Evaluation of GenAI in Biomedical Knowledge Discovery

Radmila Juris
ALMAIS, UK
radjur3@gmail.com

Eiman Almami
MWLLO, UK
almami.e@ieee.org

Ibtesim Almami
Qassim University
I.Almami@qu.edu.sa

Abstract

This paper evaluates the power of GenAI and its generated results and compares them with the more traditional way of computational reasoning for biomedical knowledge discoveries. The experiments conducted in this research focus on the problem of discovering activated/inhibited enzymes involved in diseases, but the same idea of evaluating GenAI can be applied to any problem domain. The traditional way of knowledge discovery through the power of computational logic reasoning is now challenged by the rather unexpected power of LLM and transformer architecture, such as ChatGPT, in almost any discipline of science and humanities. The results from this research are interesting: ChatGPT disappoints in biomedical knowledge discoveries because of the lack of precision and absence of reasoning. However, it is feasible to define pathways, when using GenAI, which can remedy current ChatGPT shortcomings.

Keywords: GenAI, Biomedicine, Knowledge discovery, FOL.

1. Introduction

The success of Generative AI (GenAI) technology (USTPC, 2023), based on Large Language Models (LLM) (Shanahan, 2024) and Google's Transformer Architecture (Vaswani et al., 2017) took us by surprise and in just over a year we have been overwhelmed by its applicability in science, humanities and across numerous problem domains. Despite drawbacks of the technology (Berltran et al., 2024), (Preiß et al., 2024), sometimes resulting from the issues of using predictive inference (Genkina, 2024) and plausible LLM's guessing "which word comes next" (Wooldridge, 2023), (Li et al., 2024), we must acknowledge its scientific accomplishments. The important achievements are ESM-2/ESMFold from META (Lin et al., 2022) and AlphaFold from DeepMind (Abramson et al., 2024) protein folding, where over 214 million protein structures were predicted (Jumper et al., 2021) in a short time. They created biomedical information which could have taken years to discover using traditional biomedical processes. Considering that the core of GenAI is LLM, it is not surprising that GenAI delivered significant outputs in text manipulation. However, it surprised us that LLM delivered the output we did not expect (Chaikin, 2023), and the example of protein language models, where LLM is trained on protein sequences, is one of them.

This research looks at GenAI from the perspective of knowledge discoveries in biomedical science. We know that the amount of data and knowledge generated in biomedicine is unprecedented and there are no instant, effective and accurate methods for knowledge discoveries and sharing. Considering that data/knowledge are mostly in textual format, stored in electronic files, sometimes enriched with graphs, tables and figures, it is expected that GenAI has a role to play in their discovery and dissemination. The power of knowledge stored within and extracted from the LLM could enhance discoveries and dissemination of knowledge in any discipline.

Our continuous interest in knowledge discovery and sharing in biomedicine (Almami et al., 2016a) has been influenced by discoveries in biology, chemistry, medicine, healthcare, computer science and biomedicine itself. Over the last decade we have experimented with biomedical and scientific knowledge dissemination. We have used reasoning based on First Order Logic (FOL) (Almeida, et al. 2011), and housed it within specific computational models (Shojanoori et al., 2012), (Almami, 2017) (Juric et al., 2020), (Juric et al., 2021), (Juric et al., 2023). Logic reasoning has been efficient and accurate when using existing knowledge in biomedicine and creating logic inference to secure:

A) Finding and sharing of scientific discoveries,
B) Finding the lack of research which inhibits new discoveries and
C) Defining future research pathways in knowledge discoveries.

However successful knowledge discovery with FOL and logic reasoning, based on semantic overlapping of FOL concepts proved to be, preparation for modelling these computations and implementing them with Semantic Web Technologies (SWT) (W3C, 2004) is

URI: https://hdl.handle.net/10125/109449
978-0-9981331-8-8

not trivial and unfortunately, it is time consuming. Logic reasoning is computable and creates a rather short and concise code in the SWT languages, which does not create a computationally demanding solution. However, the modelling of FOL/SWT concepts and semantic of their relationships, which secures reasoning, and populating these concepts with relevant data, can be time consuming and requires special conceptual modelling expertise.

