



The Use of Artificial Intelligence for Item Writing

**Michelle Leong, Whitney Mitchell, Elana
Tetrushvily, & Frank Igou, PhD**

Department of Psychology
& Behavioral Sciences
Louisiana Tech University





Overview

1. Introduction to AI
2. AI in Test Construction
3. Limitations of AI in Test Construction
4. Legal Implication in Employment Settings
5. Resources for AI in Test Construction



MYTH	FACT
AI is only for large businesses	AI is available for every business
AI algorithm can process any data	The quality and quantity of data matters
AI can make independent decisions	AI programs need to be taught first
AI will replace all human jobs	AI complements humans and create new jobs
AI robots will enslave people	There will be no rise of the machines



What is Artificial Intelligence(AI)?

- Complex, confusing and misunderstood term
- Broad term that describes algorithmic systems that are programmed to achieve human-defined objectives
- AI exist in common activities, experiences and interactions today
- Although definitions and concepts of AI may differ based upon the goals in which it is used (International Privacy Subcommittee of the ATP Security Committee, 2021)



What is Artificial Intelligence(AI)?

- The definitions and concepts of AI may differ based upon the goals and the domain in which it is used
 - An AI system arrives at a prediction based upon one or more observations. AI model will always return a prediction
 - In its current state of development, it's more like simulated intelligence than a non-organic intelligence

KEY
COMPONENTS OF
**ARTIFICIAL
INTELLIGENCE**

DEEP LEARNING

Computers imitating the learning and thinking methods of humans



COMPUTER VISION

Helps computers in understanding the visual world



MACHINE LEARNING

Machines learning and improving by studying data and statistics



NEURAL NETWORK

Network of artificial neurons or nodes inspired by the human brain



COGNITIVE COMPUTING

Stimulation of human thinking processes in a computerized system



NATURAL LANGUAGE PROCESSING

Enables machines to understand human language





AI in Test Construction

1. How many items do we need in our item bank to reduce exposure and maintain exam integrity?
2. What areas should be covered to ensure an accurate comprehension of knowledge?
3. How many Subject Matter Experts (SMEs) do we need to recruit?



AI in Test Construction

- AI-assisted item generation to reduce the time needed to build and bank a range of quality items,
- Allocate more time in selecting and working with SMEs
- Helps in generating higher quality items the more it learns about your exam content (feeding AI information)



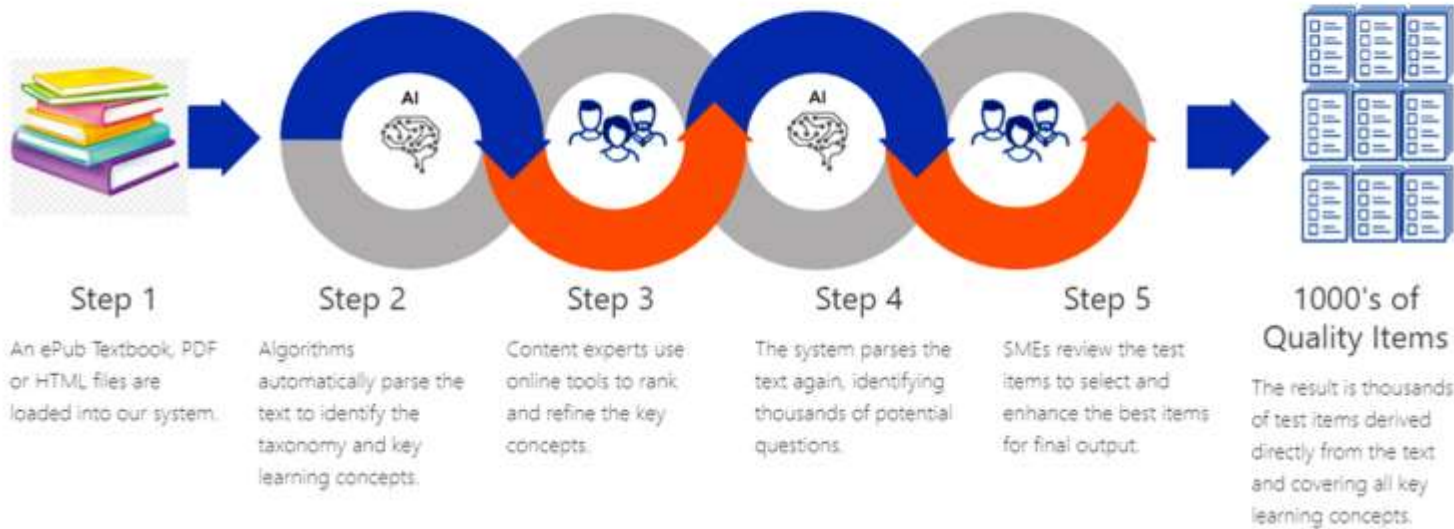
AI in Test Construction

- AI allows the best combinations of different modes and questions to assess any specific role or tasks
- This would accelerate the process towards customisation and away from generalized approaches.



