Socialis Series in Social Science
ISSN 2583-1585

Sembok & Wani, 2023

Volume 5, pp. 96-110

Received: 11th April 2023

Revised: 31st July 2023, 10th August 2023, 16th August 2023

Accepted: 1st September 2023

Date of Publication: 7th November 2023

This paper can be cited as: Sembok, T. M. T. & Wani, S. (2023). Information Retrieval, Question Answering Systems, and ChatGPT: Technology, Capability, and Intelligence. Socialis Series in Social Science, 5, 96-110.

INFORMATION RETRIEVAL, QUESTION ANSWERING SYSTEMS, AND CHATGPT: TECHNOLOGY, CAPABILITY, AND INTELLIGENCE

Tengku M. T. Sembok
Professor, Department of Computer Science, Kulliyyah of ICT, International Islamic University Malaysia, Gombak, Malaysia
tmts@iium.edu.my

Sharyar Wani
Assistant Professor, Department of Computer Science, Kulliyyah of ICT, International Islamic University Malaysia, Gombak, Malaysia
sharya@iium.edu.my

Abstract

The question is will ChatGPT replace Google? To answer the question, we have a look at the engine that powers ChatGPT. Knowing the engine that drives it, we may be able to estimate its capability and power. Information retrieval and question answering systems have been around for decades using well known technology to serve their purposes, such as keywords approach, semantic technology, predicate logic, and natural language processing. Several WH questions have been tested to assess the quality and the correctness of the answers given by ChatGPT. With respect to What, Where, and How questions, we found that ChatGPT seems to understand the questions and answer them quite well. But some facts embedded in the answers are found to be

inaccurate. These may be due to the "hallucination" problem caused by the Large Language Models (LLMs) adopted in the engine. We wonder here whether ChatGPT, which adopts the Generative Pre-trained Transformer based on LLMs, has reasoning mechanism as in the rule-based systems. In conclusion, ChatGPT represents a significant leap forward in the field of question answering systems. However, it is important to recognize that ChatGPT and Google are not direct competitors, but rather complementary tools that serve different purposes.

Keywords
Artificial Intelligence, Information Retrieval, Question Answering Systems

1. Intelligence

One of many definitions of intelligence, relevant to our discussion in this article, is: "*the general mental ability involved in calculating, reasoning, perceiving relationships and analogies, learning quickly, storing and retrieving information, using language fluently, classifying, generalising, and adjusting to new situations"* (Columbia Encyclopaedia, 2023).

There seems to be no standard definitions of intelligence agreed upon after many years of debate by the experts. But some believe that it might be observed through a behaviourism approach in deciding whether a machine possesses intelligence. Alan Turing (1950) came out with the Turing Test as an intelligence test for machines. If a machine can engage in a conversation with a human, and be able to make him/her to believe that it is human, then we can say that the machine possesses intelligence. Not all agreed with the reliability of the Turing Test. Searle (1980) is among those who disagree with the Turing Test. In his "Chinese Room" scenario, he argues that the behaviourism approach does not prove categorically that the machine understands the language in the experiment and thus possesses intelligence.

2. Chinese Room

The scenario posted by Searle (1980): There is a room with two doors with a slot on each door. Questions in Chinese can be written on a piece of paper and slip into the room through one slop. The answer is then written in Chinese, by the person in the room who does not know Chinese, on the paper and pass it outside through the other slot. The person in the room answers the questions by referring to manuals and instructions provided in the room. Assume that the

questioner is a native Chinese speaker. After getting back the paper, the Chinese native speaker discovers that the answer written in Chinese is an intelligent answer. Therefore, can we conclude that the person in the room understands Chinese and possesses intelligence? (The Mind Project, 2023).

On assumption, the person in the room who is being tasked to answer the questions, does not know Chinese. The question is, does he possess intelligence based on the intelligent answers given to the questions asked? The same goes to the Turing Test, does a computer possess intelligence if it is able to fool people that it is human? Can we equate ChatGPT to that person in the room who is being tasked to answer questions? Thus, will ChatGPT pass the Turing test and be concluded to possess intelligence?

3. Public Response on ChatGPT

ChatGPT has shown fantastic capability in mimicking humans in language handling and stunned many users. On the other hand, many also have shown negative vibes. It has received many positive and negative reviews as soon as it was launched at the end of the year 2022. (https://en.wikipedia.org/wiki/ChatGPT#cite_note-107, retrieved 9 April 2023).

