Is ChatGPT a Threat to Financial Education?

Jonathan Daigle
Monmouth University

Yulin Li
Monmouth University

Bochen Li
Monmouth University

Background

- OpenAI introduced ChatGPT in November 2022
- Other large language models (LLMs) have become available
- Concern among educators about its potential misuse among students
- Our study is closely related to Wood et al. (2023)

Data

- Sample consists of test bank questions from a principles of finance textbook

- Includes 1146 questions from eleven chapters
  - 351 dichotomous (i.e., True/False)
  - 397 multiple-choice
  - 398 open answer

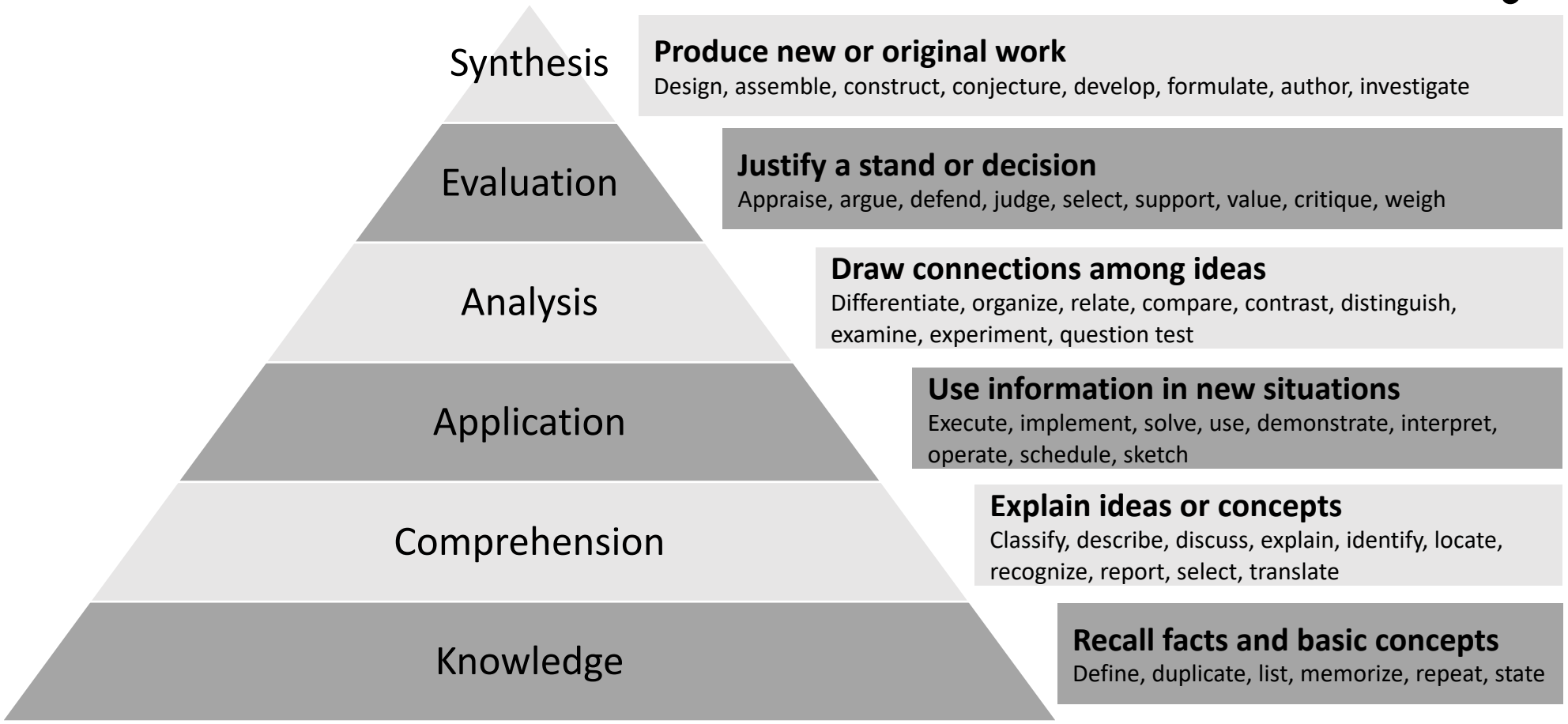
- The publisher also ranks the questions into difficulty tiers of easy, moderate, and challenging

