



Case Study 6

Responsible AI for Test Equity and Quality: The Duolingo English Test as a Case Study

Abstract (Summary):

This paper, forthcoming in the Handbook for Assessment in the Service of Learning, uses the Duolingo English Test (DET) as a case study to explore the critical role of Responsible AI (RAI) in high-stakes assessment. While AI offers efficiencies in item generation and scoring, it also presents risks, such as bias. The DET addresses these risks through four specific RAI standards—Validity and Reliability, Fairness, Privacy and Security, and Accountability and Transparency—demonstrating how these translate into practical implementation.

The paper details the development of these standards and their connection to broader RAI principles, providing concrete examples of how these practices ensure both test quality (valid score inferences) and equity (fairness for all test takers). The DET serves as a model for how companies can transparently and accountably integrate AI while mitigating ethical risks.

Theme of the Case Study:

- English proficiency test
- AI fairness and unbiased in tests
- (Human in the loop) HiTL and AI
- Responsible AI

Relevance to the Reader:

This paper's relevance to the reader centres on the crucial role of Responsible AI (RAI) in ensuring both the quality and equity of AI-powered assessments. It uses the Duolingo English Test (DET) as a compelling case study, detailing how RAI principles are implemented in practice to mitigate the risks of AI while maximizing its benefits. This is valuable for anyone involved in assessment development, AI ethics, or the use of AI in high-stakes testing. The paper provides a detailed, practical example of how RAI principles can be implemented in a real-world, high-impact setting. This is crucial given the increasing use of AI in education and assessment. It also offers insights into the process of creating an AI-powered assessment, including item generation, scoring, and security measures. It highlights the importance of human-in-the-loop approaches for maintaining quality and equity. AI ethics and governance: The four RAI standards (Validity and Reliability, Fairness, Privacy and Security, Accountability and Transparency) presented serve as a model for organizations developing and deploying AI systems, particularly in high-stakes applications. The alignment with the NIST AI RMF further strengthens the framework's validity. The paper demonstrates how a systematic application of RAI practices can contribute to achieving both test equity (fairness for all test takers) and quality (valid score inferences). This is crucial for ensuring that AI-powered assessments are not biased or discriminatory.