**Chapter 17: Statement of cash flaws**

# Generate a 40 MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a27ee5-930c-8006-84af-56cc7c4b013c>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/67b60e5e-441c-8004-8f34-1ec8b878bcd7> | **6**  <https://chatgpt.com/share/67b61242-5588-8004-bcf6-925a27b3a25e> |
| DeepSeek | **1** <https://chat.deepseek.com/a/chat/s/66fd67ae-cfa0-41c2-89b6-b04222bc824a> | **6**  <https://chat.deepseek.com/a/chat/s/ffc348be-ca12-4df6-b5a6-110c887bc25c> |
| Gemini 2 Flash | **1**  <https://g.co/gemini/share/daa091e72d6e> | **6**  <https://g.co/gemini/share/daa091e72d6e> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0** <https://chatgpt.com/share/67a5278b-466c-8006-9f1e-e3113fd29e7e> | **6**  <https://chatgpt.com/share/67a52a84-6ee0-8006-9cc1-31806f6040fe> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/67a7c532-57d0-8006-8bcd-153169d90668> | **6**  <https://chatgpt.com/share/67a7cd80-cb80-8006-9b09-69f499a7c47e> |
| Gemini 2.5 Pro Experimental | **1**  <https://docs.google.com/document/d/1Fhzno-NLsDCj7TwnNT1aCGunnXicQyk5YVIdapUEGrc/edit?tab=t.0> | **6**  <https://docs.google.com/document/d/1Fhzno-NLsDCj7TwnNT1aCGunnXicQyk5YVIdapUEGrc/edit?tab=t.0> |

1. Voting

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **13 C** | **1 M** | **7 M** | **11 M** | **16 M** | **17 M** | **20 M** | **# Fails** |
| **Acc. mentor (Q generator)** | **B** | **A** | **B** | **A** | **A** | **C** | **B** | **7** |
| **Free GPT-4o** | **A** | **B** | **A** | **B** | **B** | **B** | **A** | **0** |
| **DeepSeek** | **A** | **B** | **A** | **B** | **B** | **B** | **A** | **0** |
| **Geminai Flash** | **A** | **B** | **A** | **B** | **B** | **B** | **A** | **0** |
| **Paid GPT-4o** | **B** | **B** | **A** | **B** | **B** | **B** | **A** | **1** |
| **Wolfram 14.1** | **B** | **B** | **A** | **B** | **B** | **B** | **A** | **1** |
| **Geminai 2 Pro** | **A** | **B** | **A** | **B** | **B** | **B** | **A** | **0** |
| **Answer after voting** | **A** | **B** | **A** | **B** | **B** | **B** | **A** | **--** |

* Wrong answers are due to math issue

# Generate a 40 MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed here

<https://chatgpt.com/share/67a27c4e-b32c-8006-8f53-ef70d70c8be9>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/67b613d9-1330-8004-8c7f-e72c411baf57> | **7**  <https://chatgpt.com/share/67b6348b-da84-8004-bfb1-effcbd8ccd86> |
| DeepSeek | **0**  <https://chat.deepseek.com/a/chat/s/89427a72-4e9a-41ae-9aac-7aba475ba566> | **8**  <https://chat.deepseek.com/a/chat/s/02d88455-ec49-49f0-88e0-0fcd15fc2a7d> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/7ab71723cdbc> | **7**  <https://g.co/gemini/share/7ab71723cdbc> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/67b65eb7-6d68-8006-af33-93c57ae30665> | **7**  <https://chatgpt.com/share/67b6638e-e7d4-8006-87cf-d48234bcad93> |
| Accounting mentor | **0**  <https://chatgpt.com/share/67b65f48-cf00-8006-9e57-61431984a5a0> | **6**  <https://chatgpt.com/share/67b66780-e0cc-8006-8c4d-d53a160fc6ba> |
| Gemini 2.5 Pro Experimental | **0**  <https://docs.google.com/document/d/1Fhzno-NLsDCj7TwnNT1aCGunnXicQyk5YVIdapUEGrc/edit?tab=t.0> | **7**  <https://docs.google.com/document/d/1ct-nJRwQc6oQWWJzTMbl-u-Q4IlNFMMKyskbZWtViVU/edit?tab=t.0> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **25** | **27** | **28** | **31** | **32** | **33** | **34** | **37** | **# Fails** |
| **Wolfram (Q generator)** | **A** | **A** | **B** | **B** | **D** | **C** | **A** | **C** | **7** |
| **Free GPT-4o** | **C** | **B** | **A** | **A** | **B** | **B** | **A (Approx)** | **A** | **0** |
| **DeepSeek** | **C** | **B** | **A** | **A** | **B** | **B** | **B (Approx)** | **A** | **1** |
| **Geminai Flash** | **C** | **B** | **A** | **A** | **B** | **B** | **A (Approx)** | **A** | **0** |
| **Paid GPT-4o** | **C** | **B** | **A** | **A** | **B** | **B** | **A (Approx)** | **A** | **0** |
| **Acc. mentor** | **A** | **B** | **A** | **A** | **B** | **B** | **A (Approx)** | **A** | **1** |
| **Geminai 2 Pro** | **C** | **B** | **A** | **A** | **B** | **B** | **A (Approx)** | **A** | **0** |
| **Answer after voting** | **C** | **B** | **A** | **A** | **B** | **B** | **None** | **A** | **--** |

