Submibission - response area – evaluation function I do short answer question

First row generation is based on MiniLlama2 (1.1b)

Second row generation is based on ChatGPT-4o

Finetuned MiniLlama2 still not completed

Extra testing note: Minilama2 keeps around 20-30 tokens per second generation speed running with T80, applicable on lower level GPU servers but might not available on AWS Lambda.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune."

* **Missing Planet**: Uranus
* **Generated Text**: "Correct: 4. Based on the passage above, Could you continue the story by describing the teacher's next question or instruction?"
*  **Prompt 1**: ✅ **True** – The student listed all planets except Uranus, so this answer is partially correct, but missing one planet.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune."

* **Missing Planet**: Earth
* **Generated Text**: "Can you generate a program that checks if the provided input is a valid list of planet names in the solar system?"
*  **Prompt 2**: ❌ **False** – The student’s answer is missing Earth, so it is incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Jupiter, Saturn, Uranus, Neptune."

* **Missing Planet**: Mars
* **Generated Text**: "Teacher: Yes, Mercury is a planet. Teacher: Yes, Venus is a planet. Teacher: Yes, Earth is a planet."
*  **Prompt 3**: ❌ **False** – The student’s answer is missing both Earth and Mars, so it is incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus."

* **Missing Planet**: Neptune
* **Generated Text**: "You have a list of planets in the solar system, and each one of them is a boolean decision. You should make a logical statement based on the given."
*  **Prompt 4**: ❌ **False** – The student’s answer is missing Uranus, making it incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Mars, Saturn, Uranus, Neptune."

* **Missing Planet**: Jupiter
* **Generated Text**: "Teacher's response: Based on the passage above, Can you summarize the information about a teacher who creates a boolean decision for a question about the planets?"
*  **Prompt 5**: ❌ **False** – The student’s answer is missing Jupiter, making it incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune."

* **Missing Planet**: Uranus
* **Generated Text**: "True or False? Teacher's response: False. Based on the passage above, Is it possible to make a program that determines the number of planets?"
*  **Prompt 6**: ✅ **True** – The student’s answer is correct, listing all planets except Uranus.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune."

* **Missing Planet**: Uranus
* **Generated Text**: "Correct answer: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune. Based on the passage above, What did the…"
*  **Prompt 7**: ✅ **True** – The student’s answer is correct, listing all planets except Uranus.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Jupiter, Saturn, Uranus, Neptune."

* **Missing Planet**: Mars
* **Generated Text**: "You make the correct answer. Then, student asks you about the color of the sky. You make the correct answer. Then, student asks you about the weight…"
*  **Prompt 8**: ❌ **False** – The answer is missing Mars, making it incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Jupiter, Saturn, Uranus, Neptune."

* **Missing Planet**: Mars
* **Generated Text**: "Based on the passage above, How does the author determine which planets are part of the solar system based on their size?"
*  **Prompt 9**: ❌ **False** – The answer is missing Mars, making it incorrect.

 **Prompt**: "You are a physics teacher that makes boolean decisions on student's response on question: 'list planets in the solar system? Student's response: Mercury, Venus, Earth, Jupiter, Saturn, Uranus, Neptune."

* **Missing Planet**: Mars
* **Generated Text**: "Teacher's response: I will consider only those students who can correctly answer 'Mercury', 'Venus', 'Earth', 'J…"
*  **Prompt 10**: ❌ **False** – The answer is missing Mars, making it incorrect.