In this paper, we investigate whether the GenAI can address the problem of complexity of computations based on logic reasoning, when applying it in knowledge discovery in biomedicine. We expect that GenAI would help in its fast knowledge discoveries, without creating a time-consuming computational model based on FOL. In other words, we hope that GenAI can replace our tested model based on logic reasoning and give similar if not the same results in a fraction of the time.

Our investigation focuses on ChatGPT (Rayola, 2023) due to its undisputed popularity, number of parameters (Nimaee, et al., 2024) and size. The experiments which illustrate biomedical knowledge discovery are placed in the domain of activation/ inhibition of enzymes in various diseases, which helps in the management of diseases and drug discoveries.

The paper is structured as follows.

The background section 2 overviews the work related to GenAI and its use in creating, discovering and sharing biomedical knowledge. It also defines the methodology used in this paper, explains co-relations between FOL/SWT models and GenAI text generation and lists a set of publications of the related work.

Section 3 overviews earlier experiments and their results from (Almami et al., 2016), where FOL and the SWT languages OWL (ref) and SWRL (ref) are used for the discovery and analysis of activated/inhibited enzymes in various diseases. Section 4 runs similar experiments with ChatGPT and uses its predictive inference to get the same answers on enzyme analysis, as in section 3. Prompts imposed on ChatGPT (Genkina, 2024) were supposed to be questions which are very similar to the questions used in the experiments with FOL. The results of asking ChatGPT questions on enzyme activation/inhibition for a particular patient are very interesting and discussed/concluded in section 5.

2. The Background

LLM based on transformer architectures have changed the course, current vision for and expectations from AI, by offering unexpected, fast and attractive results of predictive inference, with an absence of reasoning, typical of human minds. The scale of LLM, number of their parameters and size of training data are excessive. The more we place within the LLM, the better the predictive inference. Disciplines from science and humanities are using GenAI for different purposes.

In biomedicine, GenAI has changed the way we search, collect, process and present data relevant for biomedical discoveries and scientific research. We talk about predicting chemical or protein structures, managing biomedical literature, dealing with biological sequences and arranging brain signals as time series (Bi et al., 2024). Therefore, it is likely that LLM and transformer architecture can assist in biomedical knowledge discoveries.

There are software tools which help in efficient literature research and information seeking in biomedicine (Jin et al., 2024). We started developing collaborative AI agents, which combine human creativity and expertise with AI's ability to analyse large datasets, navigate hypothesis spaces, and execute tasks, such as control of phenotypes, and the design of cellular circuits for the development of new therapies (Gao et al., 2024). BiomedGPT: A Unified Biomedical Generative Pre-trained Transformer for Vision, Language, and Multimodal Tasks has already been available (Zhang et al., 2024), and opportunities and challenges for using LLM in biomedicine and health are undisputed (Tian et al., 2023), (Jungwirth and Haluza, 2023), (Poon et al., 2023). Healthcare-customized LLMs like Med-PaLM, BioGPT, and DeepHealth, are still evolving, but we can confidently talk about Expert-Level Medical Question Answering with LLM (Singhal et al., 2023), (Sai et al., 2024) and explore prompt engineering (Ahmed et al., 2024) with Retrieval-Augmented Generation (RAG) (Mostert, 2023) for providing valuable means in accelerating biomedical discovery and improving health.

2.1. The Problem

It is very well known that the number of experiments and research projects in biomedicine is growing as we write this paper, and they leave mostly written, graphical and media documentation of their research and results all over academic and information sources, industry and media. It is almost impossible to collect. collate and summarise all these research/experimental outputs of biomedical knowledge generation because they are scattered across continents, scientific journals, internet blogs, webpages allocated to research, white papers from Industry, government bodies which regulate scientific research, conferences, and public lectures. There is no universal "place" where we can keep discoveries on biomedicine updated and share them. This is one of

the biggest problems in knowledge, its discoveries and sharing. Consequently, we should ask these questions,

Can GenAI finally be the answer to this problem?

Can ChatGPT deliver fast. i.e. instant results which are similar to running computations with FOL/SWT reasoning in knowledge discoveries?

Can ChatGPT cut preparation time in knowledge discoveries, answer ad-hoc questions and thus secure knowledge sharing?

Could we achieve knowledge discovery and dissemination by using ChatGPT only (i.e. avoiding the tested logic reasoning)?