Some celebrate the release and introduction of the chatbot as the breakthrough of the year and consider it as the best, and might change how people work and think (Roose, 2022; Lock, 2022; Thompson, 2022; Piper, 2022). Others considered it as a “mass destruction” that could result in cheating in schools, disinformation, and military chaos (Karp, 2023). Some made comment on its technology as lossy JPEG (Chiang, 2023), and worse still as a faux science and incompetence linguistic system (Chomsky et al., 2023). Thousands of public figures signed an open petition for immediate pause of giant AI experiments like ChatGPT (Hurst, 2023).

Finally, Brian Hood, mayor of Hepburn Shire Council, plans to take legal action against ChatGPT over false information generated. ChatGPT erroneously claimed that he was jailed for bribery during his tenure at a subsidiary of Australia's national bank. Hood was not jailed for bribery. Should Hood proceed with the lawsuit, it would be the first public defamation case against ChatGPT (BBC, 2023).

People are starting to ask how ChatGPT will be used to influence the US presidential election in 2024.

4. The Engine Tells All

ChatGPT has impressed users as well as researchers with its capabilities while also worrying regulators about the negative implications for society. We are interested in knowing the engine that powers ChatGPT which exhibits tremendous capabilities in answering questions from the users. But, yet can also generate wrong answers. This is what we are concerned with. We have tested with a few questions and some answers given by ChatGPT were obviously factually wrong. Two examples of the questions are given below. In Figure-1, on the Malaysian Prime Minister, the factual error is that his wife did not study in the University of Malaya with him. In Figure-2, on the author of this article himself, the truth is he never was in military service as a brigadier general even though he had served in the National Defence University of Malaysia.

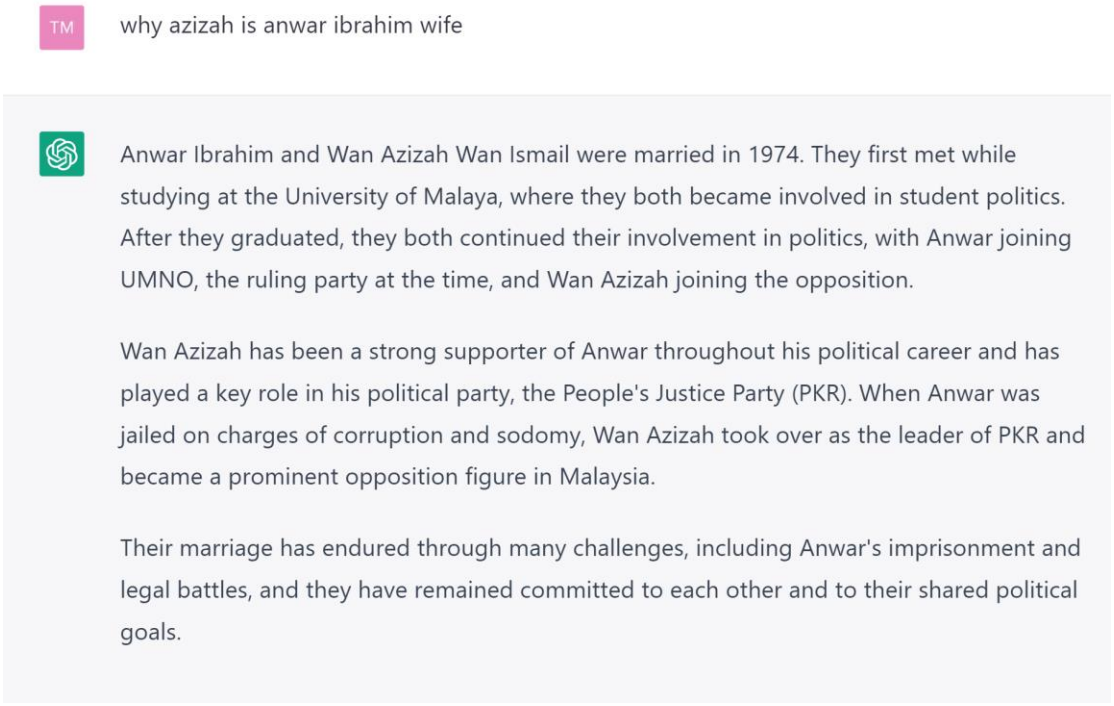


Figure 1: *Factual Error: Azizah did not study in University of Malaya with Anwar Ibrahim (Source: Authors' Own illustration)*

