and relevance, feature engineering, error analysis, model evaluation, and user interface design.[9] Leveraging domain expertise will help increase accurate and actionable outcomes, with experts in the loop to enhance understanding, align with goals, and manage risk.

- **Use AI as decision support**. Emphasize and safeguard the pivotal role of human judgment in the evaluation process. Ensure that AI and ML systems are deployed to complement and enhance, not replace, human decision-making capabilities. This approach helps maintain essential human qualities that are vital for nuanced and well-informed decision-making (e.g., empathy, teamwork, situational awareness).

- **Develop understanding of AI**. Provide training for administrators to expand their understanding of AI and ML. Focus training on basic understanding of the AI techniques used, as well as the critical integration of human judgment in interpreting outcomes, to enhance the effectiveness and sensitivity of the process. Explore the references below as a starting point for content and as a bridge to more technical collaborators such as data scientists or software engineers.

Monitor and Evaluate

Implementing AI ethically in selection requires ongoing vigilance through regular monitoring and evaluation. Just as with any element of the selection process, it is essential to provide ongoing evidence that demonstrates the validity, reliability, and fairness of AI tools and systems. This persistent oversight ensures that AI implementations stay aligned with the institution's goals and ethical standards and helps manage legal and data security risks.

From principle to practice:

- **Establish standards for ongoing evaluation**. Implement standards for classification metrics, interpretability, and data/concept drift.[10] Additionally, monitor both user and applicant reactions to ensure the system is perceived as effective, fair, understandable, and aligned with institutional goals for the tool. This approach will help to identify areas for improvement and promote compliance with ethical standards.

- **Monitor and adjust after each cycle**. At the conclusion of each application cycle, conduct a thorough review of the AI system and its outputs. Assess whether and when adjustments are needed in the model, its application, or in the training provided to those who operate it. This regular monitoring ensures that the AI tool continues to align with institutional goals and adapts to any changes in the broader context of selection.

- **Remain responsive**. Keep the AI system current and relevant by adapting to significant changes in institutional curriculum and goals and in the applicant pool. Being proactive will help to maintain the system's effectiveness and alignment with each institution's evolving needs.

- **Document development and implementation**. Create a technical report or other documentation describing the steps and decisions involved in the development and implementation of AI systems. The report documentation should align with the Liaison Committee on Medical Education (LCME®) and other relevant requirements to support thorough evaluations and audits. Additionally, develop a standard operating procedure in accordance with the LCME requirements to ensure clarity and consistency for admissions officers.

AI in Admissions and Selection Technical Advisory Committee

We would like to thank the AI in Admissions and Selection Technical Advisory Committee for sharing their expertise, guidance, and best practices to develop these principles for responsible AI.

Technical Advisory Committee

Member	Title	School
Graham Keir, MD	Neuroradiology Fellow	Weill Cornell Medical Center, New York Presbyterian
Ioannis Koutroulis, MD, PhD, MBA	Associate Dean of MD Admissions	The George Washington University School of Medicine and Health Sciences
Richard Landers, PhD	John P. Campbell Distinguished Professor of Industrial-Organizational Psychology	University of Minnesota
Arun Mahtani, MD, MS	Cardiology Fellow	Virginia Commonwealth University
Fred Oswald, PhD	Herbert S Autrey Chair in Social Sciences, Director of Graduate Studies	Rice University
Kelly Trindel, PhD	Chief Responsible AI Officer	Workday
Laurah Tuner, PhD	Assistant Dean for Assessment and Evaluation	University of Cincinnati

AAMC Staff

Name	Title
Dana Dunleavy, PhD	Senior Director, Admissions and Selection Research and Development
Rebecca Fraser, PhD	Director, Content Development, Admissions and Selection Research and Development
Derek Mracek, PhD	Manager, Analytics and Evaluation, Admissions and Selection Research and Development
Jayme Bograd	Director, Pilot Administration, Admissions and Selection Research and Development
Melissa Lee	Senior Analyst, Content Development, Admissions and Selection Research and Development

Sources Cited

1. Ferguson E, James D, Madeley L. Factors associated with success in medical school: Systematic review of the literature. BMJ. 2022;324(7343):952-957. doi: 10.1136/bmj.324.7343.952

2. Campion ED, Campion MA, Johnson J, et al. Using natural language processing to increase prediction and reduce subgroup differences in personnel selection decisions. J Appl Psychol. 2024;109(3):307-338. doi: 10.1037/apl0001144

3. Data and Trust Alliance. Algorithmic bias safeguards. Published December 8, 2021. Updated July 9, 2024. https://dataandtrustalliance.org/work/algorithmic-safety-mitigating-bias-in-workforce-decisions