* Questions 25, 27, 33 (Accounting different understandings)
* Questions 28, 31, 32, 34, 37 (Math Issue)

# Generate a 40 MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1GsQ9JDRkWtuFs3VWrVnS4tl0mMSY36gSRKWLLilXz24/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/680ba9e4-4ab4-8007-b537-4107c52967b3> | **0**  <https://chatgpt.com/share/680ba9e4-4ab4-8007-b537-4107c52967b3> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/7abb2038e144> | **1**  <https://g.co/gemini/share/7abb2038e144> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6812293d-c3b8-800f-bd9f-8a9b3a397635> | **5**  <https://chatgpt.com/share/6812293d-c3b8-800f-bd9f-8a9b3a397635> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/68126804-bc48-800f-ba6f-db84bfcf5e3f> | **0**  <https://chatgpt.com/share/68126804-bc48-800f-ba6f-db84bfcf5e3f> |
| Accounting mentor | **0**  <https://chatgpt.com/share/68127593-ff6c-800f-9035-7a078f5454a6> | **0**  <https://chatgpt.com/share/68127593-ff6c-800f-9035-7a078f5454a6> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/31224732f113> | **1**  <https://g.co/gemini/share/31224732f113> |

1. Voting

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **5m** | **6m** | **10m** | **14m** | **16m** | **# Fails** |
| **DeepSeek (Q generator)** | **a** | **a** | **a** | **a** | **a** | **0** |
| **Free GPT-4o** | **a** | **a** | **a** | **a** | **a** | **0** |
| **Accounting mentor** | **a** | **a** | **a** | **a** | **a** | **0** |
| **Geminai Flash** | **a** | **a** | **a** | **a** | **b** | **1** |
| **Paid GPT-4o** | **b** | **b** | **b** | **b** | **b** | **5** |
| **Wolfram 14.1** | **a** | **a** | **a** | **a** | **a** | **0** |
| **Geminai 2.5 Pro** | **a** | **a** | **a** | **a** | **b** | **1** |
| **Answer after voting** | **A** | **A** | **A** | **A** | **A** | **--** |

* Question 16m (Accounting different understandings)

# Generate hard 20-MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a37629-9594-8006-8fe5-793b9f203d62>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/67be4bbd-38b0-8004-b58e-4cadcfe94277> | **3**  <https://chatgpt.com/share/67be4dad-ffe8-8004-9cd3-11c826c54736> |
| DeepSeek | **0**  <https://chat.deepseek.com/a/chat/s/0a9ea89b-6331-41d0-b455-4099bc24b193> | **5**  <https://chat.deepseek.com/a/chat/s/3b40a2b1-0940-4205-b9e8-2fd36598a747> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/7ab71723cdbc> | **6**  <https://g.co/gemini/share/7ab71723cdbc> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/67be36b8-c278-8006-95a5-75c754422456> | **5**  <https://chatgpt.com/share/67be3bfc-bf3c-8006-9eee-8d86750c235b> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/67be3d37-9378-8006-b6d8-ffb0aa7cb66b> | **4**  <https://chatgpt.com/share/67be4073-9000-8006-ba7d-6e118741f448> |
| Gemini 2.5 Pro Experimental | **0**  <https://docs.google.com/document/d/1GSb6VP_kJshFbJfR30WU_-Pd94iMP1xGcuMPiY81Zqw/edit?usp=sharing> | **5**  <https://docs.google.com/document/d/1ndYRHCyKpOqpSN8uuGDI1IROy3vd512skHVIL4D6hAg/edit?usp=sharing> |

1. Voting

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **16** | **17** | **18** | **19** | **20** | **# Fails** |
| **Accounting mentor (Q generator)** | **B** | **A** | **C** | **B** | **C** | **5** |
| **Free GPT-4o** | **C** | **B** | **A** | **B (Approx)** | **C** | **4** |
| **DeepSeek** | **A** | **None** | **A** | **None** | **None** | **1** |
| **Gemini Flash** | **A** | **B (Approx)** | **A** | **C** | **A** | **3** |
| **Paid GPT-4o** | **A** | **B (Approx)** | **A** | **C** | **C** | **3** |
| **Wolfram 14.1** | **A** | **A (Approx)** | **A** | **A** | **B** | **3** |
| **Geminai 2 Pro** | **A** | **None** | **A** | **None** | **None** | **0** |
| **Answer after voting** | **A** | **(None)** | **A** | **None** | **(None)** | **--** |

* In Question 17, 19, 20, the LLMs that choose an answer have a math issue. The right answer is none
* The previous questions 17, 19, 20 are conceptually wrong solved