AI in Test Construction

AI-Powered Content Generation Technology





Limitations of AI in Test Construction



Limitations of AI in Test Construction

- Problematic Datasets → Problematic Models
 - a. The incoming data contain built-in bias
 - b. Training data can unintentionally reinforce harmful stereotypes, exclude certain groups, and contain toxic sources.
 - c. Models trained on problematic datasets can inherit harmful stereotypes and biases.
 - i. Eg: biased towards Western cultures, social biases and stereotypes.



Limitations of AI in Test Construction

- Risk of Misinformation
 - a. AI generated content models can provide inaccurate or false answers, leading to misleading information.
 - i. Eg: inaccurate medical dosages and false traffic rules.



Limitations of AI in Test Construction

- Mitigation:
 - a. OpenAI developed pre-training techniques to address filter-induced biases.
 - b. Regularly updating models and collecting new training data is crucial to ensure accuracy and relevance.



Collaborate with SMEs and Stakeholders

- Using AI in personnel selection does not mean that you can replace human judgment and interaction.
- You still need to communicate, collaborate and review the items and tests created with SMEs, stakeholders, such as candidates, hiring managers, and employees
- You need to inform and educate them about the purpose, process, and benefits of using AI in hiring.



Legal Implications of AI Use in Employment Testing & Hiring



Legislation & Regulations

- Emergent regulations in both the EU and US which will increasingly put guidelines around how any solutions are constructed and used.
- A responsible AI approach should become the expected baseline for users.



Unintentional Discrimination

- Increase in numbers of discriminatory classification and decision support algorithms
- Avoiding discrimination is challenging in machine learning
- Forms of discriminatory distortion is inevitable
 - Even when the data is free from bias



Unintentional Discrimination

- Lack of access to protected data classes by their ML teams, making efforts to validate their model for bias or discrimination extremely difficult.
- Possible biases inherited by the algorithms from human prejudices and collection artifacts hidden in the training data
 - May lead to unfair or wrong decisions



Recommendations

1. Inform job applicant during the hiring process that AI is used to assess candidates and make hiring decisions
2. Obtain informed consent from job applicant before using AI to evaluate them



Resources for AI in Test Construction



OpenAI's GPT-4

GPT-4 is a large multimodal model that, while less capable than humans in many real-world scenarios, exhibits human-level performance on various professional and academic benchmarks including generating test questions.

It can be fine-tuned on specific domains or used in a prompt-based manner to generate questions. (<https://chat.openai.com/>)



ChatGPT

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Here is a list of their availability:

- Andrew: 11 am to 3 pm
- Joanne: noon to 2 pm, and 3:30 p
- Hannah: noon to 12:30 pm, and 4

Based on their availability, there is
them are available, which is from 4
scheduled at 4 pm.

GPT-4

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm
Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm
Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm



IBM Watson

IBM Watson is an AI platform that provides a range of services, including natural language understanding and generation. It can be used to generate test questions based on given criteria or to analyze and assess existing questions.

(<https://www.ibm.com/watson>)



Google Cloud Natural Language API

Google Cloud Natural Language API offers a set of natural language processing tools. While it doesn't specifically focus on test item generation, it can be used to extract information from texts, classify content, and grammatically correct questions.

<https://cloud.google.com/natural-language/>



Hugging Face

Hugging Face is a platform that provides a wide range of pre-trained models for natural language processing, including question generation. Their models can be fine-tuned on specific information or datasets and used to generate questions based on input text or prompts.

(<https://huggingface.co/>)



Talview

Talview is an AI-powered recruitment platform that offers a test generator feature. It uses natural language processing techniques to create test questions and assessments based on specific skills or job roles and can be used for structured oral interviews.

