CSI5101 - Knowledge Representation
Winter 2024

Assignment 3 - Reasoning

Submitted by

Romario Vaz : 300308477
Saranya Krishnasami : 300321456

Index

Study 1 - Deductive reasoning

1. **Definition and importance for humans + Examples (5 points)**

   Deductive reasoning is the logical approach where one deduces a specific conclusion from a set of general ideas and observations (premises) (Bhandari, 2023). It is important to humans because it helps in analyzing complex situations, problem-solving, and making informed decisions from data (Miller & Tiwari, 2023).

   In general, deductive reasoning is not difficult for humans because people implicitly use deductive reasoning in their daily life. For instance, if the windows in your house are closed and you see and smell smoke coming out of your kitchen, you implicitly deduce that there is likely a fire in the kitchen. However, when involving complex premises and abstract concepts, or when there are several pieces of information that need to be processed together to draw a conclusion, or when there are many negative statements or double-negatives (or even triple-negatives), the process of deductive reasoning would become more challenging. It would be helpful if humans had some training in formal logic to make deductions.

   The following examples demonstrate deductive reasoning:

   Example 1 (from Practiceaptitudetests.com):

   Given the premises "All football players are fit and healthy" and "All famous sports players are football players," which of the following is the logical deduction?

   i. All football players are famous sports people

   ii. All famous people are fit and healthy

   iii. All famous sports players are fit and healthy

   iv. All fit and healthy people are football players

   v. All football players are men

   The logical answer is iii. To reach this conclusion, we need to use logical deduction. Since all famous sports players are football players, and since all football players are

fit and healthy, it stands to reason that all famous sports players are fit and healthy. The other choices cannot be logically deduced from the two premises. For this example, no specific domain knowledge, real-world knowledge, mathematical knowledge, processes or rules are required to be known. In fact, real-world knowledge would contradict the logical deduction because it is a well-known fact that not all famous sports players are football players. Further, although it is not required, a knowledge of predicate logic could help a reader arrive at the same conclusion that choice iii is correct.

Example 2 (from aptitude-test.com)

Mike finished ahead of Paul. Paul and Brian both finished before Liam. Owen did not finish last. Who was the last to finish?

The correct answer is Liam. The chain of reasoning is that since Mike finished before Paul, he was not last. Since both Paul and Brian finished before Liam, neither Paul nor Brian was last. It is stated that Owen did not finish last. By the process of elimination, only Liam remains in the list of candidates, and hence he must have been the last to finish.

The real world knowledge required for this question is an understanding of how races work, and therefore also the understanding of "finishing ahead" and "finishing before". No other specific training in formal logical deduction is required.

2. **Importance in AI research (past, present, future) (7 points)**

Past:

Deductive reasoning in AI traces back to the mid-1900s. John McCarthy's seminal work 'Programs with Common Sense' (McCarthy, 1959), laid the groundwork for logical AI, emphasizing the importance of applying mathematical logic to nonmathematical domains. This evolved into Automated Deduction/Automated Theorem Proving, emerged as a critical area within AI, became vital for AI, focusing on the development of systems capable of performing logical deductions. Such systems use formal languages and rules of inference, such as modus ponens and resolution, to derive conclusions from premises. However, for more non-trivial

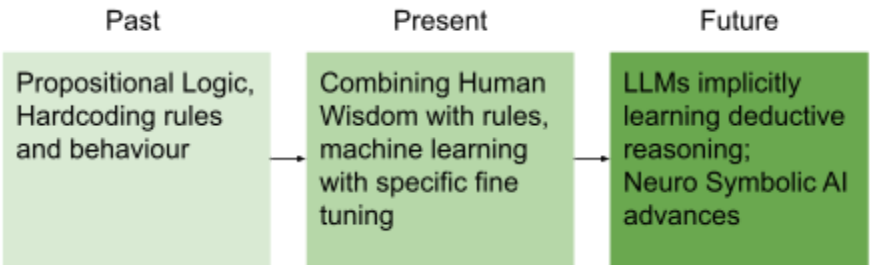
deductions, there was a combinatorial explosion and such deductions became too complex for AI systems of the time to handle (Davis, 2001).

Present:

Researchers at NYU and Google AI are advancing the field of deductive reasoning in AI, particularly through the exploration of large language models' (LLMs) capabilities in generalizing deductive reasoning tasks beyond their training examples. Their work focuses on enhancing LLMs' understanding of various deduction rules and assessing their ability to apply these rules in complex reasoning scenarios by fine-tuning with in-context learning (ICL) and chain-of-thought (CoT) examples (Shenwai, 2023). NEC is also making strides in the realm of deductive reasoning with AI, specifically through the development of Logical Thinking AI. This AI is designed to assist in decision-making processes by not only learning from numerical data but also incorporating human wisdom from sources like operation manuals. This approach enables the AI to provide logical reasoning for its suggestions, enhancing trust and understanding in its decisions (NEC, 2018).

Future:

We see the future of deductive reasoning in AI to focus on Neuro-Symbolic AI (combining learning capabilities of neural networks with the logical reasoning capabilities of symbolic AI) (Kamruzzaman et al., 2022) as well as research into devising large language models to implicitly possess sufficient deductive reasoning skills to not have to be specifically fine-tuned for deductive reasoning. Some unresolved problems involve struggles with ambiguities, uncertainty, and commonsense reasoning, and combining vision and text to form deductions (Zhang et al., 2024).