Data

- The publisher also identified each question as belonging to one of Bloom's Taxonomy levels
  1) Knowledge – recalling facts and basic concepts
  2) Comprehension – explaining ideas or concepts
  3) Application – using information in new situations
  4) Analysis – drawing connections among ideas
  5) Evaluation – justifying a stand or decision
  6) Synthesis - producing new or original work

- Each question was individually submitted to ChatGPT-3.5 (chat.openai.com)
- The answer provided by ChatGPT was compared to each question's solution
- Prompts were not saved or used for training

Bloom's Taxonomy



Results

Table 1: ChatGPT's Performance with Textbook Test Bank Questions

Category	n	No. Correct	No. Incorrect	% Correct	% Incorrect
Overall	1146	612	525	54.2%	45.8%
Question Format					
Dichotomous	351	271	80	77.2%	22.8%
Multiple Choice	397	206	191	51.9%	48.1%
Open Answer	398	144	254	36.2%	63.8%
Question Type					
Conceptual	748	477	271	63.8%	36.2%
Mathematical	398	144	254	36.2%	63.8%

Results

Topic Area	n	No. Correct	No. Incorrect	% Correct	% Incorrect
Overview of Finance	66	54	12	81.8%	18.2%
Financial Markets	33	27	6	81.8%	18.2%
Financial Statements and Cash Flows	138	76	62	55.1%	44.9%
Financial Ratios	133	87	46	65.4%	34.6%
Time Value of Money	164	68	96	41.5%	58.5%
Interest Rates	82	47	35	57.3%	42.7%
Bond Valuation	92	45	47	48.9%	51.1%
Risk and Return	147	79	68	53.7%	46.3%
Stock Valuation	89	44	45	49.4%	50.6%
Cost of Capital	94	50	44	53.2%	46.8%
Capital Budgeting	108	44	64	40.7%	59.3%

Results

Table 1: ChatGPT's Performance with Textbook Test Bank Questions

Category	n	No. Correct	No. Incorrect	% Correct	% Incorrect
Difficulty					
Easy	408	289	119	70.8%	29.2%
Moderate	612	287	325	46.9%	53.1%
Challenging	126	45	81	35.7%	64.3%
Bloom's Levels					
Knowledge	247	205	42	83.0%	17.0%
Comprehension	293	183	110	62.5%	37.5%
Application	168	59	109	35.1%	64.9%
Analysis	368	145	223	39.4%	60.6%
Evaluation	64	27	37	42.2%	57.8%
Synthesis	6	2	4	33.3%	66.7%

Table 2: Regression Results					
	Dependent Variable: ChatGPT Answered Correct				
	(1)	(2)	(3)	(4)	(5)
Dichotomous	0.379***				0.202***
	(0.034)				(0.058)
Multiple Choice	0.142***				0.130***
	(0.036)				(0.045)
Conceptual		0.254***			
		(0.031)			
Easy			0.328***		0.223***
			(0.049)		(0.050)
Moderate			0.100**		0.044
			(0.047)		(0.045)
Comprehension				-0.194***	-0.116***
				(0.039)	(0.041)
Application				-0.437***	-0.238***
				(0.053)	(0.069)
Analysis				-0.433***	-0.218***
				(0.038)	(0.061)
Evaluation				-0.376***	-0.111
				(0.071)	(0.086)
Synthesis				-0.508**	-0.366*
				(0.239)	(0.217)
Constant	0.568***	0.576***	0.609***	0.917***	0.588***
	(0.055)	(0.059)	(0.066)	(0.049)	(0.091)
Observations	1146	1146	1146	1146	1146
Controls for Chapters	Yes	Yes	Yes	Yes	Yes
R^2	0.139	0.103	0.105	0.154	0.186
Adjusted R^2	0.130	0.095	0.095	0.143	0.172
Significance levels					*p<0.1; **p<0.05; ***p<0.01

Conclusions

- Our findings indicate ChatGPT performed the best with general knowledge, conceptual, and dichotomous questions (i.e., True/False).
- However, it often struggles with moderate and challenging questions and questions requiring levels above Bloom's definition of knowledge.
- Implications
  - If any assessments are given to students that do not require proctoring (i.e., take-home exams), then the professor should cautiously use dichotomous or knowledge-based questions.
  - Given ChatGPT's struggles with mathematical questions, professors should ask for detailed explanations of any work provided back by the students.