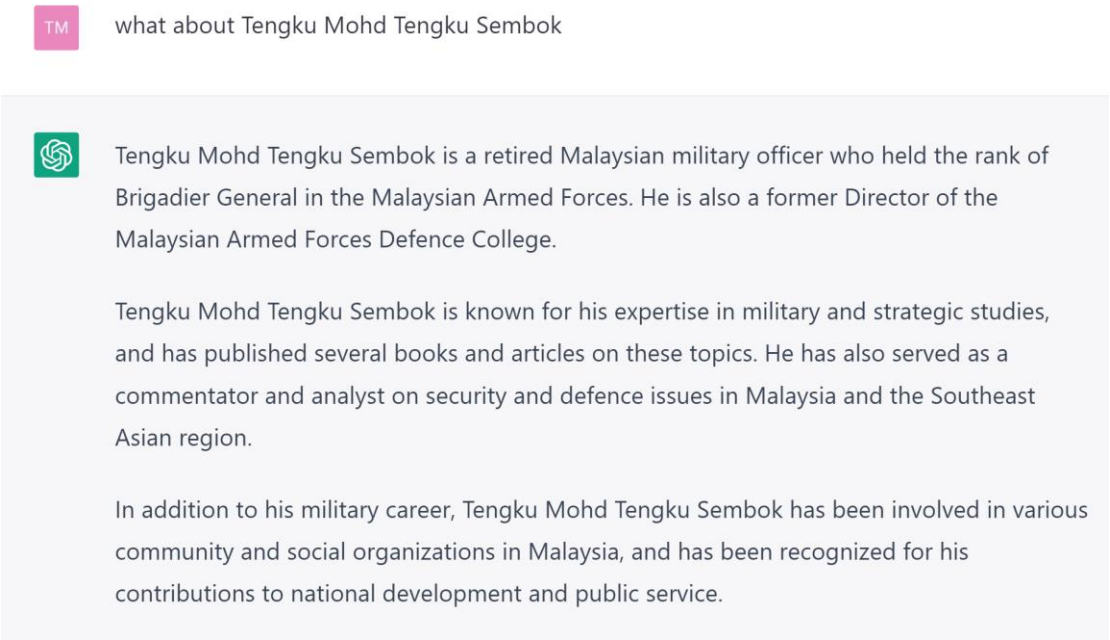


Figure 2: *Factual Error: Sembok was never a Malaysian Military Officer (Source: Authors' Own illustration)*

Knowing the engine that powers the chatbot might explain these phenomena. We have a quick look at the fundamental technology that runs related systems with ChatGPT, i.e., information retrieval systems and question answering systems which are close to our work. Information retrieval systems have dominated information seeking services online for decades led by search engines Google, Bings, Yahoo, and others. Knowing the technology behind any intelligent system, we might be able to understand the capability as well as its limitations and shortcomings.

The aim of this article is to have a quick look at the basic engine used by information retrieval systems and related question-answering systems, that we are familiar with in our work, and compare with the engine that powers ChatGPT.

5. Information Retrieval

Information Retrieval (IR) is defined broadly as the study of how to determine and retrieve *relevant* documents for a particular information request from a corpus of stored documents. Surrogates are used to represent the documents in the corpus using knowledge representation formalism. The efficiency and effectiveness of an information retrieval system is very much dependent on its knowledge representation (Mizzaro, 1997).

Most of the conventional IR systems used a keyword approach in indexing and retrieving the relevant documents. In these systems, a document is represented by an independent collection of keywords or terms, in a vector space. And, a query is similarly represented. *Di* and *Q* below, represent a document *i*, a query, respectively.

Di = (t1, wi1; t2, wi2; ...; tn, win); e.g., Di = (question, 0.9; answer,1.0; system, 0.8, ...).

Q = (t1, q1; t2, q2; ...; tj, qn); e.g., Q = (question, 1.0; answer, 0; system, 0.7, ...).