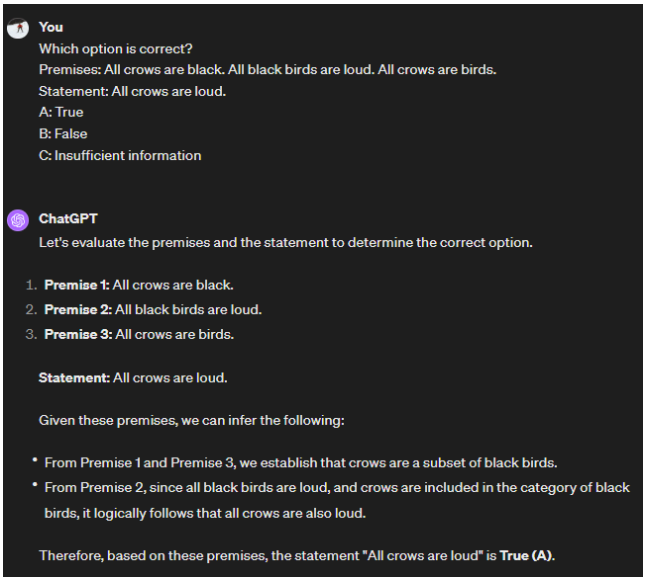


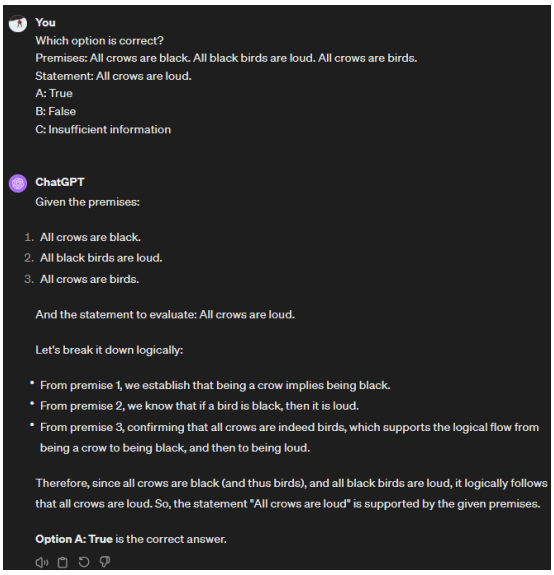
3. Importance in industry

Within AI-based products and services, deductive reasoning enables the product to make logical decisions based on rules and facts, as in mathematical theorem proving and legal reasoning, and can provide interpretability (and trust) based on logical deductions. The Coq tool uses deductive reasoning to prove mathematical theorems through interactive proof methods, decision algorithms, etc. (The Coq Proof Assistant, (n.d.)). There also exist medical expert systems that attempt to optimize medical decision making by observing symptom data, readings of vitals, etc. (Kattan, 2001).

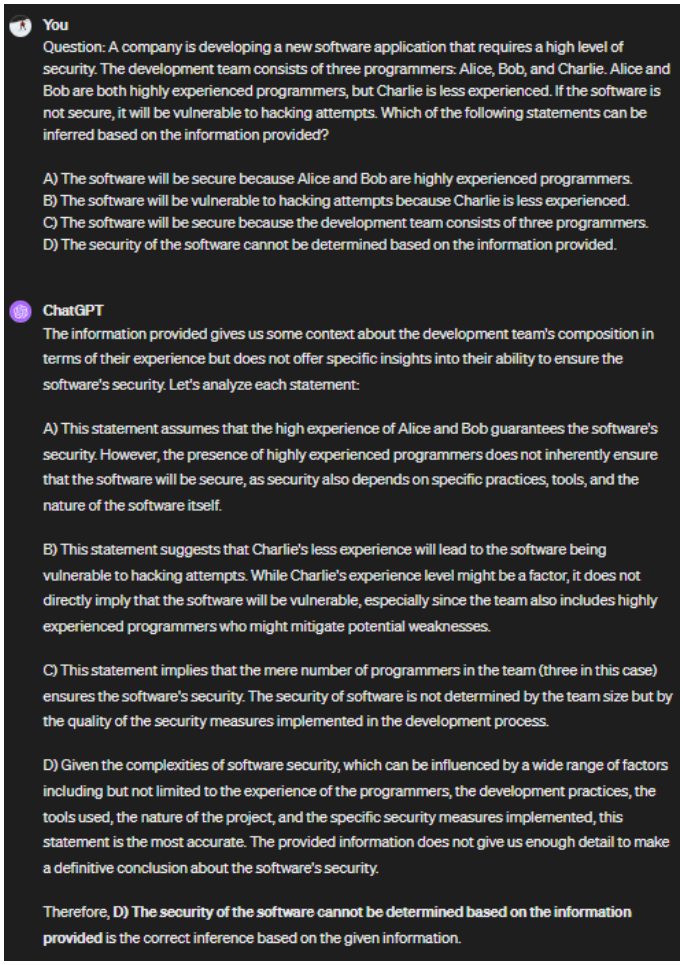
However, such tools are of somewhat limited use in industry. As previously mentioned, with more deductive rules such systems would face a combinatorial explosion and therefore be too complex and hard to scale. Moreover, current research is focused on data-driven machine learning techniques which have proven to be quite powerful and do not require explicit programming of several deductive rules. Another potential factor for the limited use in the industry is that deductive reasoning is relatively easy for humans to do by themselves, which mitigates the incentive to build a deductive reasoning system.

4. Current coverage of this reasoning in LLMs





Example 1 (from www.aptitude-test.com) (image above) is solvable with simple predicate logic. Both times, ChatGPT was able to answer the questions correctly and explain its reasoning in a human-understandable manner. The answers are fairly consistent and only differ in the phrasing.



You

Question: A company is developing a new software application that requires a high level of security. The development team consists of three programmers: Alice, Bob, and Charlie. Alice and Bob are both highly experienced programmers, but Charlie is less experienced. If the software is not secure, it will be vulnerable to hacking attempts. Which of the following statements can be inferred based on the information provided?

A) The software will be secure because Alice and Bob are highly experienced programmers.
B) The software will be vulnerable to hacking attempts because Charlie is less experienced.
C) The software will be secure because the development team consists of three programmers.
D) The security of the software cannot be determined based on the information provided.

ChatGPT

The information provided gives us some context about the development team's composition in terms of their experience but does not offer specific insights into their ability to ensure the software's security. Let's analyze each statement:

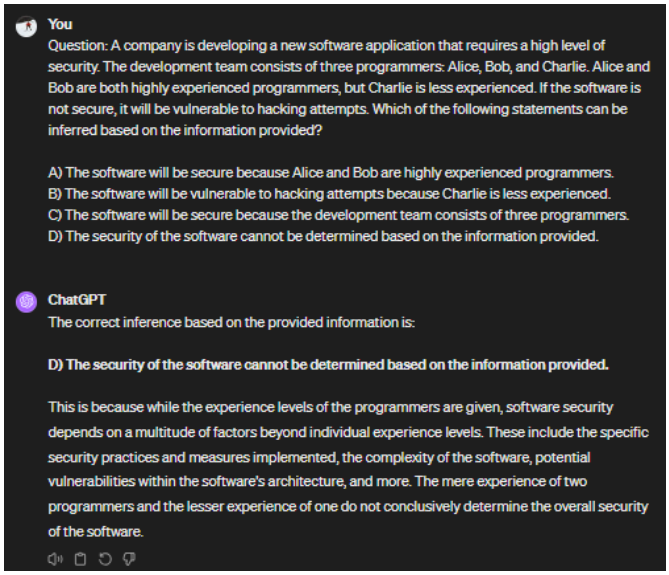
A) This statement assumes that the high experience of Alice and Bob guarantees the software's security. However, the presence of highly experienced programmers does not inherently ensure that the software will be secure, as security also depends on specific practices, tools, and the nature of the software itself.