# Generate hard 20-MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a37a4e-a328-8006-ab30-4b50d2be3fa8>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/67c7e60f-6d68-8004-9d83-8aaa2f78f6d1> | **3**  <https://chatgpt.com/share/67c7e60f-6d68-8004-9d83-8aaa2f78f6d1> |
| DeepSeek | **0**  <https://chat.deepseek.com/a/chat/s/c9233abf-a921-43d7-afab-04196646bb1a> | **3**  <https://chat.deepseek.com/a/chat/s/c9233abf-a921-43d7-afab-04196646bb1a> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/03c98e6c3536> | **6**  <https://g.co/gemini/share/03c98e6c3536> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6800202e-3388-8006-9e41-e7ec1e90d024> | **4**  <https://chatgpt.com/share/6800202e-3388-8006-9e41-e7ec1e90d024> |
| Accounting mentor | **0**  <https://chatgpt.com/share/67fc1d8d-6364-8006-ad02-dc9ca06af453> | **5**  <https://chatgpt.com/share/67fc1d8d-6364-8006-ad02-dc9ca06af453> |
| Gemini 2.5 Pro Experimental | **0**  <https://docs.google.com/document/d/13HlbJaqeJMzQE4JWiVyJMR9V3nxxmnXgRlnUL77A2Ko/edit?tab=t.0> | **4**  <https://docs.google.com/document/d/1qBtr-0PcwgFlzenI1_p_ItJvdaHjRia64frIwfsGT2Y/edit?tab=t.0> |

1. Voting

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **11** | **13** | **15** | **16** | **19** | **20** | **# Fails** |
| **Wolfram (Q generator)** | **B** | **B** | **A** | **C** | **C** | **A** | **3** |
| **Free GPT-4o** | **A** | **B** | **B** | **A** | **C** | **A** | **0** |
| **DeepSeek** | **A/D** | **B** | **B** | **A** | **C** | **A** | **0** |
| **Geminai Flash** | **A/D** | **B** | **B** | **A** | **A** | **A** | **1** |
| **Paid GPT-4o** | **A** | **A** | **B** | **A** | **C** | **A** | **1** |
| **Accounting mentor** | **A** | **A** | **B** | **A** | **C** | **B** | **2** |
| **Geminai 2 Pro** | **A/D** | **A** | **B** | **A** | **C** | **A** | **1** |
| **Answer after voting** | **A** | **B** | **B** | **A** | **C** | **A** | **--** |

Questions 13, 20 (Accounting different understandings)

Question 11 has duplicate choices

# Generate hard 20-MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1UG9IL06KOIPIGPsUBfz8IU7dXhZPTvxpYXLiX7jdC3U/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **3**  <https://chatgpt.com/share/680ba9e4-4ab4-8007-b537-4107c52967b3> | **4**  <https://chatgpt.com/share/680ba9e4-4ab4-8007-b537-4107c52967b3> |
| Gemini 2 Flash | **3**  <https://g.co/gemini/share/e523adbd9f87> | **4**  <https://g.co/gemini/share/e523adbd9f87> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **4**  <https://chatgpt.com/share/6812293d-c3b8-800f-bd9f-8a9b3a397635> | **6**  <https://chatgpt.com/share/6812293d-c3b8-800f-bd9f-8a9b3a397635> |
| Wolfram 14.1 | **5**  <https://chatgpt.com/share/68126804-bc48-800f-ba6f-db84bfcf5e3f> | **3**  <https://chatgpt.com/share/68126804-bc48-800f-ba6f-db84bfcf5e3f> |
| Accounting mentor | **3**  <https://chatgpt.com/share/68127593-ff6c-800f-9035-7a078f5454a6> | **4**  <https://chatgpt.com/share/68127593-ff6c-800f-9035-7a078f5454a6> |
| Gemini 2.5 Pro Experimental | **5**  <https://g.co/gemini/share/a420252e7511> | **4**  <https://g.co/gemini/share/a420252e7511> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **1c** | **5c** | **6c** | **7c** | **9c** | **10c** | **2m** | **5m** | **6m** | **7m** | **8m** | **9m** | **10m** | **# Fails** |
| **DeepSeek (Q generator)** | **b** | **c** | **a** | **a** | **d** | **c** | **b** | **c** | **d** | **c** | **a** | **b** | **a** | **8** |
| **Free GPT-4o** | **d** | **c** | **a** | **a** | **c** | **a** | **c** | **c** | **none** | **d** | **a** | **a** | **a** | **1** |
| **Geminai Flash** | **b** | **a** | **a** | **a** | **c** | **a** | **c** | **c** | **C (Approx)** | **d** | **a** | **a** | **a** | **1** |
| **Paid GPT-4o** | **d** | **a** | **a** | **a** | **c** | **a** | **b** | **a** | **C (Approx)** | **d** | **b** | **a** | **c** | **4** |
| **Wolfram 14.1** | **d** | **a** | **d** | **a** | **c** | **a** | **c** | **c** | **none** | **d** | **a** | **b** | **a** | **2** |
| **Acc. mentor** | **d** | **b** | **a** | **a** | **c** | **c** | **c** | **c** | **none** | **d** | **a** | **a** | **a** | **2** |
| **Geminai 2 Pro** | **d** | **b** | **a** | **d** | **c** | **a** | **c** | **c** | **C (Approx)** | **d** | **a** | **a** | **a** | **2** |
| **Answer after voting** | **D** | **A** | **A** | **A** | **C** | **A** | **C** | **C** | **None** | **D** | **A** | **A** | **A** | **--** |

Questions 10 (Accounting different understandings)