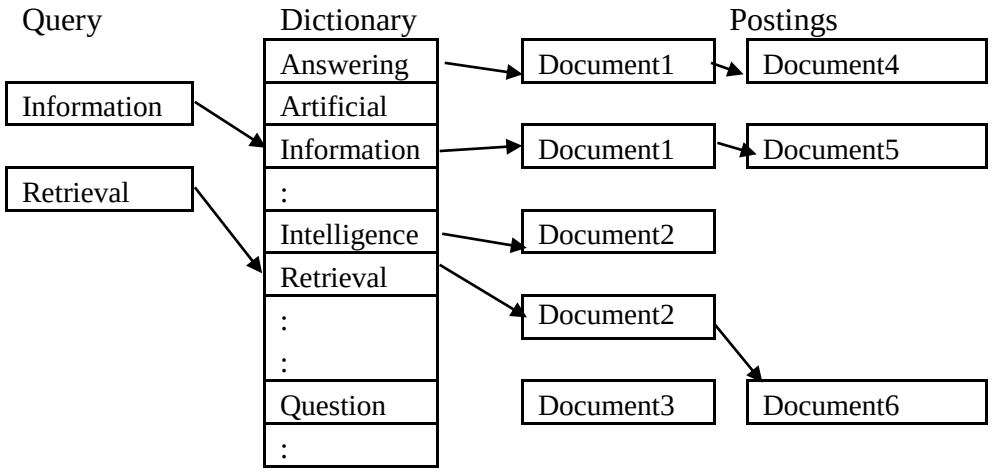
Where:

Di represents a vector for document *i* comprises words, *ti* to *tn*, with respective weight, *wi*, to *wn* to represent the importance of each word in the document *Di*.

Q represents a query vector composed of words *t1* to *tn* with respective weights *q1* to *qn* representing the importance of the respective words in the query.

A similarity matching between documents and query is performed based on certain formula such as Inner Product Function as given below:

Similarity (*Di, Q*) = $\sum_{j=1,n}$ (*qj . wij*); e.g. similarity value between *Di* and *Q*: Similarity (*Di,Q*) = (0.9 x 1 + 0 + 0.8 x 0.7)



Indexing: Inverted File

Figure-3: *Document Index in Information Retrieval*
(Source: Authors' Own illustration)

The documents will be presented to the user sorted based on their similarity values. The more the similarity value the more likely it to be relevant to the user. Knowing this basic technique in making judgement on relevancy of documents, provide us information on how the system works. Basically, it is how information retrieval systems, such as Google, operate, in representing most likely relevant documents to our query. These systems normally use indexing mechanisms to efficiently implement the storing and retrieving process as depicted in the Figure-3 above. The aim here is to show the transparency of how the system works.

6. Question Answering Systems

Users might want to get a specific answer for a specific question. This brings us to Question Answering Systems. Question answering process comprises "reading and comprehension" tasks by the system to retrieve answers for the question from the contents of the relevant documents. This involves knowledge on the syntax, semantic and pragmatic domains of natural language understanding (McGuinnes, 2004;-Sembok et al., 2013).

The common approach to natural language understanding is to use meaningful knowledge representation of documents which can incorporate syntactic and semantic knowledge as well as world knowledge in areas related to the question. For example if there are three keywords *boy, cut,* and *ribbon*, then the phrase "boy cut the ribbon" can be represented by *boy(x), ribbon(y) and cut(x,y),* using logical knowledge representation. We have experimented using predicate logic representation to incorporate the semantic in the surrogates, as illustrated in Figure-4. (Kadir et al., 2009; 2012) (Nau, 2017)

The architecture given in Figure-4 illustrates that the document surrogates are in predicate logic representation which incorporate logical deductive reasoning capability that can be used to answer questions (Sembok et al., 2013). That implements transparency in the system and the ability to explain the basis and support in arriving at the answer. In other words, the source of information to derive the answer can be traced back to which documents and sentences that contribute to answering.