(<https://www.talview.com/>)



Hiring Platform

For quick, confident, and bias-free talent decisions

Exam & Proctoring Platform →

For simpler, more secure exam proctoring

Hiring Platform →

For quick, confident, and bias-free talent decisions

Exam & Proctoring Platform

For simpler, more secure exam proctoring



AI Proctoring

AI-powered remote proctoring software



Secure Browser

Prevent unauthorized access, control test environment



Certifast AI

100% AI-generated course and exam certification platform



Record & Review

AI proctoring with human review



Exam Engine

Speed, versatile, and secured online exams



Proctoring Comparison

Discover your ideal remote proctoring solution



Live Remote Proctoring



Interview Insights

Effective, AI-supported, bias-free candidate interviews



Behavioral Insights

Understand candidate behavior for better hiring



Drive Management

Minimize no-shows, enhance candidate experience



Recruitment Chatbot

Automate tedious recruitment tasks with hiring chatbot



Video Interviews

AI video interviewing for hassle-free, fast-track recruitment



Linguistic Insights

AI-Powered Voice Assessment for Evaluating Spoken English



Assessments

Speedy, versatile platform for secure online assessments



Live Interviews

Efficiently interview, review, and collaborate



Candidate Verification

Protect your reputation, say no to proxy candidates



Coding Tests

Effortlessly assess modern programming languages



Intelligent Scheduling

Conflict-free, automated interview scheduling



Hiring with MS Teams

Increase collaboration in your hiring process



Cautions and Limitations

Keep in mind that this is relatively new technology.

Even though these platforms can assist in generating test items, it's important that test developers and subject matter experts review any questions generated by AI to ensure accuracy, relevance, item quality, potential bias and ensure fairness.



ChatGPT Generated Items

In the following example, ChatGPT was used to draft items for an **Entry-level Police Officer exam**

The prompts used were **“entry-level police officer”, “selection exam”, “multiple-choice items”, “different levels of Bloom’s Taxonomy”** and **“use O*NET as the job content source”**

Bloom’s Taxonomy was added to vary the level of item difficulty and cognitive load.



ChatGPT Generated Items

Bloom's Taxonomy Level: Knowledge

What is the legal BAC (Blood Alcohol Content) limit for operating a vehicle in your jurisdiction?

- A. 0.05%
- B. 0.08%
- C. 0.10%
- D. 0.15%

Which of the following amendments in the U.S. Constitution protects against unreasonable searches and seizures?

- A. First Amendment
- B. Second Amendment
- C. Fourth Amendment
- D. Fifth Amendment



ChatGPT Generated Items

Bloom's Taxonomy Level: Comprehension

Based on the provided scenario, determine the correct course of action during a traffic stop: Scenario: A driver is stopped for a minor traffic violation and appears nervous, avoiding eye contact.

- A. Issue a warning and let the driver go without further investigation.
- B. Issue a citation for the traffic violation and ask the driver if they have anything illegal in the vehicle.
- C. Request backup and conduct a thorough search of the vehicle.
- D. Observe the driver's behavior for any signs of impairment or suspicious activity and document your observations.



ChatGPT Generated Items

Bloom's Taxonomy Level: Comprehension

Which of the following statements best describes the role of community policing?

- A. Community policing involves strict enforcement of laws and regulations to maintain order.
- B. Community policing focuses on developing partnerships with community members to address crime and public safety issues.
- C. Community policing aims to establish a militarized presence in high-crime neighborhoods to deter criminal activity.
- D. Community policing primarily relies on surveillance technology to monitor criminal behavior.



ChatGPT Generated Items

Bloom's Taxonomy Level: Analysis

Officer Johnson arrives at the scene of a car accident. What steps should he take to ensure the safety of everyone involved and secure the scene?

- A. Provide first aid to injured parties and call for backup
- B. Photograph the vehicles involved and interview witnesses
- C. Issue traffic tickets to the drivers involved in the accident
- D. Move the vehicles to the side of the road to clear traffic congestion



ChatGPT Generated Items

Limitations

- The previous slides contained simple examples of what is possible with the free version of ChatGPT.
- This was a simplified example of what is currently possible using a common, free, very basic AI platform
- The free version of ChatGPT can draft 100's of items in minutes and can be linked to specific content sources link O*NET
- However, currently, ChatGPT often gets facts wrong. Any items generated may need editing and subject matter expert review.



Questions? Comments?

Feel free to address any questions or comments about AI and item generation to:

Frank Igou, PhD.

figou@latech.edu

(318) 257-5455

Also, please feel to share any cutting-edge work you may be doing with AI and testing.

**John Connor
watching y'all make
friends with AI**



Thank You for Your Time!

Questions?



Presenters:

Michelle Leong – mcl041@latech.edu

Elana Tetrushvily – ete004@latech.edu

Whitney Mitchell – wdr008@email.latech.edu

Frank Igou – figou@latech.edu