Knowledge discoveries are expected to be feasible in GenAI because they are directly associated with questions, we may ask ChatGPT and use prompting. Knowledge dissemination is similar. Through prompting mechanisms and possible RAG, we can add knowledge in the common pool of data and share it. These will have a positive impact on GenAI model training and LLM tuning. GenAI maybe ideal for knowledge discoveries and sharing, but it should be assessed for its efficiency and accuracy.

2.2. FOL/SWT Models vs. GenAI

It is very difficult to compare reasoning with FOL/SWT and text generation with ChatGPT for knowledge discoveries, because their differences are significant. It is not only that they differ in the type of inference. These are two different types of software technologies which could not fit any benchmarking, because they work on different principles. One creates text based on the most plausible guesses on "which word comes next" and the other is using semantic overlapping between textual terms to create results in knowledge discoveries through logic reasoning. However, they have something in common which is not always obvious. They both create questions/answers situations where we impose questions, and the technology produces answers.

FOL/SWT models secure logic reasoning to answer *competency questions* upon SWT models. Competency question is a term coined by SWT standards and they are essential in defining SWT models. In other words, before we create SWT models we must have clearly defined competency questions, and logic inference brings answers to these questions.

GenAI and ChatGPT is not far away from this principle. We have questions and ChatGPT is expected to give us the answers.

This is the only common ground we may have when using both, FOL/SWT and ChatGPT. These two technologies might be impossible to compare, but they may either complement or replace each other. What is important to note is that we must achieve knowledge discovery and dissemination and the answers we have been able to get with FOL/SWT models, are also expected from ChatGPT.

2.3. The Methodology

To illustrate the way of assessing ChatGPT for its efficiency and accuracy in knowledge discoveries we used a methodology suitable for biomedical scientists. This would mean that prompting would be defined with terminologies used by the biomedical profession. Also, considering that we will ask ChatGPT the same questions as competency questions defined for FOL/SWT logic reasoning, then ChatGPT prompting shouldn't deviate from them.

ChatGPT is constantly trained and tuned but might not have the latest biomedical research result incorporated. Therefore, we must make sure that we ask questions which do have answers from the corpus of knowledge created in the last decade and exclude everything created in the last 6 months. Finally, any additional prompting should be done with one goal in mind: we need a fast or maybe instant answer to our question(s) which could be close to answers obtained with logic reasoning, and thus excessive prompting might be counter-productive.

To summarise the methodology will allow to

a) Define competency questions in FOL/SWT model, illustrate logic reasoning, run the reasoning rules and show their results, which are answers to competency questions.
b) Take the competency questions from a) and ask ChatGPT for its answers. If they are not what we expected, then a moderate prompting may be required to get them as close to answers from a) as possible.
c) The choice of prompts should be determined by a biomedical scientist, because general prompt engineering skills might be counterproductive in attempts to replace FOL/SWT reasoning with ChatGPT answers.

2.4. Related work

At the time of writing this paper there were no publications which compared results of logic reasoning in knowledge discoveries with the results given by GenAI. However, we have found that logic reasoning attracted some attention of authors working with GenAI. These authors proposed the synergy of

a) logic reasoning through FOL implemented with the SWT languages, and
b) inference given by LLM prevalent in GenAI. for many reasons.

In (Breit et al., 2023) the authors give a survey of applications in which intelligent systems combine learning technologies and Machine Learning (ML) with techniques developed by the SWT community, for resolving a range of domains specific problems. In (d'Amato, C., 2020) ML has been introduced in the Semantic Web for solving problems such as link and type prediction, ontology enrichment and completion. In (He, et al., 2023) the authors explore LLMs for ontology alignment and (Kido, H. 2023) explores data generated models of formal logic and claims that probabilistic reasoning on logical formulae is a reasonable alternative to a logical consequence relation.

3. FOL and SWT for Biomedical Knowledge Discovery

Biomedical knowledge discovery using a logic reasoning process is shown in Figure 1 and adapted from (Almami et al., 2016a), .(Almami, 2017) Software technologies for implementing the reasoning process consists of SWT modelling concepts and languages OWL and SWRL. Therefore, the reasoning process from Figure 1 uses SWRL enabled OWL ontologies and gives answers to competency questions. It can be deployed as a computational model and support the discovery of information and data in biomedicine, which otherwise might not have been accessible or available. At the same time, semantic overlapping, as a backbone of knowledge description in SWRL enabled OWL ontologies can assist in finding *"what is missing in biomedical research"* and potentially direct us towards new discoveries (Juric et al., 2018). For example, this can become a springboard for personalising treatments of chronic diseases, drug discoveries and their repurposing (Juric et al., 2023).

