# Validity Issues and Ethics Concerns of AI-Assisted Assessment

\*The data collection of this chapter is assisted by ChatGPT 4.o.

## Learning Objectives:

1. Provide an overview of AI's role in assessment, emphasizing the importance of validity and ethics in AI-based assessment systems.
2. Explore the concept of validity in AI assessment, including definitions, types, and challenges, with real-world examples and case studies.
3. Investigate possible solutions to the ethical concerns surrounding AI in assessment, covering bias, fairness, privacy, transparency, accountability, and regulatory frameworks.
4. Offer strategies and guidelines to address validity and ethics concerns in AI-based assessment, highlighting collaborative efforts and future trends in the field.

## Introduction

 Read this article online and discuss this question in the title with your peers.

Do you ever imagine that one day you could be accused of cheating based on an AI detection tool as a university student?

This nightmare became a reality for several students at UC Davis in 2023. As reported by USA Today, an AI detection tool flagged these students' work as potentially plagiarized, leading to false cheating accusations and significant distress. This incident not only highlights the growing presence of AI in educational assessment but also underscores the critical importance of addressing validity and ethics in these systems.

This story well illustrates the significance of our topic: the validity issues and ethical concerns of human-led AI assessment. Human-led AI assessment is an approach where AI tools assist in evaluating and analyzing data, but the final decisions and interpretations are made by human experts, ensuring that the assessment process remains guided by human judgment and ethical considerations. It is a prime example of how AI, while powerful, can falter in high-stakes environments such as education. The technology, meant to uphold academic integrity, instead created a scenario where its own validity was questioned, and ethical considerations were thrust into the spotlight.

In this chapter, I aim to delve into the intricacies of AI's role in assessment, focusing on the paramount importance of validity and ethics. AI has revolutionized various fields, including education, where it promises efficiency and objectivity. However, as the UC Davis incident illustrates, these systems are not infallible and can lead to significant errors and ethical dilemmas. Through a problem-based approach, I will explore the concepts of validity in AI assessments, investigating definitions, types, and the challenges we face in ensuring that these systems measure what they are supposed to measure accurately and fairly.

Validity in AI assessment is multifaceted, encompassing construct validity, content validity, and criterion validity, among others. Each type presents unique challenges, particularly in the context of AI, where algorithms may misinterpret data or apply inappropriate standards. Real-world examples, such as the UC Davis case, provide a stark reminder of the consequences of neglecting these validity concerns.

Moreover, the ethical landscape of AI in assessment is fraught with validity issues such as bias, fairness, privacy, transparency, and accountability. The false accusations at UC Davis stem from a lack of transparency and accountability in the AI tool's decision-making process. These ethical concerns must be addressed to build trust and ensure that AI systems in education serve all students equitably. Bias in AI, for instance, can perpetuate existing inequalities, while privacy issues can arise from the extensive data collection required for AI systems to function effectively.

In addressing these concerns, this chapter will offer strategies and guidelines to enhance both the validity and ethics of AI-based assessments. Collaborative efforts between educators, technologists, and policymakers are essential to develop robust frameworks that govern the use of AI in education. Future trends in AI assessment will also be discussed, highlighting the potential for advancements that prioritize ethical considerations and rigorous validation processes.

By starting with the real-world vignette of the UC Davis incident, this chapter aims to engage readers with a tangible example of the significant impact that validity and ethical issues in AI assessment can have on students' lives. Through this lens, I will provide an in-depth exploration of the topic, equipping educators, students, and policymakers with the knowledge and tools to navigate the complex landscape of AI in educational assessment.

## New discipline------The science of evaluation for AI models

 What does the science of evaluation for AI aim for?

   It is aimed to assesshow well AI systems perform tasks they're designed to do.

This discipline is crucial for ensuring that AI technologies are effective, reliable, and ethical in their applications.

#### Objectives of the Science of Evaluation for AI:

1.     Performance Assessment:

* Its primary goal is to evaluate the performance of AI systems in executing their intended tasks. For instance, in healthcare, an AI system designed for diagnosing diseases would be evaluated by comparing its diagnoses against the consensus of expert human practitioners. This involves measuring the accuracy, precision, recall, and F1 score to quantitatively determine the system's performance.

2.     Reliability and Fairness:

* Assessing AI's reliability entails examining how consistently the system performs under various conditions. Fairness is another critical factor, ensuring that the AI does not favor any group unfairly and operates equitably across diverse populations.

3.     Transparency and Accountability:

* Transparency in AI systems means making the decision-making processes understandable to humans. This involves elucidating how the AI arrives at its conclusions, which is essential for building trust and ensuring accountability.

4.     Handling Real-World Situations:

* AI systems must be evaluated for their ability to manage diverse and challenging real-world scenarios. This includes testing their robustness and adaptability to unforeseen situations.

What problems in our learning and life can we use the science of evaluation to investigate?