4. Keir G, Hu W, Filippi CG, Ellenbogen L, Woldenberg R. Using artificial intelligence in medical school admissions screening to decrease inter-and intra-observer variability. JAMIA Open. 2023;6(1). doi: 10.1093/jamiaopen/ooad011

5. Rottman C, Gardner C, Liff J, Mondragon N, Zuloaga L. New strategies for addressing the diversity-validity dilemma with big data. J Appl Psychol. 2023;108(9):1425. doi: 10.1037/apl0001084

6. H2O.ai. Admissible machine learning models. Updated July 9, 2024. Accessed June 11, 2024. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/admissible.html

7. IBM. AI Fairness 360. Accessed June 11, 2024. https://aif360.res.ibm.com/

8. Park E. The AI bill of rights: A step in the right direction. Orange County Lawyer Magazine 2023;65(2).

9. Huyen C. The human side of machine learning. In: Designing Machine Learning Systems. O'Reilly Media; 2022:chap 11.

10. Huyen C. The human side of machine learning. In: Designing Machine Learning Systems. O'Reilly Media; 2022:chap 11.

Appendix

The following journal articles, organizational resources, and technical references support the Principles for Responsible AI developed by the AAMC and the AI in Admissions and Selection Technical Advisory Committee.

Additional Relevant Journal Articles

Drum B, Shi J, Peterson B, Lamb S, Hurdle JF, Gradick C. Using natural language processing and machine learning to identify internal medicine-pediatrics residency values in applications. Acad Med. 2023;98(11):1278-1282. doi: [10.1097/ACM.0000000000005352](10.1097/ACM.0000000000005352)

Knopp MI, Warm, EJ, Weber, et al. AI-enabled medical education: AI enabled medical education: Threads of change, promising futures, and risky realities across four potential future worlds. JMIR Med Educ. 2023;9. doi: [10.2196/50373](10.2196/50373)

Mahtani AU, Reinstein I, Marin M, Burk-Rafel J. A new tool for holistic residency application review: Using natural language processing of applicant experiences to predict interview invitation. Acad Med. 2023;98(9):1018-1021. doi: [10.1097/ACM.0000000000005210](10.1097/ACM.0000000000005210)

Triola MM, Reinstein I, Marin M, et al. Artificial intelligence screening of medical school applications: Development and validation of a machine-learning algorithm. Acad Med. 2023;98(9):1036-1043. doi: [10.1097/ACM.0000000000005202](10.1097/ACM.0000000000005202)

Zhang N, Wang M, Xu H, Koenig N, Hickman L. Reducing subgroup differences in personnel selection through the application of machine learning. Pers Psychol. 2023;76(4):1125-1159. doi: 10.1111/peps.12593

Organizational Resources

Association of Test Publishers. Artificial intelligence and the testing industry: A primer. Published July 6, 2021. https://www.testpublishers.org/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_6July2021_Final%20R1%20.pdf

Department of Education Office of Educational Technology. Artificial intelligence and the future of teaching and learning: Insights and recommendations. Published May 2023. https://tech.ed.gov/files/2023/05/ai-future-of-teaching-and-learning-report.pdf

European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM/2021/206 final. Published April 21, 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Published January 2023. https://doi.org/10.6028/nist.ai.100-1

Organisation for Economic Co-operation and Development. The OECD Artificial Intelligence Principles. Published May 2019. Updated May 2024. https://oecd.ai/en/ai-principles

Society for Industrial and Organizational Psychology. Considerations and Recommendations for the Validation and Use of AI-based Assessments for Employee Selection. January 2023. https://www.siop.org/Portals/84/SIOP%20Considerations%20and%20Recommendations%20for%20the%20Validation%20and%20Use%20of%20AI-Based%20Assessments%20for%20Employee%20Selection%20010323.pdf?ver=5w576kFXzxLZNDMoJqdIMw%3d%3d

Stanford HAI. The 2019 AI Index Report. Published December 2019. https://hai.stanford.edu/research/ai-index-2019

Technical References

FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.

Association for Computing Machinery; June 12-15, 2023; Chicago, IL. https://dl.acm.org/doi/proceedings/10.1145/3593013

Hooker S. Fairness, security, and governance in machine learning. Stanford University; 2022. https://docs.google.com/presentation/d/1cshMKKSX24L0RL7LNzyOkZNQHD7N-Zyff8iffrLIVYM/edit

Montreal AI Ethics Institute. https://montrealethics.ai/

Trustworthy Machine Learning. Resources. https://trustworthyml.org/resources