B) This statement suggests that Charlie's less experience will lead to the software being vulnerable to hacking attempts. While Charlie's experience level might be a factor, it does not directly imply that the software will be vulnerable, especially since the team also includes highly experienced programmers who might mitigate potential weaknesses.

C) This statement implies that the mere number of programmers in the team (three in this case) ensures the software's security. The security of software is not determined by the team size but by the quality of the security measures implemented in the development process.

D) Given the complexities of software security, which can be influenced by a wide range of factors including but not limited to the experience of the programmers, the development practices, the tools used, the nature of the project, and the specific security measures implemented, this statement is the most accurate. The provided information does not give us enough detail to make a definitive conclusion about the software's security.

Therefore, **D) The security of the software cannot be determined based on the information provided** is the correct inference based on the given information.



You

Question: A company is developing a new software application that requires a high level of security. The development team consists of three programmers: Alice, Bob, and Charlie. Alice and Bob are both highly experienced programmers, but Charlie is less experienced. If the software is not secure, it will be vulnerable to hacking attempts. Which of the following statements can be inferred based on the information provided?

A) The software will be secure because Alice and Bob are highly experienced programmers.
B) The software will be vulnerable to hacking attempts because Charlie is less experienced.
C) The software will be secure because the development team consists of three programmers.
D) The security of the software cannot be determined based on the information provided.

ChatGPT

The correct inference based on the provided information is:

D) The security of the software cannot be determined based on the information provided.

This is because while the experience levels of the programmers are given, software security depends on a multitude of factors beyond individual experience levels. These include the specific security practices and measures implemented, the complexity of the software, potential vulnerabilities within the software's architecture, and more. The mere experience of two programmers and the lesser experience of one do not conclusively determine the overall security of the software.

Example 2 (From fintest) (refer to the images above) - ChatGPT correctly answers in both attempts. In the first attempt, it is very verbose and explains why each answer choice A-C is incorrect by accurately explaining that programming level and software security are two unrelated concepts and that there is not enough information to determine the security of the software. In the second attempt, the response is more succinct and it manages to explain how there is not enough information, without going into detail about why the other choices were wrong.

Given these responses, ChatGPT has not struggled at all with this level of deductive reasoning. Of course, more complex examples might cause it to be confused but with these test examples, at its current state, ChatGPT is definitely within a step towards "general AI". As a large language model, it was not trained specifically on deductive reasoning but rather on a large variety of texts from various domains, and it is impressive that it is able to deductively reason correctly.

References:

Bhandari, P. (2023, June 22). What is deductive reasoning? | explanation & examples. Scribbr. https://www.scribbr.com/methodology/deductive-reasoning/

Miller, J., & Tiwari, R. (2023, June 1). How do you use deductive and inductive reasoning to support your arguments?. Deductive and Inductive Reasoning: How to Support Your Arguments. https://www.linkedin.com/advice/3/how-do-you-use-deductive-inductive-reasoning-1e

Deductive reasoning tests: Practice Questions & Tips (2024). Practice Aptitude Tests. (2023, January 11). https://www.practiceaptitudetests.com/deductive-reasoning-tests/

Deductive reasoning test. Aptitude Test. (n.d.). https://www.aptitude-test.com/deductive-reasoning.html

McCarthy, John, 1959, “Programs with common sense”, in Proceedings of the Teddington Conference on the Mechanization of Thought Processes, London: Her Majesty’s Stationary Office, 75–91.

Davis, Martin (2001). "The Early History of Automated Deduction". Robinson & Voronkov 2001.

Shenwai, D. S. (2023, December 15). Researchers from NYU and Google Ai Explore Machine Learning’s frontiers in advanced deductive reasoning. MarkTechPost. https://www.marktechpost.com/2023/12/14/researchers-from-nyu-and-google-ai-explore-machine-learnings-frontiers-in-advanced-deductive-reasoning/

Logical thinking AI that interprets the result logically with reasons. NEC. (2018, December 12). https://www.nec.com/en/global/rd/technologies/201807/index.html

Kamruzzaman , Md. K. M., Zhoulabel, L. L., Aaron, A., & Hitzlerlabelauthor .[Department of Computer Science, P. (2022.). Neuro-Symbolic Artificial Intelligence. ar5iv. https://ar5iv.labs.arxiv.org/html/2105.05330

Zhang, Y., Bai, H., Zhang, R., Gu, J., Zhai, S., Susskind, J., & Jaitly, N. (2024). How Far Are We from Intelligent Visual Deductive Reasoning? arXiv [Cs.AI]. Retrieved from http://arxiv.org/abs/2403.04732

The coq proof assistant. Welcome! | The Coq Proof Assistant. (n.d.). https://coq.inria.fr/

Kattan, M. W. (2001). Expert Systems in Medicine. In N. J. Smelser & P. B. Baltes (Eds.), International Encyclopedia of the Social & Behavioral Sciences (pp. 5135–5139). doi:10.1016/B0-08-043076-7/00556-8

Deductive reasoning test: Free practice questions (2023). Fintest. (2023, March 13).
https://www.fintest.io/tests/deductive-reasoning-test/

Study 2 - Temporal reasoning

1. **Definition and importance for humans + Examples**

   The choice of reasoning for study 2 is Temporal reasoning. Temporal reasoning refers to the cognitive process of understanding and reasoning about time-related concepts, events, and sequences. It involves the ability to extract, assimilate, and manipulate temporal information to reconstruct a series of events for reasoning purposes (Gentner 2024).

   This cognitive skill is crucial for humans as it requires various higher-order cognitive processes such as planning, problem-solving, memory retrieval, and decision-making. Understanding temporal relationships allows individuals to organize events chronologically, predict future outcomes based on past experiences, and make informed choices (Chen 2007).

   For humans, temporal reasoning is significant as it is the basis of various cognitive functions and decision-making processes. Temporal reasoning involves understanding time-related concepts, events, and sequences. While basic skills develop naturally, complex tasks such as predicting outcomes or understanding intricate patterns can be challenging.