The science of evaluation for AI is a multidisciplinary field that addresses the complex challenges to ensure that AI systems are trustworthy, ethical, and effective. By focusing on comprehensive evaluation metrics and methodologies, this discipline aims to enhance the reliability and impact of AI technologies across various domains.

####   Science of Evaluation for AI addresses the Key Issues such as:

1.     Detecting and Combating AI-Generated Disinformation:

* One significant issue is the ability of AI to generate and spread disinformation. Evaluation frameworks must be developed to detect such disinformation and implement strategies to combat it effectively.

2.     Evaluation and Measurement Tools for Trustworthy AI:

* Developing tools and methodologies to evaluate AI systems for trustworthiness. This includes metrics and benchmarks that ensure the AI operates as intended and maintains integrity.

3.     Streamlining Data Model Creation:

* Utilizing platforms, MLOps systems, and tools to streamline the creation of data models. This is essential for efficient AI development and deployment.

4.     Customized Local Models:

* Creating and evaluating customized local models that cater to specific needs and contexts. This involves ensuring these models are optimized and effective for their intended applications.

5.     Setting Benchmarks and Managing Risks:

* Establishing benchmarks for AI performance and identifying potential risks. This is crucial for maintaining safety standards and managing the inherent risks associated with AI technologies.

6.     AI Safety and Risk Research:

* Conducting research focused on AI safety and risk management. This includes studying potential hazards and developing strategies to mitigate them.

7.     Healthcare AI Evaluation:

* Specifically evaluating AI applications in healthcare to ensure they meet the stringent standards required in medical practice. This involves rigorous testing and validation against clinical benchmarks.

8.     Optimization of AI Models:

* Continuously improving AI models to enhance their efficiency and effectiveness. This includes fine-tuning algorithms and processes to achieve optimal performance.

9.     AI Ethics and Equity:

* Ensuring that AI systems adhere to ethical standards and promote equity. This involves evaluating the ethical implications of AI decisions and their impact on different populations.

10.  Leveraging Technological Collaboration for Public Good:

* Promoting collaboration among technologists, policymakers, and other stakeholders to leverage AI for public good. This involves developing policies and frameworks that support the ethical and beneficial use of AI.

## Understanding validity issues in AI-assisted Assessment

   Validity refers to the degree to which an assessment accurately measures what it is intended to measure.  The concept of validity applies to any assessment be it in the context of AI checkers or AI-assisted assessment.In the context of AI-assisted assessment, ensuring validity is critical to guarantee that the conclusions drawn from the assessment results are accurate and meaningful. There are over 150 types of validity. The following are several common types of validity, each with its own focus and importance:

1.     Construct Validity: This type assesses whether the tool truly measures the concept it intends to measure. For example, if an AI tool is designed to evaluate critical thinking skills, construct validity would ensure that it accurately captures all dimensions of critical thinking.

2.     Content Validity: This type examines whether the assessment content covers the entire range of the concept being measured. An AI tool assessing mathematical ability should include a comprehensive set of problems that cover all relevant areas of mathematics.

3.     Criterion Validity: This type involves comparing the AI assessment results with other established measures or outcomes. For instance, an AI tool’s predictions of student performance could be validated by comparing them with actual grades or standardized test scores.

4.     Face Validity: Though more superficial, face validity considers whether the assessment appears to measure what it is supposed to measure. If the AI tool is user-friendly and intuitively aligned with the assessment goals, it is likely to have high face validity.

5. other types of validity such as \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Choose one type of validity from above and discuss the relevant validity issues that you are concerned with if AI-assisted assessment is applied to a course of your majors.

#### Challenges and Considerations in Ensuring Validity in AI-Assisted Assessment

To ensure validity in AI-assisted assessments poses several challenges and requires careful consideration of various factors:

1.     Data Quality and Bias: The accuracy of AI assessments heavily depends on the quality of data used to train the algorithms. Biases in training data can lead to skewed results, affecting the validity of the assessment. Ensuring diverse and representative data sets is crucial to mitigate this issue.

2.     Algorithm Transparency: The black-box nature of many AI algorithms can obscure how decisions are made, making it difficult to assess their validity. Transparent algorithms that allow stakeholders to understand and verify the decision-making process are essential for maintaining validity.

3.     Dynamic Learning Environments: Educational settings are dynamic, with constantly evolving curricula and student needs. AI tools must be adaptable and regularly updated to ensure they remain valid and relevant in changing environments.

4.     Contextual Factors: The context in which an assessment is administered can significantly impact its validity. Factors such as the testing environment, the stakes of the assessment, and individual student differences must be considered when evaluating validity.

5.     Ethical Implications: Ethical considerations, including fairness, privacy, and accountability, intersect with validity issues. Ensuring that AI assessments do not unfairly disadvantage any group and that students' data are protected is essential for maintaining both ethical standards and validity.

####  Food for Discussion

·       Debating the Balance: How can we balance the need for sophisticated AI algorithms with the necessity for transparency and understandability? Should simpler, more interpretable models be preferred over complex, opaque ones?

