

# Artificial Intelligence

## Assignment 1 (Solution)

Matthias Horbach

[mhorbach@uni-koblenz.de](mailto:mhorbach@uni-koblenz.de)

Institute of Web Science and Technologies

Department of Computer Science

University of Koblenz-Landau

Submission until: 02.05.2020, 23:59

Tutorial on: 03.05.2020

## 1 Limits of current AIs

(6 Points)

Pick a chat bot of your choice. I would advise ChatGPT, because it makes this exercise most interesting, but any one will do. Find a question that the chat bot answers confidently but incorrectly. State both question and answer. Explain how you came up with your question, what the chat bot did wrong, and what you believe to be the reason why the chat bot failed.

**Sample solution:** One example in the lecture were Sally-Anne questions, where the AI has to distinguish between different people's state of mind. ChatGPT can answer most simple such problems. Once a third person is introduced, it forgets the focus of the problem and just explains the state of the world, not how Sally interprets it.

**Question:** *Sally turns the oven to 150 degrees. Two minutes later, she leaves the kitchen to answer the phone. While she is gone, Anne turns the oven up to 180 degrees. Just before the phone had rung, Pete had turned the oven off. When Sally comes back a few minutes later, what temperature does she expect the oven to be at?*

**Answer:** *If Pete turned off the oven just before the phone rang, then the oven should be close to 150 degrees when Sally comes back, assuming that the oven was not on for a long time before Anne turned it up to 180 degrees.*

*The amount of time that the oven was at 180 degrees depends on how long Anne left it on before Pete turned it off. However, if Pete turned off the oven quickly after Anne turned up the temperature, then the oven may have only been at 180 degrees for a short period of time.*

*Assuming that Sally was only gone for a few minutes, the oven may still be warm but not hot when she returns. It is important to note that leaving the oven unattended can be a fire hazard and is not recommended.*

ChatGPT is also not great when it comes to logical reasoning. I assumed that it would not be able to solve even problems on this first exercise sheet. I believe that the main problem is that there is no wiggle room when doing maths/logics. Either it's completely right or completely wrong, and tiny changes in the problem may have a huge impact on the solution. In this sense, truth tables are like a Winograd scheme for logics. Note that the final column below is incorrect. I also had to fix the line breaks in the table.

**Question:** *Please write me a LaTeX truth table for the formula  $\neg(\neg P \vee \neg\neg P)$ .*

**Answer:** *Sure! Here's the truth table for the formula  $\neg(\neg P \vee \neg\neg P)$ :*

$P$	$\neg P$	$\neg\neg P$	$\neg(\neg P \vee \neg\neg P)$
T	F	T	F
F	T	F	T

[...]

ChatGPT also does not understand how objects are arranged in space. I tried a classic question that many humans struggle with, concerning two books on a shelf. Books stand in such a way that the front of the first book is right next to the back of the second book, so the first page of the first book and the last page of the second book are separated by only two covers and no pages at all. I had actually expected the AI to “know” this problem from its training, but it is completely oblivious to the central argument that books in a shelf are stacked in reverse order.

**Question:** *The three volumes of “Lord of the Rings” stand on a book shelf. A book worm sitting on page one of the first book wants to eat its way to the last page of the second book. How many covers does it need to chew through?*

**Answer:** *The book worm would need to chew through 4 covers: the cover of the first book, the back cover of the first book, the front cover of the second book, and the back cover of the second book. The third volume is not relevant to the question as the book worm only wants to eat through the first two books.*

## 2 Modeling in propositional logic

(10 Points)

Assume the following propositions:

- $A$  = “Alice visits the party“
- $B$  = “Bob visits the party“
- $C$  = “Charlie visits the party“
- $D$  = “Daniel visits the party“

Formulate the following sentences in propositional logic over the signature  $\mathcal{At} = \{A, B, C, D\}$ :

- Alice visits the party and Bob doesn't.
- Daniel and Bob visit the party if and only if Charlie also visits the party.
- If Alice and Bob visit the party, then Charlie does as well—but only if Daniel is not visiting the party.
- Charlie visits the party if and only if not both Bob and Alice visit the party, or at least Daniel visits the party.
- If Alice visits the party, then Bob and Charlie do as well—if Alice does not visit the party, then Charlie and Daniel visit the party.