   Example 1: Evolution of Neural Network Models

   Arrange the following milestones in the evolution of neural network models in chronological order:

   A. Development of the Mark 1 Perceptron

   B. Introduction of backpropagation for artificial neural networks

   C. Publication of a functional learning algorithm for a multilayer Perceptron

   D. Creation of the Neocognitron

   E. Implementation of a deep network with eight layers trained using the group method

The correct order of the events A, C, D, B, E. To answer the question, domain knowledge is crucial to know the evolution of the neural network models in the order of the events. As well, for an expert this could also be a tacit knowledge where the person has worked on the models as models evolved and has the sequence of evolution as implicit knowledge gained through his work experience.

Example 2: (from https://www.examsbook.com/)

The chairman entered the Assembly Hall at 10 min. before 12.30 hours to conduct the interview. The chairman was 20 min. earlier than the managing director. The chairman was 30 min. late as per the time fixed. What was the time fixed for the interview?

(A) 12:50

(B) 12:10

(C) 12:40

(D) 12:00

The answer is (B) 12:10, to effectively solve the problem at hand, a combination of procedural, declarative, and implicit knowledge is essential. Procedural knowledge is necessary for performing calculations and operations related to time, such as subtracting or adding time intervals. Declarative knowledge encompasses understanding the facts and information provided in the problem statement, including the specific times involved and their relationships. Implicit knowledge plays a crucial role in intuitively grasping the significance of being late for an appointment and recognizing the importance of punctuality in professional contexts.

2. **Importance in AI research (past, present, future) (7 points)**

   Past: The study of temporal reasoning in AI dates back to the early days of AI research. One of the seminal works in this area is Allen's interval algebra, introduced in the 1980s, which formalized temporal relations between events. Early articles like "Reasoning About Time" by James Allen (1983) laid the groundwork for understanding temporal reasoning and its applications in AI.The representation of time and the ability to reason about temporal relationships have been essential components in the development of AI systems, enabling them to process and interpret time-related data effectively (Chittaro, 2000).

   Present: Temporal reasoning continues to hold a crucial role in present AI research, especially with the advent of large language models (LLMs) that excel in understanding and predicting time-sensitive contexts. Current AI research focuses on tasks like temporal relation extraction, temporal knowledge graph reasoning, and temporal question answering, leveraging advanced NLP methods to enhance temporal reasoning capabilities. The application of temporal logic in AI has enabled the formal specification and verification of real-time systems, reactive systems, and hybrid systems, showcasing the practical relevance of temporal reasoning in contemporary AI applications (Olex, 2021). By mastering temporal reasoning, AI systems can better interpret sequences of events, anticipate future outcomes, and adapt to changing circumstances, thus enhancing their effectiveness and

applicability across a wide range of domains including robotics, autonomous vehicles, natural language processing, and financial forecasting.

Future: In future AI research, the importance of temporal reasoning is expected to continue growing significantly. Researchers are likely to focus on advancing temporal reasoning capabilities to address more complex tasks that require multi-step reasoning and prediction based on future timestamps. There is a shift towards enhancing explainable temporal reasoning to provide clear explanations for predictions based on intricate temporal relationships, which will be crucial for improving the transparency and interpretability of AI systems. Additionally, the exploration of emerging trends in temporal representation and reasoning, along with its integration with other established areas like temporal databases and logic programming, will shape the future direction of AI research in temporal reasoning. This evolution underscores the ongoing relevance and expanding scope of temporal reasoning in shaping the future landscape of AI technologies (Chenhan 2023).

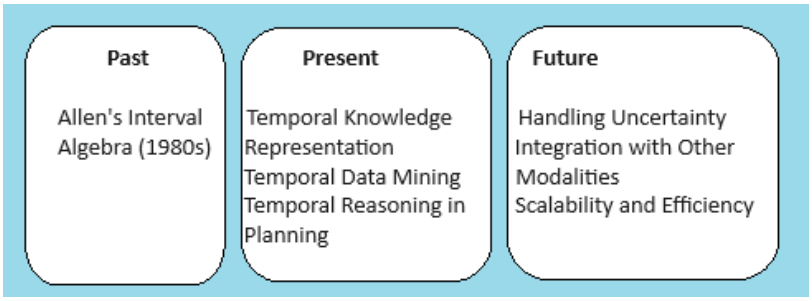


Figure 2.0 Key concepts regarding the past, present, and future trajectory of verbal reasoning.

3. **Importance in industry (4 points)**

Temporal reasoning is integral to numerous industries due to its capability to process and understand time-dependent data. Its importance within AI-based products, technologies, and services lies in its ability to:

Personalization and Contextual Understanding: In sectors like retail and marketing, temporal reasoning helps AI systems understand customers' evolving preferences and behaviors over time, enabling personalized recommendations and targeted advertising campaigns.

Real-time Decision Making: In dynamic environments such as finance, healthcare, and transportation (Autonomous vehicles), temporal reasoning enables AI systems to make rapid decisions based on current temporal data, ensuring timely responses to changing conditions.

Technologies and products that incorporate temporal reasoning include,

Autonomous Vehicles and Robotics: Technologies like self-driving cars and industrial robots rely on temporal reasoning to perceive and navigate dynamic environments, predicting the movements of other vehicles or obstacles in real time.

Financial Trading Algorithms: Algorithmic trading systems incorporate temporal reasoning to analyze market data, identify patterns, and execute trades automatically based on predefined rules and strategies

Predictive Analytics Software: AI-driven predictive analytics platforms, including IBM Watson Analytics, SAS Predictive Analytics, and Google Cloud Prediction API, leverage temporal reasoning to forecast future trends and make data-driven decisions

There is some *limited use of temporal reasoning in certain industries*, it could be due to factors such as,

Data Availability and Quality: Industries with sparse or low-quality temporal data may find it challenging to leverage temporal reasoning effectively for decision-making or predictive analysis.

Complexity and Interpretability: Temporal reasoning algorithms may be computationally intensive or difficult to interpret, limiting their adoption in industries where simplicity and transparency are paramount.

4. **Current coverage of this reasoning in LLMs (4 points)**

Testing chatgpt with the example 1 on the evolution of neural networks, from figure 2.1, the tool was not able to rank the developments in the right order. When prompted to use the years to order them , still the chatbot was not able to rank,it it lacked mathematical reasoning to order the years in chronological order.