**Chapter 15: Long Term Liabilities**

# Generate a 40 MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a28393-caf8-8006-9cbd-b6251dac6e16>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-free | **0**  https://chatgpt.com/share/68169341-2a34-8002-9a26-34aae06e18bb | **4**  <https://chatgpt.com/share/68169341-2a34-8002-9a26-34aae06e18bb> |
| DeepSeek | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **5**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/b3154fbaa124> | **4**  <https://g.co/gemini/share/b3154fbaa124> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/68129bca-04e4-800f-9f83-f9fd6aacdc85> | **5**  <https://chatgpt.com/share/68129bca-04e4-800f-9f83-f9fd6aacdc85> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/681b89e8-313c-8006-8f5a-4e17d2bab099> | **3**  [**https://chatgpt.com/share/681b89e8-313c-8006-8f5a-4e17d2bab099**](https://chatgpt.com/share/681b89e8-313c-8006-8f5a-4e17d2bab099) |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/2ab7c0b116cd> | **4**  <https://g.co/gemini/share/2ab7c0b116cd> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **22** | **23** | **24** | **29** | **30** | **35** | **36** | **37** | **39** | **# Fail** |
| **Accounting mentor (Q generator)** | **c** | **c** | **b** | **c** | **a** | **c** | **a** | **b** | **a** | **3** |
| **Free GPT-4o** | **b** | **d** | **b** | **c** | **a** | **None** | **none** | **a** | **a** | **1** |
| **DeepSeek** | **b** | **d** | **b** | **c** | **a** | **b** | **b** | **a** | **a** | **2** |
| **Gemini Flash** | **b** | **d** | **b** | **c** | **a** | **a (Approx)** | **b** | **b (Approx)** | **a** | **1** |
| **Paid GPT-4o** | **b** | **c** | **c** | **d** | **c** | **c (Approx)** | **b (Approx)** | **b (Approx)** | **a** | **4** |
| **Wolfram 14.1** | **c** | **c** | **b** | **c** | **a** | **c (Approx)** | **b (Approx)** | **c (Approx)** | **b** | **4** |
| **Gemini 2.5 Pro** | **b** | **d** | **b** | **c** | **a** | **c (Approx)** | **b** | **c** | **a** | **1** |
| **Answer after voting** | **B** | **D** | **B** | **C** | **A** | **None** | **None** | **None** | **A** | **--** |

* Some questions are ambiguous so the LLMs make assumptions.
* In question 37, the LLMs take different decisions about the evaluated value is it gain or loss

# Generate a 40 MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed here

<https://chatgpt.com/share/67a281e8-6104-8006-92e3-f1f4d3c5a101>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/68015b7a-5e04-8007-8259-f32a06512db5> | **3**  <https://chatgpt.com/share/68015b7a-5e04-8007-8259-f32a06512db5> |
| DeepSeek | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **4**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/180a567b29a9> | **4**  <https://g.co/gemini/share/180a567b29a9> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6812ae6d-4474-800f-8ed5-24ca5d02d9a4> | **4**  <https://chatgpt.com/share/6812ae6d-4474-800f-8ed5-24ca5d02d9a4> |
| Accounting mentor | **0**  <https://chatgpt.com/share/6812b3d7-1a30-800f-b79e-d0ca786cfc2e> | **2**  <https://chatgpt.com/share/6812b3d7-1a30-800f-b79e-d0ca786cfc2e> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/88ab93bdb86d> | **4**  <https://g.co/gemini/share/88ab93bdb86d> |

1. Voting

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **2m** | **3m** | **5m** | **10m** | **14m** | **# of wrong** |
| **Wolfram (Q generator)** | **b** | **d** | **c** | **c** | **b** | **4** |
| **Free GPT-4o** | **b** | **d** | **d** | **c** | **c** | **5** |
| **DeepSeek** | **a** | **a** | **a** | **b** | **b** | **0** |
| **Geminai Flash** | **a** | **a** | **a** | **b** | **b** | **0** |
| **Paid GPT-4o** | **a** | **a** | **a** | **b** | **b** | **0** |
| **Accounting mentor** | **b** | **d** | **d** | **c** | **c** | **5** |
| **Gemini 2 Pro** | **a** | **a** | **a** | **b** | **b** | **0** |
| **Answer after voting** | **A** | **A** | **A** | **B** | **B** | **--** |

some LLMs make right calculations but wrong choices

Generate a 40 MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1H-EGwORTp9YNPfceYSw--n_71EZgXhbgOI9zCeCQOjY/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1> | **0**  https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1 |
| Gemini 2 Flash | **1**  [**https://g.co/gemini/share/128e1c00cce7**](https://g.co/gemini/share/128e1c00cce7) | **0**  <https://g.co/gemini/share/128e1c00cce7> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6813f7cd-9ec0-800f-aa99-03f071a93341> | **0**  <https://chatgpt.com/share/6813f7cd-9ec0-800f-aa99-03f071a93341> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/68126dff-d39c-800f-a0d9-7bd462248678> | **0**  <https://chatgpt.com/share/68126dff-d39c-800f-a0d9-7bd462248678> |
| Accounting mentor | **0**  <https://chatgpt.com/share/68137f0c-c044-800f-a51c-bb4a10d342d9> | **0**  <https://chatgpt.com/share/68137f0c-c044-800f-a51c-bb4a10d342d9> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/4c372fd17cba> | **0**  <https://g.co/gemini/share/4c372fd17cba> |