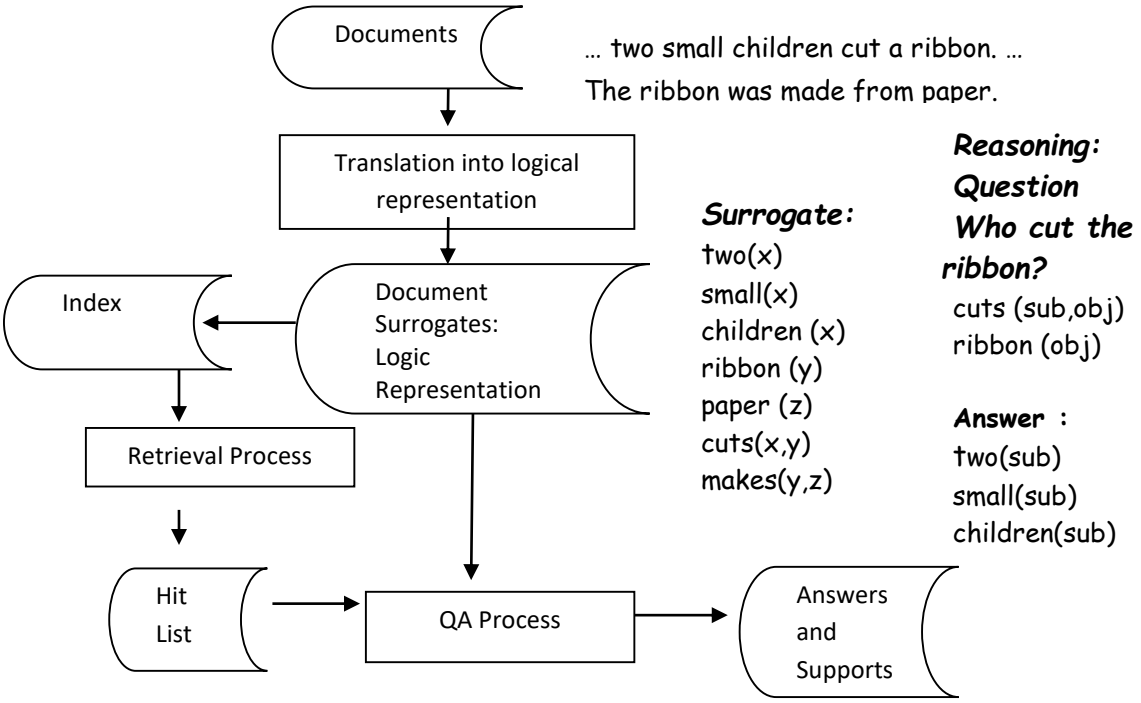


Figure 4: *Architecture of a Question Answering System with Document Surrogate in Predicate Logic to Incorporating Deductive Reasoning*
(Source: Authors' Own illustration)

7. Knowledge Representation Models

We have highlighted two models of knowledge representation as used in information retrieval and question answering systems. Based on the models used, we can assess the strength and the limitation of the systems powered by the models. Next, let us have a look at the model that powers ChatGPT in order to assess its strength and limitations. Two basic questions can be asked regarding the capability of the model used in this study; they are:

1) Does the system have reasoning capability, such as deduction or induction reasoning?

2) Can the system provide an explanation on the answer given by tracing back on the source of knowledge that contributes to the answer?

8. ChatGPT

ChatGPT uses a variant of Large Language Models called generative pre-trained transformer (GPT). At its core is a sophisticated Artificial Neural Network engine that has been fine-tuned using supervised and reinforcement learning techniques that power the model. It uses statistics to analyse vast amounts of data, learning the patterns and connections between words and phrases. Basically, given an unfinished sentence with missing words, the algorithm can suggest the most suitable words to complete the sentence based on its very large training data set. Figure-5 illustrates the GPT model predicting the missing word (Vaswani et al., 2017).

Although the underlying technology behind ChatGPT has shown fantastic capability in mimicking humans in language handling, but as reported, it suffers from shortcomings such as hallucination, bias, toxicity, and fairness in its output (reference). The main reason this happened is because the generation engine of words is based upon the statistics on proximity and frequency of words occurring in the corpus of the training data set. Figure-6 illustrates the underlying basic algorithm of Large Language Models. Decisions on the choices of words are based upon billions of parameters represented by nodes and links in the neural networks embedded in the Encoder and Decoder of GPT. Each link has weight to denote the import of the link which is derived from the training data set.