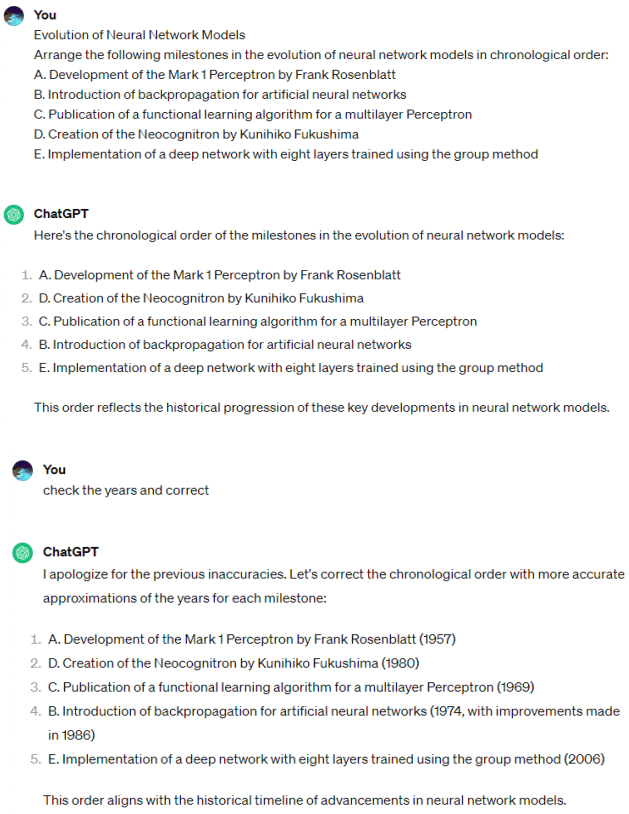


Figure 2.1 ChatGPT answering example 1 question on Evolution of NN

Tested the example 2 which is a temporal reasoning question that involves breaking down the question step by step and doing mathematical reasoning such as subtraction or adding time. From Figure 2.2, it's seen that on ChatGPT the answer given by the tool was option (A) 12:50 both the times, which is incorrect. The chatbot also provided the reasoning for the problem which revealed that the chatbot failed at the final mathematical reasoning step. Without being trained on the task, chatgpt did pretty well, able to understand the context, able to perform decent reasoning and it failed at the last step that shows that it may need more training to perform complex tasks. LLM are not good at doing mathematical reasoning compared to natural language processing tasks such as Summarization or question answering. Probably more training and even reinforcement learning to correct the mistakes would be the way forward. At the rate GPT learns, correcting its mistakes every step counts and it's a step towards AGI.

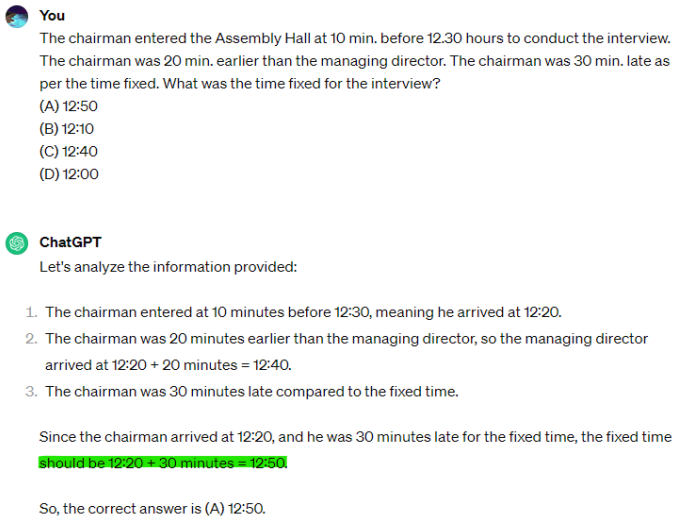


Figure 2.2 Chat GPT answering a Temporal reasoning question

References:

Gentner, D., & Boroditsky, L. (2024, February 19). The Development of Temporal Concepts: Linguistic Factors and Cognitive Processes. Frontiers in Psychology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6290033/

Chen, Y., & Wang, L. (2007, January 14). Temporal reasoning with medical data - A review with emphasis on medical natural language processing. Frontiers in Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7357407/

Chittaro, L., & Montanari, A. (2000). Temporal representation and reasoning in artificial intelligence: Issues and approaches. Annals of Mathematics and Artificial Intelligence, 28, 47–106. https://doi.org/10.1023/A:1018900105153

Olex, A & Bridget, T. (2021). Review of Temporal Reasoning in the Clinical Domain for Timeline Extraction: Where we are and where we need to be. https://www.sciencedirect.com/science/article/pii/S1532046421001131

Chenhan, Y, Qianqian, X, Jimin H, & Sophia A, Back to the Future: Towards Explainable Temporal Reasoning with Large Language Models. https://doi.org/10.48550/arXiv.2310.01074

Allen interval algebra - https://en.wikipedia.org/wiki/Allen%27s_interval_algebra

Study 3 - Verbal Reasoning

1. **Definition and importance for humans + Examples (5 points)**

The reasoning chosen for study 3 is Verbal reasoning. Verbal reasoning involves the ability to understand, evaluate and logically interpret information presented in written or spoken language. This type of reasoning is important to humans for several reasons including Communication, problem-solving, decision-making, education and learning.

Verbal reasoning can be considered relatively easy for humans, as it is a fundamental aspect of daily communication and cognitive function. However, honing and improving verbal reasoning skills may require some training, especially in more complex scenarios or for specific purposes such as standardized tests or academic pursuits. Few examples used to test verbal reasoning in humans are analogies, reading comprehension, sentence completion etc.

Example 1: (from Verbal reasoning practice tests)

Find the two statements that together prove: Sandy is a country music singer.

1: Sandy sings in a band.

2: Sandy likes country music the best.

3: Difficult music such as country and classical requires a strongly trained voice.

4: Sandy's band plays country songs.

5: Sandy has attended singing lessons for many years.

A: 1 & 3 B: 2 & 5 C: 3 & 5 D: 1 & 4 E: 3 & 4

The answer for this question is (D) 1 & 4, both statements logically follows that if Sandy sings in a band that performs country songs, then Sandy is indeed a country music singer. This conclusion is derived from the knowledge that singers in bands typically perform the genre of music associated with their band's repertoire . Knowledge required to solve this is World-Knowledge to understand the association between singers and the genres they perform.Specific Domain Knowledge, for familiarity with the characteristics of country

music and the typical activities of country music singers. Logical Reasoning, ability to analyze statements and draw conclusions based on the information provided. Also ability to understand and comprehend the question in the English language.

