

Automatic Item Generation

- Computer-based testing (CBT) has been growing more popular alongside the advancement of educational technology. As the demand for CBT rises, it is vital that large item banks are also being created. The traditional method of producing items is **expensive, time consuming, and lacks scalability**.
- Automatic Item Generation (AIG) is an **augmented intelligence approach** that combines the expertise of **subject matter experts** with **computational power** to efficiently produce large numbers of items.



- Prior to the rise in large language models, AIG depended on specialized software to integrate SME instructions with computational capabilities. With the advancements of large language models, specifically OpenAI's GPT models, there is potentially no longer a need for specialized software. GPT-4 may be a more accessible solution to generating items and facilitating a novel approach to AIG.

Item Evaluation

1. Is the question clear and unambiguous?
2. Is the stem focused on a single and meaningful problem?
3. Are alternatives homogeneous (e.g., length, complexity)?
4. Is the question free from verbal clues to answer?
5. If applicable, is the source material necessary for the question?
6. Is the question free from spelling and grammar mistakes?
7. Is the keyed response properly identified?
8. Does the question avoid using correct responses or words repeated in the stem and alternatives?
9. Are all the alternatives plausible?
10. Is there only one correct response?
11. Is the item aligned with the template objective?

Rating Categories:

Accept Item

- All 11 of the criterion are satisfied
- No changes to the item are required

Minor Revisions

- 1-3 of the criterion are not satisfied
- Minor changes to the item are required

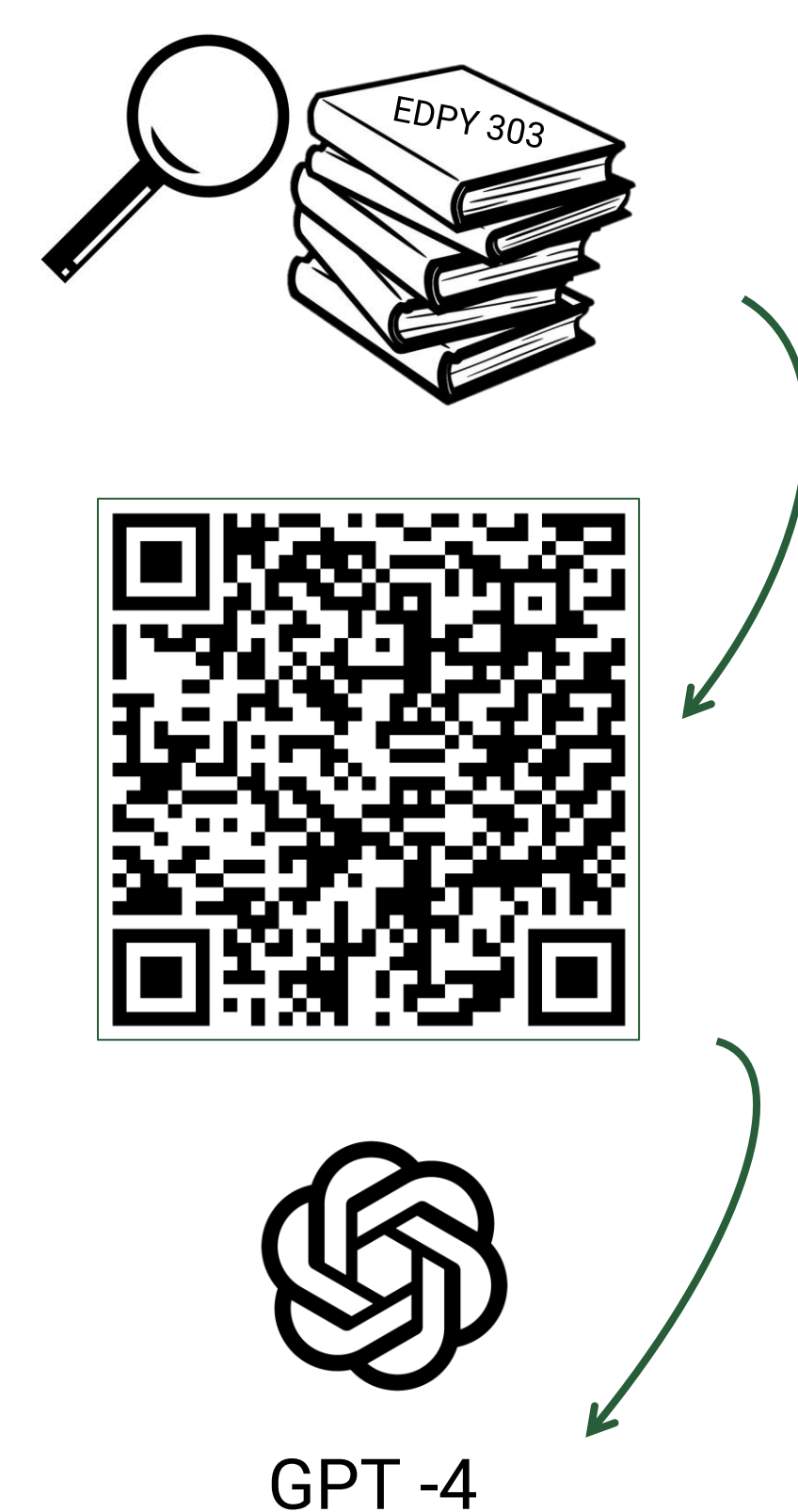
Reject

- 4+ of the criterion are not satisfied
- Requires change of more than 20% of the item

Methods

Gierl, Lai, and Tanygin (2021) developed a three-step method to implement AIG.

- 1) The subject-matter expert creates a cognitive model by identifying the content that will be used to produce new items.
- 2) The subject-matter expert creates an item model by extracting the content from the cognitive model and identifying the features that can be manipulated to create new items.
- 3) A computer uses the item model to generate a large number of new items.



Results

- Thirty-five cognitive models were created to cover material in EDPY 303 – Education Assessment. Each cognitive model focused on a single concept.
- Thirty items were generated per each of the cognitive models. In total, 1050 items were generated.
- Five items per cognitive model were randomly selected to be evaluated. Seven items were used for practice evaluation.
- 77/168 Items were accepted with no revisions required.
- 53/168 of the items required minor revisions.

Discussion & Conclusion

- Following the three step AIG method, GPT-4 can be leveraged in generating a large bank of items. Subject-matter experts can create specific instructions on the important characteristics of items and generate many parallel items.
- The quality and characteristics of the items are highly dependent on the quality of the prompt. Subject-matter experts can create specific instructions on the important characteristics of items and generate many parallel items.
- The creation of cognitive models requires a high level of investment from the subject matter experts.