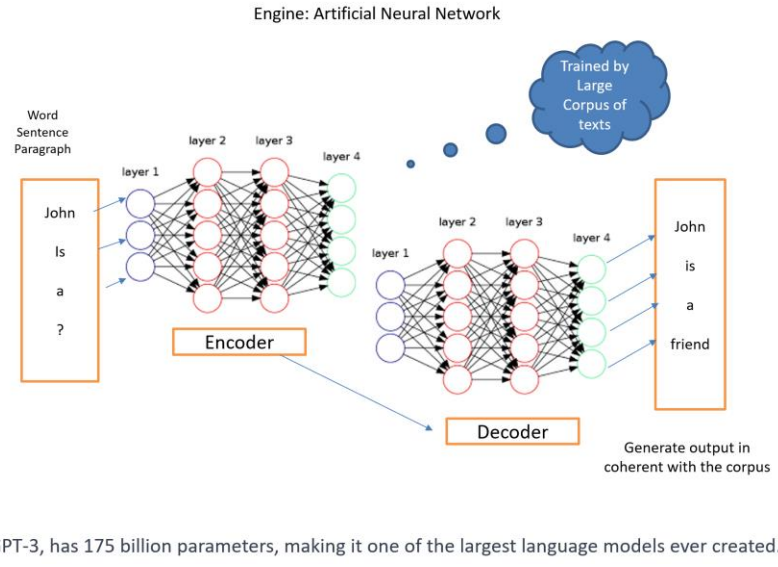


Figure 5: *Generative Pre-trained Transformer: Powered by Artificial Neural Network (Source: Authors' Own illustration)*

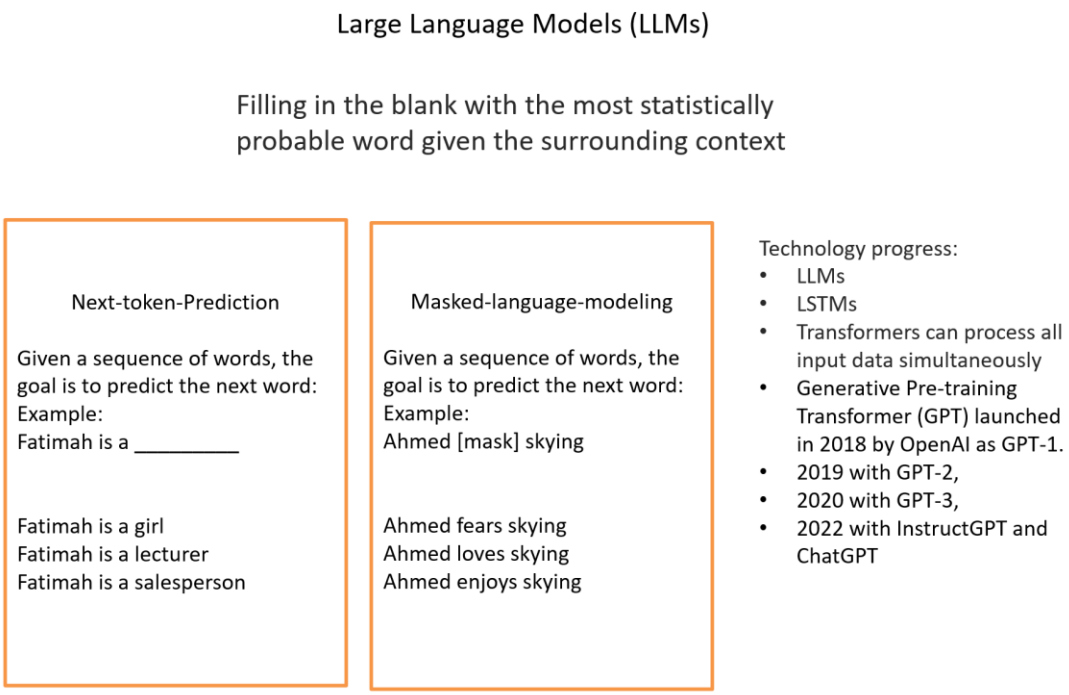


Figure 6: *LLMs: Filling in the Missing Word Based on Statistically Probable Word (Source: Authors' Own Illustration)*

9. Experiment on a Set of Questions

A set of 19 questions from our previous experiment in question answering system on Al-Quran knowledge base has been selected from the experiment by Wani & Sembok (2018). Seeking information from the holy Muslim scripture is becoming more popular in the digital world because more people are seeking spiritual guidance in this troubling era of social media (Razzaq, 2018). These questions have been tested against ChatGPT and the answers obtained are compared to the results obtained from the previous experiment. The question answering sessions has been carried out exclusively with ChatGPT separately from Google. The answers obtained are classified into 3 categories as follows: 1) correct answer-correct reference; 2) correct answer-wrong reference; 3) wrong answer-wrong reference.

The comparison between the two performances obtained is summarised in the Table-1 below. It shows that ChatGPT has a problem in tracing the source of the answers.