Example 2: (from Verbal reasoning practice tests)

In ancient Egypt, pyramid workers would say "*gengro yatju*" when wishing other pyramid workers "good building". Upon the pyramid workers' return, the children would ask "*Ern yatju gengro*?", which meant "Was building good"? The pyramid workers would either reply "*Ern gengro*" ("was good building") or "*Ern gangro yatju*!"

If gangro symbolises the opposite of gengro, what does "*Ern gangro yatju*" mean?

A: was bad building B: was hard building C: was fun building D: was cold building E: was happy building

For this question the answer is (A) was bad building. Understanding the ancient Egyptian phrases "gengro," "yatju," and "gangro" requires a blend of linguistic, semantic, and cultural knowledge. Linguistic expertise is essential for interpreting the meanings of these words based on vocabulary and grammar rules. Semantic knowledge aids in grasping the relationships between words, such as recognizing that "gangro" signifies the opposite of "gengro." Cultural insights into ancient Egyptian customs and language conventions provide context for interpreting these phrases accurately within their historical setting. Additionally, inferential reasoning skills play a crucial role in deducing the meaning of expressions like "Ern gangro yatju'' by logically inferring its opposite connotation to "Ern gengro."

2. **Importance in AI research (past, present, future)**

   Past: The study of verbal reasoning in AI can be traced back to the development of expert systems in the 1970s and 1980s, which aimed to mimic human expertise in specific domains through verbal interactions. Early articles such as "Elements of a Theory of Human Problem Solving" by Allen Newell and Herbert A. Simon (1972) and "The Language of Thought" by Jerry A. Fodor (1975) laid the groundwork for understanding the role of language and reasoning in human cognition.

Present: Several large research laboratories are interested in verbal reasoning in AI. For example, at the Knowledge Representation and Reasoning (KRR) conference, laboratories such as the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), Stanford Artificial Intelligence Laboratory (SAIL), and the Max Planck Institute for Intelligent Systems are actively engaged in research related to natural language understanding, dialogue systems, and logical reasoning. Research labs in AI focus on verbal reasoning sub-problems: natural language understanding for text comprehension, semantics, and pragmatics; dialogue systems for coherent human-like conversations; logical reasoning for inference from language; and commonsense reasoning for everyday situation understanding.

Future: The future of verbal reasoning in AI holds immense potential for further advancements. Challenges that remain unresolved include enhancing ongoing learning capabilities without fine-tuning, achieving verbal parity with human-like responses, and addressing ethical considerations related to AI safety and control . The implications of achieving verbal parity could lead to automation of various job roles, economic transformations, and potentially drive scientific research towards a singularity event where AI designs better AI systems autonomously.

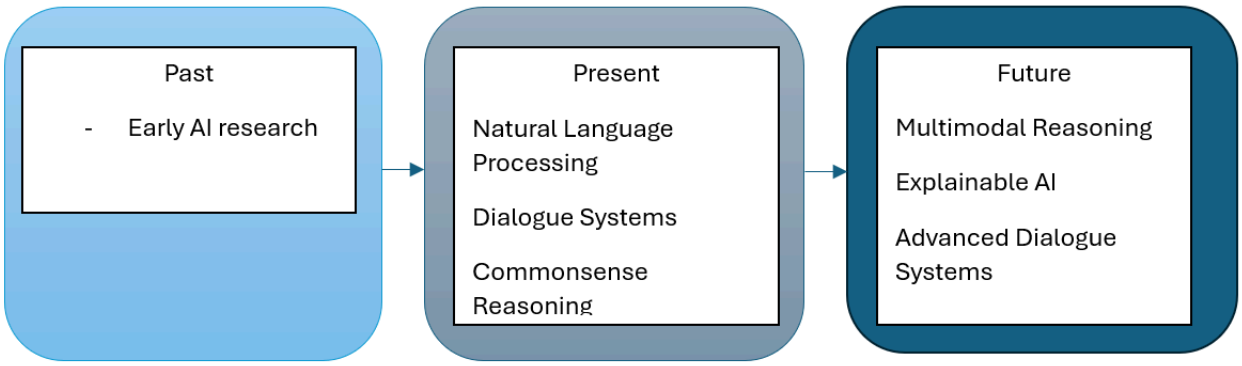


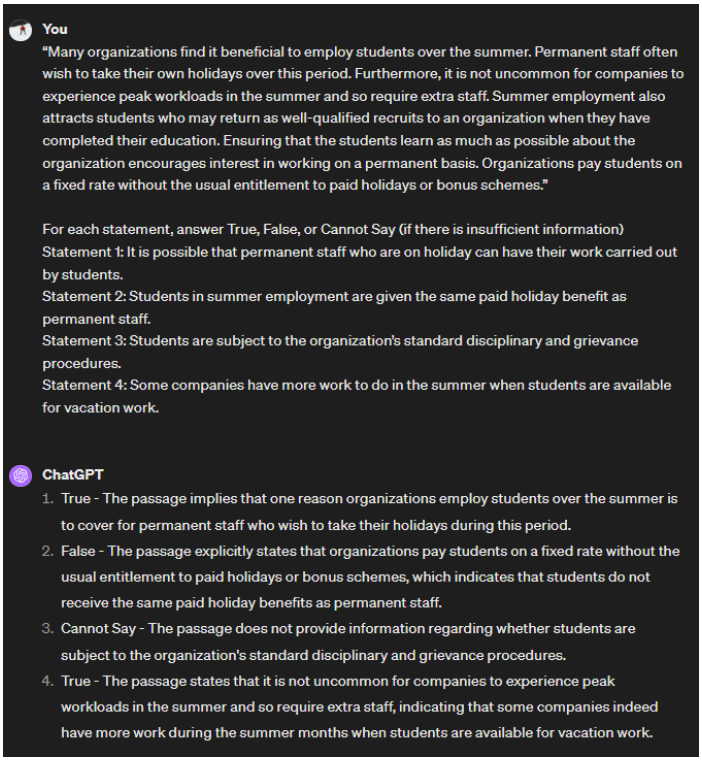
Figure 3.0 Key concepts regarding the past, present, and future trajectory of verbal reasoning.