1. Voting

|  |  |  |
| --- | --- | --- |
| **LLM / Q.** | **7c** | **# Fails** |
| **DeepSeek (Q generator)** | **b** | **0** |
| **Free GPT-4o** | **b** | **0** |
| **Geminai Flash** | **d** | **1** |
| **Paid GPT-4o** | **b** | **0** |
| **Wolfram 14.1** | **b** | **0** |
| **Acc. mentor** | **b** | **0** |
| **Geminai 2 Pro** | **b** | **0** |
| **Answer after voting** | **b** | **--** |

# Generate hard 20-MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a37da5-a9bc-8006-8c63-52c5ce3dade5>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/68169341-2a34-8002-9a26-34aae06e18bb> | **0**  <https://chatgpt.com/share/68169341-2a34-8002-9a26-34aae06e18bb> |
| DeepSeek | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash Thinking Experimental | **0**  <https://g.co/gemini/share/b3154fbaa124> | **0**  <https://g.co/gemini/share/b3154fbaa124> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/68129bca-04e4-800f-9f83-f9fd6aacdc85> | **0**  <https://chatgpt.com/share/68129bca-04e4-800f-9f83-f9fd6aacdc85> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/681b8849-c87c-8006-a1fc-082af76869af> | **0**  [**https://chatgpt.com/share/681b8849-c87c-8006-a1fc-082af76869af**](https://chatgpt.com/share/681b8849-c87c-8006-a1fc-082af76869af) |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/2ab7c0b116cd> | **0**  <https://g.co/gemini/share/2ab7c0b116cd> |

1. Voting

|  |  |
| --- | --- |
| **LLM / Q.** | **# Fails** |
| **Accounting mentor (Q generator)** | **0** |
| **Free GPT-4o** | **0** |
| **DeepSeek** | **0** |
| **Geminai Flash** | **0** |
| **Paid GPT-4o** | **0** |
| **Wolfram 14.1** | **0** |
| **Geminai 2.5 Pro** | **0** |
| **Answer after voting** | **--** |

# Generate hard 20-MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a50a37-f490-8006-b8ac-3d400667ce98>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  https://chatgpt.com/share/6816833b-fb5c-8002-a90b-024cf0a5cdfe | **7**  <https://chatgpt.com/share/6816833b-fb5c-8002-a90b-024cf0a5cdfe> |
| DeepSeek | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **5**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/68531b796c7d> | **6**  <https://g.co/gemini/share/68531b796c7d> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6812ae6d-4474-800f-8ed5-24ca5d02d9a4> | **6**  <https://chatgpt.com/share/6812ae6d-4474-800f-8ed5-24ca5d02d9a4> |
| Accounting mentor | **0**  <https://chatgpt.com/share/6812b3d7-1a30-800f-b79e-d0ca786cfc2e> | **5**  <https://chatgpt.com/share/6812b3d7-1a30-800f-b79e-d0ca786cfc2e> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/88ab93bdb86d> | **7**  <https://g.co/gemini/share/88ab93bdb86d> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **20** | **# Fails** |
| **Wolfram (Q generator)** | **b** | **b** | **c** | **b** | **a** | **c** | **b** | **b** | **d** | **7** |
| **Free GPT-4o** | **a** | **c** | **none** | **a** | **a** | **c** | **b (Approx)** | **d (Approx)** | **a (Approx)** | **2** |
| **DeepSeek** | **a** | **b (Approx)** | **d (Approx)** | **a (Approx)** | **c** | **c** | **b (Approx)** | **a (Approx)** | **d (Approx)** | **6** |
| **Gemini Flash** | **a** | **b** | **b (Approx)** | **a** | **a** | **c** | **a** | **None** | **b** | **2** |
| **Paid GPT-4o** | **a** | **b (Approx)** | **d (Approx)** | **a** | **a** | **c** | **a (Approx)** | **d (Approx)** | **None** | **2** |
| **Accounting mentor** | **b** | **c** | **c** | **b** | **c** | **b** | **c** | **b** | **c** | **9** |
| **Gemini 2 Pro** | **a** | **none** | **none** | **none** | **a** | **c** | **d (Approx)** | **d (Approx)** | **b (Approx)** | **3** |
| **Answer after voting** | **A** | **(None)** | **B (None)** | **A (None)** | **A** | **C** | **A (None)** | **D (None)** | **A (None)** | **--** |

* Questions are too complicated, and the correct answer isn’t listed in the choices.
* Each LLM uses different methods in the solutions, and this leads to different answers
* In question 20, Free Gpt4o, DeepSeek, Paid Gpt 4o, and Gemini 2.5 pro got the same answer, but they have conflicts in the choices because the correct answer is not listed in the choices
* Questions 12,13, 14, 17, 18, 20 makes the voting on the calculated answer not the chosen option