Table 1: *Performance Comparison*

Experiment	**Results**		
	Correct Answer-Correct Reference	Correct Answer-Wrong Reference	Wrong Answer-Wrong Reference
Wani's	18	0	1
ChatGPT	9	5	5

(Source: Authors' Own Illustration)

All the 19 questions and the classification of results for the questions are listed in the Appendix.

10. Conclusion

We can conclude here, the generation of answer by ChatGPT to a question is based upon the collection of words in the question and how it is propagated to words, sentences and paragraphs that form the answer to the question. Instead of looking through manuals as in the case of the person in the Chinese Room, ChatGPT is looking through the nodes and links in the neural networks to generate the answer. Many equate an artificial neural network as a "Black Box", that we cannot trace its reasoning process in the light of deductive or inductive reasoning. As in the Aristotle's deduction, *Socrates is a man, all men are mortal, and therefore Socrates is mortal*. Or, as in induction, *all flamingos I have ever seen are pink, therefore all flamingos must be pink.*

ChatGPT seems to have a kind of "Black Box" reasoning based on statistics on proximity and frequency of word occurrences from a huge training data set. But its line of reasoning cannot be traced nor does its source of information that derived the answer. Unlike, the rule-based AI where the reasoning process can be traced back to provide explanation to the answer derived. There are efforts to incorporate traceable reasoning capability and explanation ability to AI systems that are powered by Artificial Neural Network.

Thus, the war to dominate the information seeking market is still on. Whether the person in the Chinese Room or ChatGPT possesses intelligence remains to be answered.

REFERENCES

BBC (2023, April 7) "ChatGPT: Mayor starts legal bid over false bribery claim". *[News Article]*. *BBC News*. https://www.bbc.com/news/technology-65202597

Chiang, T. (2023, February 17). "ChatGPT Is a Blurry JPEG of the Web". *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

Chomsky, N., Roberts, I., Watumull, J. (2023, March 12). "Opinion | Noam Chomsky: The False Promise of ChatGPT". *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Columbia Encyclopedia. (2023, April 23). Intelligence. Infoplease. https://www.infoplease.com/encyclopedia/medicine/psychology/discipline/intelligence#:~:text=intelligence%2C%20in%20psychology%2C%20the%20general,and%20adjusting%20to%20new%20situations

Hurst, L. (2023, March 30). “Profound risk to humanity”; “Tech leaders call for 'pause' on advanced AI development". *Euronews*. https://www.euronews.com/next/2023/03/29/profound-risk-to-humanity-elon-musk-and-steve-wozniak-join-calls-to-halt-ai-development

Kadir, R. A., Sembok, T. M. T., & Zaman, H. B. (2012). Intelligent Information Access Based on Logical Semantic Binding Method. In *Advances in Knowledge Representation*. https://www.intechopen.com/chapters/36659

Kadir, R.A., & Sembok, T.M.T. (2009). “Improvement of document understanding ability through the notion of answer literal expansion in logical-linguistic approach”. *WSEAS Transactions on Information Science and Applications*, *6*, 966-975. https://www.researchgate.net/publication/265745875_Improvement_of_Document_Understanding_Ability_through_the_Notion_of_Answer_Literal_Expansion_in_Logical-linguistic_Approach

Karp, P. (2023, February 6). "MP tells Australia's parliament AI could be used for 'mass destruction' in speech part-written by ChatGPT". *The Guardian*. https://www.theguardian.com/australia-news/2023/feb/06/labor-mp-julian-hill-australia-parliament-speech-ai-part-written-by-chatgpt

Lock, Samantha (2022, December 5). "What is AI chatbot phenomenon ChatGPT and could it replace humans?". *The Guardian*. https://web.archive.org/web/20230116100346/https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans

McGuinnes, D. L. Question Answering on the Semantic Web. *IEEE Intelligent Systems* 19(1), (2004), Jan/Feb 2004: pp. 82 - 85. https://doi.org/10.1109/MIS.2004.1265890

Mizzaro, S. 1997. Relevance: The Whole History. *JASIS*, Vol.48, No.9, pp.810-832. https://doi.org/10.1023/A:1016323201283

Nau, J., Filho, A., & Passero, G. (2017). "Evaluating Semantic Analysis Methods for Short Answer Grading Using Linear Regression." *PEOPLE: International Journal of Social Sciences*, *3*(2), 437–450. https://doi.org/10.20319/pijss.2017.32.437450