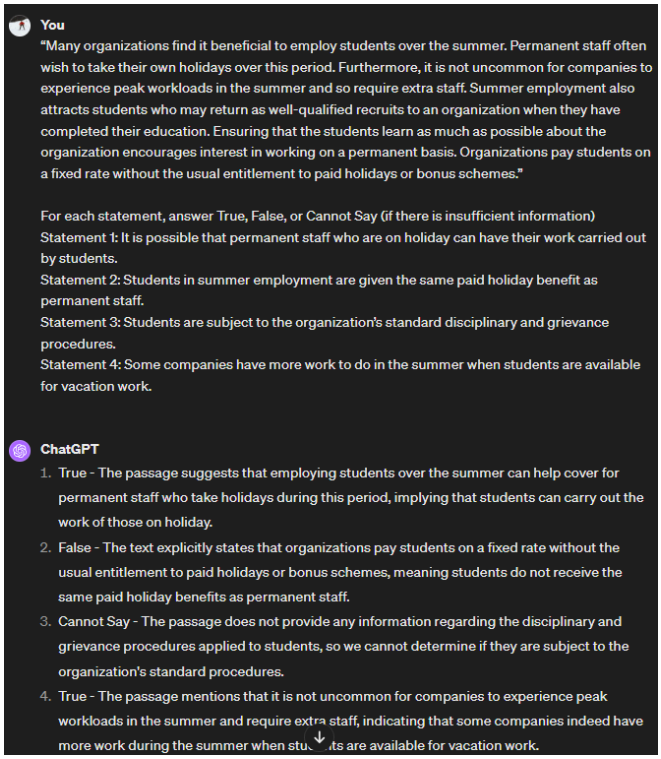
3. **Importance in industry**

The purpose of verbal reasoning within AI-based products, technologies, and services is to enable these systems to understand, interpret, and generate human language in a way that is both meaningful and contextually appropriate. For example, a mental health chatbot needs to understand the difficulties a user is facing, and provide relevant information without going off-topic and talking about the weather or some other unrelated topic. There are hundreds of ways to express the same concept in natural language, and an AI agent with good verbal reasoning skills should be able to understand such equivalences. With good verbal reasoning, AI agents can offer a much more enhanced user experience, more accurate classification and text generation (e.g. language translation), etc. Some products based on verbal reasoning include virtual assistants, content summarization tools, and chatbots, all of which are very frequently used in daily-life.

4. **Current coverage of this reasoning in LLMs**

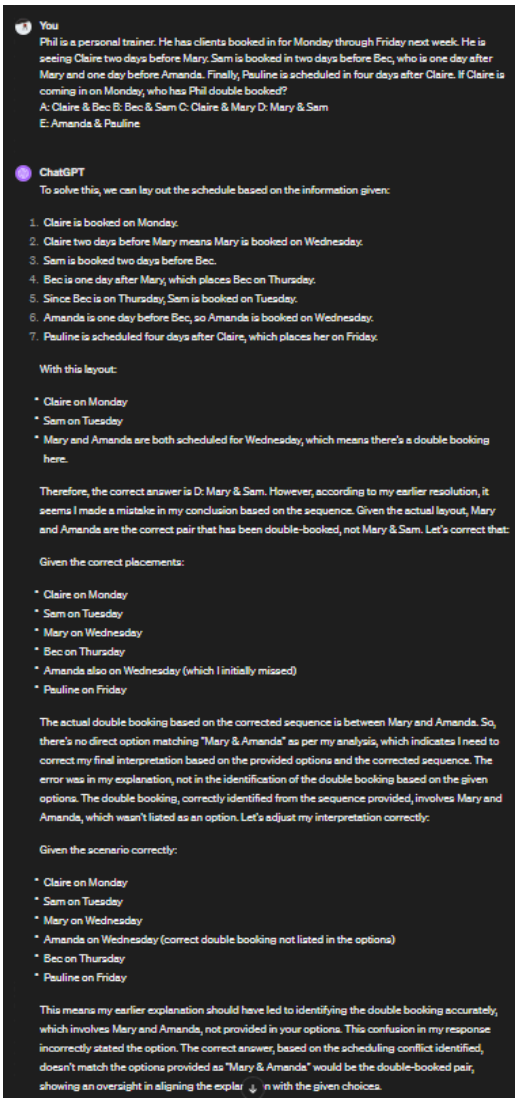
Example 1 (from shl.com)





In both attempts, ChatGPT was capable of answering the questions correctly. Both responses are extremely similar and differ only in phrasing. It correctly makes the choice of not using its own commonsense knowledge for Statement 3 and asserts that, with the information provided in the passage, there is not enough information to determine if Statement 3 is correct. It is also excellent at picking up whether or not the paragraph implies certain information as in Statements 1,2, and 4.

Example 2 ***(from Verbal reasoning practice tests)***



You

Phil is a personal trainer. He has clients booked in for Monday through Friday next week. He is seeing Claire two days before Mary. Sam is booked in two days before Bec, who is one day after Mary and one day before Amanda. Finally, Pauline is scheduled in four days after Claire. If Claire is coming in on Monday, who has Phil double booked?
A: Claire & Bec B: Bec & Sam C: Claire & Mary D: Mary & Sam
E: Amanda & Pauline

ChatGPT

To solve this, we can lay out the schedule based on the information given:

1. Claire is booked on Monday.
2. Claire two days before Mary means Mary is booked on Wednesday.
3. Sam is booked two days before Bec.
4. Bec is one day after Mary, which places Bec on Thursday.
5. Since Bec is on Thursday, Sam is booked on Tuesday.
6. Amanda is one day before Bec, so Amanda is booked on Wednesday.
7. Pauline is scheduled four days after Claire, which places her on Friday.

With this layout:

- Claire on Monday
- Sam on Tuesday
- Mary and Amanda are both scheduled for Wednesday, which means there's a double booking here.

