在这一周的实验中，Ruihui和我一起学习and研究如何在我们的测试中使用jinja template以及API。最后我们成功的搭建了环境并且能够成功运行代码。这样我们就可以通过大量的访问CGPT以得到一些数据，而不是亲自一个个去问。

Run without any prompts:Is this formula satisfiable

Use Red to mark totally correct

Use Green to mark doubts

Question 3:

Atom Count: 19

CGPT shows this formula is ‘quite complex’, and it tries to simplify the formula by applying logical equivalences and simplification, but it cannot finish this work since the formula is quite long and challenging.

Question 9：

Atom Count: 17

CGPT constructs a truth table and gives wrong answer

Question 10:

Atom Count: 21

CGPT simplifies the formula and produces correct answer.

Question 12

Atom Count: 16

CGPT shows it is likely to be satisfiable but it cannot make sure.

Question 13

Atom Count: 21

CGPT simplifies the formula and produces the correct answer.

Question 16

Atom Count: 13

CGPT can produces correct answer rapidly.

Question 18

Atom Count: 31

CGPT can produces correct answer by evaluating all possible truth values, but it doesn’t show the truth table.

Question 19

Atom Count: 16

CGPT gives the process of simplify the formula, but the result is wrong(through truth table generator).

Question 20:

Atom Count: 23

CGPT assesses the satisfiability of each component separately and shows the process, and it can give the correct answer though it says ‘it is quite complex’.

Question 23

Atom Count: 19

CGPT cannot evaluate this.

Question 23

Atom Count: 20

CGPT produces wrong truth table.

Question 26:

Atom Count: 29

CGPT recommends to use a propositional logic solver to determine it, and it cannot do the evaluation.

Question 28:

Atom Count: 16

CGPT gives wrong truth table.

Question 29:

Atom Count: 11

CGPT applies the biconditional elimination to give correct answer.

Question 30:

Atom Count: 27

CGPT recommends to use a propositional logic solver to determine it, and it cannot do the evaluation.

1. For formulas with more than 10 atoms, GPT is prone to errors or cannot answer questions.

2. It seems that CGPT is better at simplify. In these complex formulas, CGPT can simplify the formulas to get the correct answer. This method may be more acceptable to CGPT. CGPT has trouble producing truth table according to these results. We are going to try to test this with prompts during the week.

Run with prompts:Let CGPT gives a set of atoms whose expression is T

Use Red to mark totally correct

Use Green to mark doubts

Q1.

Atom Count: 24

Correct

Q2.

Atom Count: 16

Correct

Q3.

Atom Count: 17

Correct

Q4.

Atom Count: 27

Correct

Q5

Atom Count: 28

Just simplify, no determine.

Q6

Atom Count: 31

Correct

Q7

Atom Count: 25

Correct

Q8

Atom Count: 29

Correct

Q9

Atom Count: 30

Wrong answer with judgement.

Q10

Atom Count: 29

Correct

Q11

Atom Count: 12

False, wrong combe of atoms.

Q12

Atom Count: 14

Briefly judgment not satisfiable, which is wrong.

Q13

Atom Count: 26

Same to Q12

Q14

Atom Count: 13

Correct

Q15

Atom Count: 15

Correct

Q16

Atom Count: 18

Correct

Q17

Atom Count: 25

Correct

Q18

Atom Count: 26

Correct

Q19

Atom Count: 24

Correct

Q20

Atom Count: 26

Correct answer, but when giving specific atoms, there is no T or F for atom r.

Q21

Atom Count: 14

Correct answer, but it gives a wrong truth table. The combo of atoms is incorrect.

Q22

Atom Count: 19

Correct answer. Incorrect combo of atoms.

Q23

Atom Count: 28

Correct

Q24

Atom Count: 27

Correct answer. Incorrect combo of atoms.

Q25

Atom Count: 25

Correct answer. Incorrect combo of atoms.

Q26

Atom Count: 22

Correct

Q27

Atom Count: 23

Correct but wrong truth table.

Q28

Atom Count: 18

Correct.

Q29

Atom Count: 17

Correct

Q30

Atom Count: 13

Correct but wrong truth table.

Correct rate18/30

Our final conclusion is

This prompt does help, but this method seems to determine whether it is satisfiable by internally generating a truth table.

As mentioned before, CGPT seems to have difficulty generating truth tables for complex expressions. We felt simplifying might be a better approach and planned to test it.