Piper, K. (2022, December 15). "ChatGPT has given everyone a glimpse at AI's astounding progress". *Vox*. https://www.vox.com/future-perfect/2022/12/15/23509014/chatgpt-artificial-intelligence-openai-language-models-ai-risk-google

Razzaq, T. (2018). "Spirituality in The Perspective of Sacred Scriptures of The World Religion". *PEOPLE: International Journal of Social Sciences*, *4*(3), 700–710. https://doi.org/10.20319/pijss.2018.43.700710

Roose, Kevin (2022, December 5). "The Brilliance and Weirdness of ChatGPT". *The New York Times*. https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html

Searle, John R. (1980), Minds, Brains, and programs, The Behavioral and Brain Sciences 3, 417-457. https://doi.org/10.1017/S0140525X00005756

Sembok, T. M., Zaman, H. B., & Kadir, R. A. (2008). IRQAS: Information Retrieval and Question Answering System Based on a Unified Logical-Linguistic Model. In *Proceedings of the 7th WSEAS International Conference on Artificial intelligence, knowledge engineering and data bases, Cambridge, UK*. https://www.researchgate.net/publication/234778531_IRQAS_information_retrieval_and_question_answering_system_based_on_a_unified_logical-linguistic_model

Sembok, T. T., & Kadir, R. A. (2013). A unified logical-linguistic indexing for search engines and question answering. *International Journal of Mathematical Models and Methods in Applied Sciences*, *7*, 22-29. https://www.naun.org/main/NAUN/ijmmas/16-553.pdf

TheMindProject. (2023 April 8). Retrieved from https://mind.ilstu.edu/curriculum/turing_test/turing_test_curric_long.html

Thompson, D. (2022, December 8). "Breakthroughs of the Year". *The Atlantic*. https://www.theatlantic.com/newsletters/archive/2022/12/technology-medicine-law-ai-10-breakthroughs-2022/672390/

Turing, A.M. (1950), Computing Machinery and Intelligence, Mind, vol.LIX, No. 236. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing.* https://doi.org/10.48550/arXiv.1706.03762

Wani, S., Sembok, T.M.T. (2018). Constructing a knowledge base for Al-Qur'an utilising principles of human communication. CAMP'18. 26-28 March 2018. Kota Kinabalu, Malaysia. https://doi.org/10.1109/INFRKM.2018.8464823

Appendix: Performance Comparison

Questions	Wani's Result	ChatGPT's Result
What did the wife of 'Imran pledge to her Lord?	Correct Answer Correct Reference	Correct Answer Correct Reference
What was the child of 'Imran named?	Wrong Answer Wrong Reference	Wrong Answer Wrong Reference
What did the wife of 'Imran seek refuge for her child?	Correct Answer Correct Reference	Wrong Answer Wrong Reference
What did Zakaria find every time he entered the chamber?	Correct Answer Correct Reference	Correct Answer Correct Reference
What happened when Zechariah was standing in prayer in the chamber?	Correct Answer Correct Reference	Correct Answer Wrong Reference
What did Zechariah think was not possible since he had reached old age?	Correct Answer Correct Reference	Correct Answer Wrong Reference
What did Abraham said to his father Azar?	Correct Answer Correct Reference	Correct Answer Wrong Reference
What is Salsabeel?	Correct Answer Correct Reference	Correct Answer Wrong Reference

What is the story of tribe of Saba'?	Correct Answer Correct Reference	Correct Answer Correct Reference
What was the Prophet commanded to tell his wives?	Correct Answer Correct Reference	Correct Answer Wrong Reference
What was the dream of Joseph?	Correct Answer Correct Reference	Correct Answer Correct Reference
What happened in the city of Babylon?	Correct Answer Correct Reference	Wrong Answer Wrong Reference
Should we take our fathers as allies if they prefer disbelief?	Correct Answer Correct Reference	Correct Answer Correct Reference
Whose care was Mary in?	Correct Answer Correct Reference	Correct Answer Correct Reference
Who provides without account?	Correct Answer Correct Reference	Correct Answer Wrong Reference
Who did Zechariah call upon for offspring?	Correct Answer Correct Reference	Correct Answer Correct Reference
Who was Zechariah asked to remember in the morning and evening?	Correct Answer Correct Reference	Wrong Answer Wrong Reference
Who was chosen above the women of the worlds?	Correct Answer Correct Reference	Correct Answer Correct Reference
Who was given the good tidings about Jesus?	Correct Answer Correct Reference	Correct Answer Correct Reference