# Generate hard 20-MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1pvez-i9nWISTeOMh0hWUvjHOkMHhzW-RRp2E5tpiJYY/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1> | **4** <https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af14> |
| Gemini 2 Flash | **0**  [**https://g.co/gemini/share/128e1c00cce7**](https://g.co/gemini/share/128e1c00cce7) | **4**  <https://g.co/gemini/share/128e1c00cce7> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **2**  <https://chatgpt.com/share/6813f7cd-9ec0-800f-aa99-03f071a93341> | **6**  <https://chatgpt.com/share/6813f7cd-9ec0-800f-aa99-03f071a93341> |
| Wolfram 14.1 | **2**  <https://chatgpt.com/share/68126dff-d39c-800f-a0d9-7bd462248678> | **4**  <https://chatgpt.com/share/68126dff-d39c-800f-a0d9-7bd462248678> |
| Accounting mentor | **2**  <https://chatgpt.com/share/68137f0c-c044-800f-a51c-bb4a10d342d9> | **5**  <https://chatgpt.com/share/68137f0c-c044-800f-a51c-bb4a10d342d9> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/8355a1be01ff> | **5**  <https://g.co/gemini/share/8355a1be01ff> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **4c** | **10c** | **1m** | **2m** | **3m** | **5m** | **7m** | **9m** | **10m** | **# Fails** |
| **DeepSeek (Q generator)** | **c** | **a** | **b** | **A (Approx)** | **a** | **b** | **d** | **b** | **b** | **5** |
| **Free GPT-4o** | **a** | **a** | **b** | **b** | **c** | **b** | **b** | **d** | **b** | **2** |
| **Gemini 2.5 Flash** | **c** | **a** | **b** | **A (Approx)** | **c** | **b** | **b** | **none** | **D (Approx)** | **1** |
| **Paid GPT-4o** | **a** | **d** | **d** | **b** | **c** | **a** | **b** | **b** | **D (Approx)** | **4** |
| **Wolfram 14.1** | **a** | **d** | **b** | **A (Approx)** | **c** | **b** | **b** | **none** | **D (Approx)** | **1** |
| **Acc. mentor** | **a** | **d** | **b** | **c** | **c** | **b** | **b** | **none** | **D (Approx)** | **2** |
| **Gemini 2.5 Pro** | **c** | **a** | **b** | **none** | **c** | **b** | **b** | **none** | **D (Approx)** | **1** |
| **Answer after voting** | **A** | **A** | **B** | **None** | **C** | **B** | **B** | **None** | **None** | **--** |

* In questions 3, 7; DeepSeek changes the right answer after solving the question. Somehow, it modifies the question itself to match his wrong answer

**Chapter 14: Corporates**

# Generate a 40 MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a28592-1c50-8006-b3d5-fa54683fa177>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o free | **0**  https://chatgpt.com/share/681695e8-3e70-8002-9c58-f9a6831a4800 | **2**  <https://chatgpt.com/share/681695e8-3e70-8002-9c58-f9a6831a4800> |
| DeepSeek | **2**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **2**  https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J\_X9ZdMxs/edit?usp=sharing |
| Gemini 2 Flash Thinking Experimental | **1**  <https://g.co/gemini/share/b3154fbaa124> | **2**  <https://g.co/gemini/share/b3154fbaa124> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6812a171-2fe4-800f-893b-7d4f947de036> | **2**  <https://chatgpt.com/share/6812a171-2fe4-800f-893b-7d4f947de036> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/68129153-4d90-800f-b431-dff61b704e32> | **2**  <https://chatgpt.com/share/68129153-4d90-800f-b431-dff61b704e32> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/0ceed17829f7> | **2**  <https://g.co/gemini/share/0ceed17829f7> |

1. Voting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **12** | **19** | **29** | **33** | **# Fails** |
| **Accounting mentor (Q generator)** | **c** | **a** | **b** | **c** | **2** |
| **Free GPT-4o** | **c** | **a** | **a** | **d** | **0** |
| **DeepSeek** | **b/c** | **b** | **a** | **d** | **2** |
| **Geminai Flash** | **b** | **a** | **a** | **d** | **1** |
| **Paid GPT-4o** | **c** | **a** | **a** | **d** | **0** |
| **Wolfram 14.1** | **c** | **a** | **a** | **d** | **0** |
| **Geminai 2 Pro** | **c** | **a** | **a** | **d** | **0** |
| **Answer after voting** | **C** | **A** | **A** | **D** | **--** |

# Generate a 40 MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed here

<https://chatgpt.com/share/67a28965-0c5c-8006-9256-4123c6abb94e>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/681b8625-76d0-8004-8fb9-82f0daa7b6ef> | **0**  <https://chatgpt.com/share/681b8625-76d0-8004-8fb9-82f0daa7b6ef> |
| DeepSeek | **3**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **0**  https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J\_X9ZdMxs/edit?usp=sharing |
| Gemini 2 Flash Thinking Experimental | **1**  <https://g.co/gemini/share/f3ecd60aadbb> | **0**  <https://g.co/gemini/share/f3ecd60aadbb> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/6813febf-7ff0-800f-92c7-ac1cbf59e442> | **0**  https://chatgpt.com/share/6813febf-7ff0-800f-92c7-ac1cbf59e442 |
| Accounting mentor | **0**  <https://chatgpt.com/share/6814073d-9a24-800f-9b9b-50673574acaa> | **0**  <https://chatgpt.com/share/6814073d-9a24-800f-9b9b-50673574acaa> |
| Gemini 2.5 Pro Experimental | **1**  <https://g.co/gemini/share/f1d0984741a7> | **0**  <https://g.co/gemini/share/f1d0984741a7> |