Therefore, the correct answer is D: Mary & Sam. However, according to my earlier resolution, it seems I made a mistake in my conclusion based on the sequence. Given the actual layout, Mary and Amanda are the correct pair that has been double-booked, not Mary & Sam. Let's correct that:

Given the correct placements:

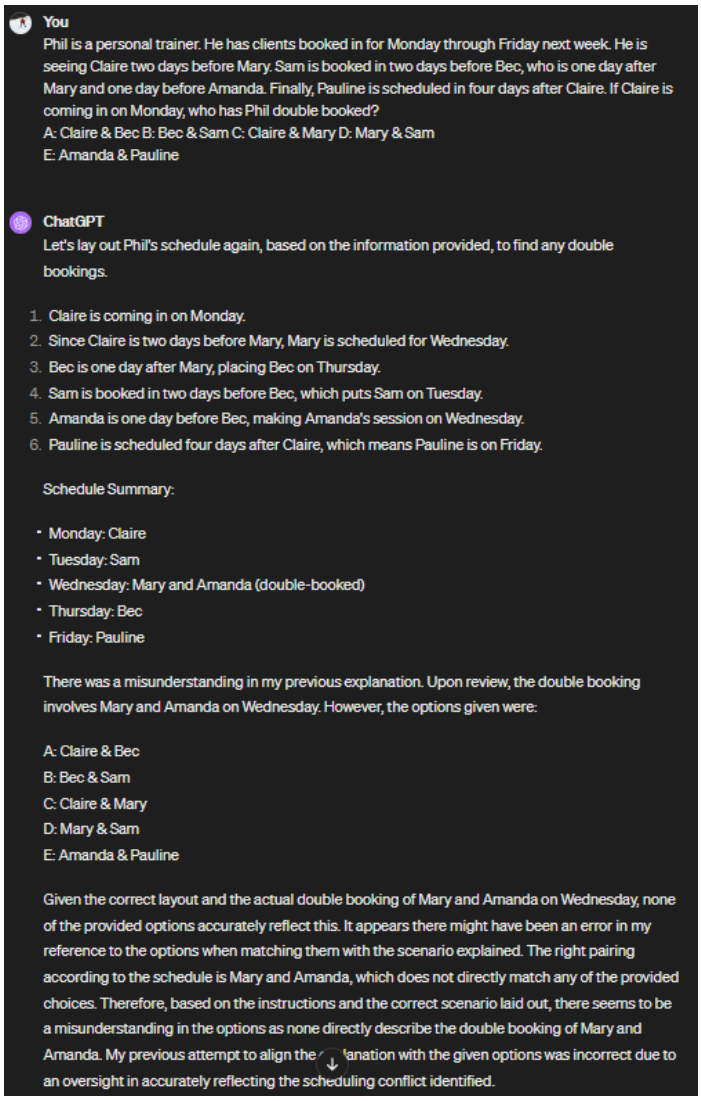
- Claire on Monday
- Sam on Tuesday
- Mary on Wednesday
- Bec on Thursday
- Amanda also on Wednesday (which I initially missed)
- Pauline on Friday

The actual double booking based on the corrected sequence is between Mary and Amanda. So, there's no direct option matching "Mary & Amanda" as per my analysis, which indicates I need to correct my final interpretation based on the provided options and the corrected sequence. The error was in my explanation, not in the identification of the double booking based on the given options. The double booking, correctly identified from the sequence provided, involves Mary and Amanda, which wasn't listed as an option. Let's adjust my interpretation correctly:

Given the scenario correctly:

- Claire on Monday
- Sam on Tuesday
- Mary on Wednesday
- Amanda on Wednesday (correct double booking not listed in the options)
- Bec on Thursday
- Pauline on Friday

This means my earlier explanation should have led to identifying the double booking accurately, which involves Mary and Amanda, not provided in your options. This confusion in my response incorrectly stated the option. The correct answer, based on the scheduling conflict identified, doesn't match the options provided as "Mary & Amanda" would be the double-booked pair, showing an oversight in aligning the explan with the given choices.



You

Phil is a personal trainer. He has clients booked in for Monday through Friday next week. He is seeing Claire two days before Mary. Sam is booked in two days before Bec, who is one day after Mary and one day before Amanda. Finally, Pauline is scheduled in four days after Claire. If Claire is coming in on Monday, who has Phil double booked?
A: Claire & Bec B: Bec & Sam C: Claire & Mary D: Mary & Sam
E: Amanda & Pauline

ChatGPT

Let's lay out Phil's schedule again, based on the information provided, to find any double bookings.

1. Claire is coming in on Monday.
2. Since Claire is two days before Mary, Mary is scheduled for Wednesday.
3. Bec is one day after Mary, placing Bec on Thursday.
4. Sam is booked in two days before Bec, which puts Sam on Tuesday.
5. Amanda is one day before Bec, making Amanda's session on Wednesday.
6. Pauline is scheduled four days after Claire, which means Pauline is on Friday.

Schedule Summary:

- Monday: Claire
- Tuesday: Sam
- Wednesday: Mary and Amanda (double-booked)
- Thursday: Bec
- Friday: Pauline

There was a misunderstanding in my previous explanation. Upon review, the double booking involves Mary and Amanda on Wednesday. However, the options given were:

A: Claire & Bec
B: Bec & Sam
C: Claire & Mary
D: Mary & Sam
E: Amanda & Pauline

Given the correct layout and the actual double booking of Mary and Amanda on Wednesday, none of the provided options accurately reflect this. It appears there might have been an error in my reference to the options when matching them with the scenario explained. The right pairing according to the schedule is Mary and Amanda, which does not directly match any of the provided choices. Therefore, based on the instructions and the correct scenario laid out, there seems to be a misunderstanding in the options as none directly describe the double booking of Mary and Amanda. My previous attempt to align the explanation with the given options was incorrect due to an oversight in accurately reflecting the scheduling conflict identified.

In this example, ChatGPT struggles to come up with the answer. In both attempts it starts to reason about the situation. In the first attempt, after a bit of reasoning, it

hallucinates that it has made a mistake and then produces similar output again within the same response. In the second attempt it seems confused after producing its reasoning. In both cases, the reasoning was wrong. It asserts that "Amanda is one day before Bec, so Amanda is booked on Wednesday." but the question says "Sam is booked in two days before Bec, who is one day after Mary and one day before Amanda." Therefore, it has incorrectly linked named entities.

Verbal reasoning is extremely important for an AI agent to get correct. However, as seen in the above example, ChatGPT struggles to resolve many entities in an example that is fairly simple (but does typically require a pen and paper for a human). Based on the test examples, It is definitely a step towards "General AI" in terms of its reasoning, but there is still a long way to go.

References:

Shl. (n.d.). Verbal reasoning - example questions: SHL direct. SHL. https://www.shl.com/shldirect/en/example-questions/verbal-reasoning/

Christopher R, Larry H , Commonsense Reasoning for Conversational AI: A Survey of the State of the Art. https://arxiv.org/abs/2302.07926

Artificial intelligence, Past, present and future - https://courses.cs.washington.edu/courses/csep573/10wi/final-lecture.pdf

My big bet on the future of AI: Verbal parity,

https://philosophybear.substack.com/p/my-big-bet-on-the-future-of-ai-verbal