**Sample solution:**

- $A \wedge \neg B$
- $C \Leftrightarrow (D \wedge B)$
- $(A \wedge B \wedge \neg D) \Rightarrow C$
- $C \Leftrightarrow (\neg(B \wedge A) \vee D)$
- $(A \Rightarrow (B \wedge C)) \wedge (\neg A \Rightarrow (C \wedge D))$

### 3 Reasoning in propositional logic

(14 Points)

Let  $\text{At} = \{P, Q, R\}$  be a propositional signature. Determine which of the following formulas are valid and/or satisfiable and/or unsatisfiable. Justify your answer, e.g. using interpretations or equivalences.

- (1)  $(P \wedge Q) \Rightarrow (P \vee Q)$
- (2)  $(P \vee Q) \Rightarrow (P \wedge Q)$
- (3)  $\neg(P \wedge \neg\neg P)$
- (4)  $Q \Rightarrow \neg Q$
- (5)  $Q \wedge \neg Q$
- (6)  $\neg(\neg P \vee \neg\neg P)$
- (7)  $((P \Rightarrow Q) \wedge (\neg P \Rightarrow R)) \Rightarrow (Q \vee R)$

**Sample solution:** There are lots of ways to answer these questions. Here are a few:

- (1)  $(P \wedge Q) \Rightarrow (P \vee Q) \equiv \neg(P \wedge Q) \vee (P \vee Q) \equiv \neg P \vee \neg Q \vee P \vee Q \equiv \neg P \vee P \vee \neg Q \vee Q \equiv \top \vee \top \equiv \top$ .

So the formula is valid. Hence it is also satisfiable.

- (2) is satisfiable, because it is satisfied by  $PQR$ : Under this interpretation, the right hand side of the implication is true, thus the same holds for the whole implication. It is not valid, because it is not satisfied by  $\bar{P}QR$ : Under this interpretation, the right hand side of the implication is false while the left hand side is true, thus the whole implication is false.

- (3) The interpretation of  $Q$  and  $R$  is obviously not relevant. Partial truth table for  $P$ :

$P$	$\neg P$	$\neg\neg P$	$P \vee \neg\neg P$	$\neg(P \vee \neg\neg P)$
true	false	true	true	false
false	true	false	false	true

So the formula is satisfiable, because it is satisfied, e.g., by  $\bar{P}QR$ .

It is not valid, because it is not satisfied by  $PQR$ .

- (4)  $\text{Mod}(Q \Rightarrow \neg Q) = \text{Mod}(\neg Q \vee \neg Q) = \text{Mod}(\neg Q)$

So the formula is satisfied by  $P\bar{Q}R$ .

It is not valid, because it is not satisfied by  $PQR$ .

- (5)  $\text{Mod}(Q \wedge \neg Q) = \text{Mod}(Q) \cap \text{Mod}(\neg Q) = \text{Mod}(Q) \cap (\text{Int}_{\text{At}} \setminus \text{Mod}(Q)) = \emptyset$

So the formula is unsatisfiable.

- (6)  $\neg(\neg P \vee \neg\neg P) \equiv \neg\neg P \wedge \neg\neg\neg P \equiv P \wedge \neg P \equiv \perp$

So the formula is unsatisfiable.

- (7) For any interpretations making  $Q$  or  $R$  true, the right hand side of the implication is true, and thus so is the whole formula. The only other interpretations are  $P\bar{Q}\bar{R}$  and  $\bar{P}\bar{Q}\bar{R}$ . Both make the left hand side of the implication false and thus the whole formula true.

Since every interpretation is a model of the formula, the formula is valid. Hence it is also satisfiable.

## Important hints

- Always include all names of all group members that helped solving the excercises **on your PDF**. Only those will receive points for solving the excercises.
- By handing in this sheet, you confirm that you solved these excercises yourself. If the situation occurs that two groups have identical solutions, both groups will get zero points.
- Your SVN-Repositories can be accessed via

<https://svn.uni-koblenz.de/westteaching/ai23/>

followed by your group name. SVNs will be available from 28.04. onwards.

- **Format:** All solutions must be contained in PDF documents (including source code). Additionally, source code must be provided as plain files that are readable via a standard text editor.
- Please make sure that all your programs can be run without errors. Comments on your souce code will be in the annotated PDF that we create during excercise corrections.
- Do not use any mutated vowels or special characters in your source code. Also, do not use those or spaces in file names.