1. Voting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **2** | **7** | **13** | **17** | **# Fails** |
| **Wolfram (Q generator)** | **a** | **a** | **d** | **d** | **0** |
| **Free GPT-4o** | **a** | **b** | **d** | **d** | **1** |
| **DeepSeek** | **d** | **b** | **b** | **d** | **3** |
| **Geminai Flash** | **a** | **a** | **b** | **d** | **1** |
| **Paid GPT-4o** | **a** | **b** | **d** | **d** | **1** |
| **Accounting mentor** | **a** | **a** | **d** | **d** | **0** |
| **Geminai 2 Pro** | **a** | **a** | **d** | **b** | **1** |
| **Answer after voting** | **A** | **A** | **D** | **D** | **--** |

# Generate a 40 MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1frB8_xRKNoW4sRQYZ-27iuf_cN0d45gdUL5FPlA2liI/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1> | **1**  https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1 |
| Gemini 2 Flash | **1**  [**https://g.co/gemini/share/4f1b3f4ff280**](https://g.co/gemini/share/4f1b3f4ff280) | **2**  <https://g.co/gemini/share/4f1b3f4ff280> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/68140ce0-358c-800f-bd1d-3cc1fbf8021a> | **3**  <https://chatgpt.com/share/68140ce0-358c-800f-bd1d-3cc1fbf8021a> |
| Wolfram 14.1 | **0**  <https://chatgpt.com/share/681295e4-ea80-800f-a9fc-d167b5b1334f> | **1**  <https://chatgpt.com/share/681295e4-ea80-800f-a9fc-d167b5b1334f> |
| Accounting mentor | **0**  <https://chatgpt.com/share/68139307-63ac-800f-a9b7-1fad935120ee> | **1**  <https://chatgpt.com/share/68139307-63ac-800f-a9b7-1fad935120ee> |
| Gemini 2.5 Pro Experimental | **1**  <https://g.co/gemini/share/8fb530e1a093> | **2**  <https://g.co/gemini/share/8fb530e1a093> |

1. Voting

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **10c** | **12c** | **3m** | **6m** | **14m** | **19m** | **# Fails** |
| **DeepSeek (Q generator)** | **b** | **b** | **c** | **b** | **b** | **b** | **1** |
| **Free GPT-4o** | **b** | **b** | **a** | **b** | **b** | **b** | **0** |
| **Geminai Flash** | **b** | **a/b** | **a** | **a** | **b** | **b** | **2** |
| **Paid GPT-4o** | **b** | **b** | **a** | **b** | **c** | **d** | **2** |
| **Wolfram 14.1** | **b** | **b** | **a** | **b** | **b** | **b** | **0** |
| **Acc. mentor** | **b** | **b** | **a** | **b** | **b** | **b** | **0** |
| **Geminai 2 Pro** | **d** | **b** | **a** | **a** | **b** | **b** | **2** |
| **Answer after voting** | **B** | **B** | **A** | **B** | **B** | **B** | **--** |

# Generate hard 20-MCQs-quiz from “Accounting mentor”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a50d9d-890c-8006-9f0e-8f6f2184f153>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/681695e8-3e70-8002-9c58-f9a6831a4800> | **4**  <https://chatgpt.com/share/681695e8-3e70-8002-9c58-f9a6831a4800> |
| DeepSeek | **1**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **6**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash Thinking Experimental | **0**  <https://g.co/gemini/share/bb4a9d34e2e6> | **5**  <https://g.co/gemini/share/b3154fbaa124> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **1**  <https://chatgpt.com/share/6812a171-2fe4-800f-893b-7d4f947de036> | **5**  <https://chatgpt.com/share/6812a171-2fe4-800f-893b-7d4f947de036> |
| Wolfram 14.1 | **1**  <https://chatgpt.com/share/681b8519-027c-8006-86e5-89f1f7249da6> | **6**  <https://chatgpt.com/share/681b8519-027c-8006-86e5-89f1f7249da6> |
| Gemini 2.5 Pro Experimental | **1**  <https://g.co/gemini/share/0ceed17829f7> | **5**  <https://g.co/gemini/share/0ceed17829f7> |

1. Voting

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **2** | **12** | **14** | **15** | **17** | **18** | **19** | **# Fails** |
| **Accounting mentor (Q generator)** | **b** | **c** | **c** | **b** | **b** | **b** | **a** | **6** |
| **Free GPT-4o** | **a** | **b** | **None** | **c** | **d** | **d** | **c** | **0** |
| **DeepSeek** | **a** | **b** | **A (Approx)** | **c** | **d** | **d** | **c** | **0** |
| **Geminai Flash** | **b** | **b** | **c** | **c** | **d** | **d** | **c** | **1** |
| **Paid GPT-4o** | **a** | **b** | **c** | **c** | **d** | **d** | **c** | **0** |
| **Wolfram 14.1** | **a** | **b** | **None** | **c** | **d** | **d** | **c** | **0** |
| **Geminai 2 Pro** | **a** | **b** | **c** | **c** | **d** | **d** | **c** | **0** |
| **Answer after voting** | **A** | **B** | **(None)** | **C** | **D** | **D** | **C** | **--** |

* Question 14 should have answer of 33,000 but it isn’t in the choice list. So, each LLM modify the IBT term to choose an answer from the list.  
  So the question isn’t clear.