·       Role of Human Oversight: What is the appropriate level of human oversight in AI-assisted assessments to ensure validity without undermining the efficiency and objectivity that AI brings?

·       Adapting to Change: How can AI systems be designed to adapt to the evolving nature of educational environments while maintaining high validity standards?

·       Mitigating Bias: What strategies can be implemented to ensure that AI tools are trained on unbiased data and that their application does not perpetuate existing inequalities?

·       Ethical Frameworks: What ethical frameworks should be established to guide the development and implementation of AI-assisted assessments, ensuring that validity and fairness are upheld?

By understanding and addressing these validity issues, educators, researchers, and policymakers can work towards creating AI-assisted assessment systems that are accurate, reliable, and equitable, ultimately enhancing the educational experience for all students.

Ethics Concerns in AI-Based Assessment

A.   General Ethical Implications of AI-assisted Assessment

B.    Ethical Guidelines for Designing and Implementing HCI of AI-Based Assessment Systems

## Real-World Examples and Discussion

Group work: Discuss or debate the relevant validity issues and ethics concerns of human-led AI assessment in these real-world examples. Your discussion can follow three key questions of formative assessment: where we are now? Where we are going? How to get there? (Black & Wiliam, 2009)

1. How can we ensure that AI-based assessment tools are both valid and fair?

Proctorio Controversy: Proctorio, an AI-based proctoring software widely used in educational institutions for remote exams, faced scrutiny over its validity and ethical implications. Students reported instances of racial bias and privacy concerns, alleging that the software unfairly flagged certain behaviors as cheating based on cultural differences.

1. What are the potential ethical concerns of using AI in assessment according to your observation in your life? and how can they be addressed?

           Amazon's Recruiting Tool: In 2018, Amazon scrapped an AI-based recruiting tool after discovering bias against female candidates. The system, trained on resumes submitted over a ten-year period, consistently downgraded resumes containing the word "women's" or from all-female colleges. This incident underscores the ethical considerations in AI-based assessment, emphasizing the need for transparency and accountability in algorithmic decision-making processes.

3.     Illustrate with one example from your real-world experience on how bias has manifested in AI-assisted assessment systems, and what steps can be taken to mitigate it?

Bias in Automated Essay Scoring: Automated essay scoring systems, which use AI algorithms to evaluate written responses, have been criticized for perpetuating bias. Research has shown that these systems often favor essays written in a certain style or language, disadvantaging students from diverse backgrounds. Such findings underscore the importance of ongoing validation and refinement of AI-based assessment tools to minimize bias and ensure fairness.

4.     What role do you think should you play in managing your use of AI-assisted assessment?

Facial Recognition in Grading: Some universities have experimented with AI-powered facial recognition technology to monitor student engagement and assess their understanding during lectures. However, concerns have been raised regarding the validity of using facial expressions as indicators of comprehension and the potential invasion of students' privacy. This example illustrates the complexity of integrating AI into educational assessment while upholding ethical standards and ensuring validity.

1. What role do you think should educational institutions play in regulating the use of AI-assisted assessment?

AI in College Admissions: Colleges and universities are increasingly using AI algorithms to assist in the admissions process, raising concerns about fairness and transparency. Critics argue that reliance on AI may exacerbate existing inequalities in access to higher education, as the algorithms could inadvertently favor applicants from privileged backgrounds. This case highlights the ethical dilemmas inherent in using AI to make high-stakes decisions that profoundly impact individuals' lives and futures.

References

Black, P., & Wiliam, D. (2003). ‘In praise of educational research’: Formative assessment. British educational research journal, 29(5), 623-637.

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). The cowriter project: Teaching a robot how to write. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts(pp. 269-269).

Liang, L., (2017). The bearing of macro and micro culture on digital activity: EFL teachers' perception of computer-based matriculation oral language test reform. Learning, Culture and social Interaction (SSCI).12, 32-44.

Liang, L., Tognolini, J., Hendry, G., & Mantai, L. (2022). A review of tertiary formative assessment using digital technology in the past decade: what has been facilitated?. Proceeding in 8th International Conference on Higher Education Advances (pp. 119-126). Universitat Politècnica de València.<http://dx.doi.org/10.4995/HEAd22.2022.14371>

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. ACM Transactions on Interactive Intelligent Systems (TiiS), 11(3-4), 1-45.

Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., ... & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. ACM Computing Surveys, 55(13s), 1-42.

Reddy, S., Rogers, W., Makinen, V. P., Coiera, E., Brown, P., Wenzel, M., ... & Kelly, B. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ health & care informatics, 28(1).

van Wissen, A., Gal, Y. A., Kamphorst, B. A., & Dignum, M. V. (2012). Human–agent teamwork in dynamic environments. Computers in Human Behavior, 28(1), 23-33.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223-235.

Read this online at <https://edtechbooks.org/introduction_to_ai_and_ethics_in_higher_education/sgdcfqmfey>