

Using AI for faster and better scoring



Our human-in-the-loop process for scoring

With AI support, we enhance the scale and speed of our gold-standard assessment scoring — with human-expert checks and balances every step of the way.

First, we decide what to score.

Then, we build and validate the AI scoring models.

Next, we deploy and continuously evaluate and refine the models.

1

Identify the scoring parameters, and establish what's appropriate for AI scoring.

2

All experts agree on the protocol for evaluating the AI scoring.

3

Review with the Technical Advisory Committee (TAC) and incorporate modifications to the evaluation protocol.

4

Human-score a set of test items.

5

Build and train the AI scoring algorithm that will replicate and reflect our human scoring.

6

Run psychometric analyses on a pilot test of AI-scored items to ensure that they're reliable and valid.

7

Conduct fairness and bias reviews. Review results with the TAC and implement their recommendations.

8

Deploy the AI scoring models.

9

Employ review processes where every response is evaluated by both the AI and a human. If they disagree, a second human review is called for.

10

Conduct periodic or continual operational reviews to monitor for fairness and drift.



Learn more at ets.org/toefl

Copyright © 2024 by ETS. TOEFL and TOEFL IBT are registered trademarks of ETS in the United States and other countries. The Eight-Point logo is a trademark of ETS. All other trademarks are property of their respective owners. andk000038