# Generate hard 20-MCQs-quiz from “Wolfram”

The prompt and the generated questions are listed in this file

<https://chatgpt.com/share/67a51237-4f3c-8006-83e5-1c3d3c37330c>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  [**https://chatgpt.com/share/681b83ad-41b4-8004-b059-6993181c0f1d**](https://chatgpt.com/share/681b83ad-41b4-8004-b059-6993181c0f1d) | **2**  <https://chatgpt.com/share/681b83ad-41b4-8004-b059-6993181c0f1d> |
| DeepSeek | **0**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> | **5**  <https://docs.google.com/document/d/1dOhfFC71di7uCxijbqNbqzfSrg1tITFax9J_X9ZdMxs/edit?usp=sharing> |
| Gemini 2 Flash Thinking Experimental | **0**  <https://g.co/gemini/share/1b9c303eb722> | **5**  <https://g.co/gemini/share/1b9c303eb722> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **0**  <https://chatgpt.com/share/6814033b-5a38-800f-ab5d-70d505ca852f> | **4**  <https://chatgpt.com/share/6814033b-5a38-800f-ab5d-70d505ca852f> |
| Accounting mentor | **0**  <https://chatgpt.com/share/6814073d-9a24-800f-9b9b-50673574acaa> | **5**  <https://chatgpt.com/share/6814073d-9a24-800f-9b9b-50673574acaa> |
| Gemini 2.5 Pro Experimental | **0**  <https://g.co/gemini/share/f1d0984741a7> | **5**  <https://g.co/gemini/share/f1d0984741a7> |

1. Voting

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **12** | **15** | **16** | **17** | **19** | **# Fails** |
| **Wolfram (Q generator)** | **c** | **c** | **c** | **c** | **b** | **4** |
| **Free GPT-4o** | **b** | **b** | **c** | **a** | **a** | **0** |
| **DeepSeek** | **b** | **b** | **a** | **a** | **a** | **1** |
| **Geminai Flash** | **b** | **b** | **B (Approx)** | **a** | **a** | **1** |
| **Paid GPT-4o** | **b** | **b** | **c** | **a** | **a** | **0** |
| **Accounting mentor** | **b** | **b** | **a** | **a** | **a** | **1** |
| **Geminai 2 Pro** | **b** | **b** | **none** | **a** | **a** | **1** |
| **Answer after voting** | **B** | **B** | **None (B)** | **A** | **A** | **--** |

Generate hard 20-MCQs-quiz from “DeepSeek”

The prompt and the generated questions are listed here

<https://docs.google.com/document/d/1sKm99MPJOx0dcytuXNjOkyvYJZlrFS0JVgdB8MJLZJg/edit?usp=sharing>

LLMs voting

1. Free plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **2**  <https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1> | **3**  https://chatgpt.com/share/680bb2c5-d080-8007-9dc9-626484058af1 |
| Gemini 2 Flash | **0**  <https://g.co/gemini/share/4f1b3f4ff280> | **4**  <https://g.co/gemini/share/4f1b3f4ff280> |

1. Paid plans

|  |  |  |
| --- | --- | --- |
| **LLM** | **Number of wrong answers** | |
| Concept based | Math based |
| GPT-4o | **2**  <https://chatgpt.com/share/68140ce0-358c-800f-bd1d-3cc1fbf8021a> | **3**  <https://chatgpt.com/share/68140ce0-358c-800f-bd1d-3cc1fbf8021a> |
| Wolfram 14.1 | **1**  <https://chatgpt.com/share/681295e4-ea80-800f-a9fc-d167b5b1334f> | **1**  <https://chatgpt.com/share/681295e4-ea80-800f-a9fc-d167b5b1334f> |
| Accounting mentor | **1**  <https://chatgpt.com/share/68139307-63ac-800f-a9b7-1fad935120ee> | **3**  <https://chatgpt.com/share/68139307-63ac-800f-a9b7-1fad935120ee> |
| Gemini 2.5 Pro Experimental | **2**  <https://g.co/gemini/share/9987731515e4> | **2**  <https://g.co/gemini/share/9987731515e4> |

1. Voting

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LLM / Q.** | **6c** | **8c** | **9c** | **4m** | **6m** | **7m** | **8m** | **9m** | **10m** | **# Fails** |
| **DeepSeek (Q generator)** | **a** | **b** | **b** | **b** | **a** | **a** | **b** | **b** | **b** | **3** |
| **Free GPT-4o** | **b** | **a** | **b** | **a** | **b** | **a** | **b** | **b** | **c** | **5** |
| **Gemini Flash** | **a** | **b** | **none** | **a** | **a** | **b** | **b** | **c** | **c** | **4** |
| **Paid GPT-4o** | **c** | **c** | **b** | **b** | **a** | **a** | **a** | **c** | **c** | **2** |
| **Wolfram 14.1** | **c** | **b** | **b** | **b** | **a** | **a** | **b** | **c** | **b** | **1** |
| **Acc. mentor** | **c** | **b** | **b** | **b** | **a** | **b** | **b** | **c** | **c** | **1** |
| **Gemini 2 Pro** | **c** | **c** | **b** | **b** | **a** | **b** | **b** | **c** | **b** | **3** |
| **Answer after voting** | **C** | **B** | **B** | **B** | **A** | **A** | **B** | **C** | **C** | **--** |

* The different answers in Questions 9, 10 are because of adding new terms in the law  
  This is accounting issue not related to the Math

Findings in the created questions

For MCQs from Wolfram and Accounting mentor:

* Duplicate answers in Math Questions
* Wrong solutions because of wrong math calculations but the accounting concepts are right