The reasoning process in Figure 1 is specifically tailored to answer competency questions related to

"***activation/inhibition of enzymes which can control the development of diseases".***

The concepts, such as ontological classes PATIENT, DISEASE, ENZYME and HEALTH MANIFESTATION, have been populated with individuals from biomedical literature/research relevant for enzyme activation/inhibition and focused on ***Transglutaminase 2 enzyme*** because of its implication in the pathogenesis of human diseases.

3.1. Experiment 1

In the first experiment, our logic reasoning should be able to answer the following competency question

"*Which enzyme should be either activated or inhibited for a person who might have Alzheimer's disease (we are not sure yet) but the person experiences memory loss and the doctor suspects brain cell damaged after examining the patient's scan. The activated/inhibited enzyme could help in the control of the progress of the disease*".

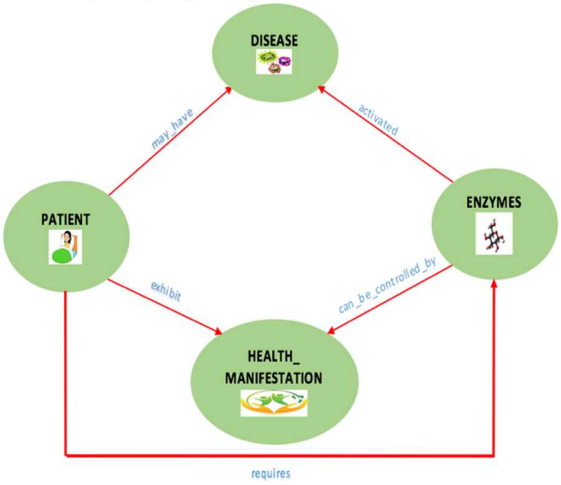

Figure 1. The reasoning process for answering the competency question

These main ontological classes from Figure 1, relate to each other through object properties (named red unidirectional lines). The logic inference created by the process in Figure 1, which can answer the competency question above, is in the format of the **object property named REQUIRES** (thicker red arrow). In other words, the reasoning model from Figure 1 infers relationship between individuals of PATIENT and ENZYME classes.

Figure 2 shows the reasoning, i.e. inference of object property REQUIRED has been programmed in SWRL language and it corresponds directly to the semantic of the competency question.

Figure 3 shows the result of the SWRL rule from Figure 2, which answers the competency question. Patient BOB who has exhibited memory loss and brain cell damage and might have Alzheimer's, and would benefit from the activation/inhibition of enzyme TG_2_(tissue_TG_tTG_cTG)'. The inference in this logic reasoning is the relationship (object property) between Bob and enzyme TG_2.

3.2. Experiment 2

In the second experiment, our logic reasoning should be able to answer this competency question:

"Which MEDICATION should a particular PERSON take if doctors suspect Alzheimer's, because

the patient health problems include memory loss and brain cell damage?"

Answers to this question will discover if there exists a medication for the specified health problems. If it does, the medication will be listed.

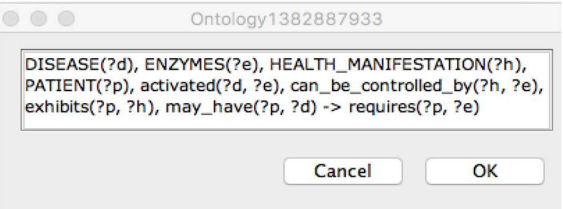

Figure 2. Reasoning rule for answering competency question from experiment 1

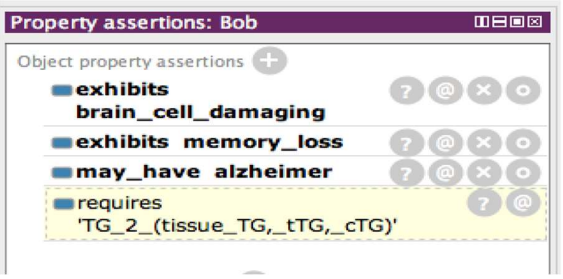

Figure 3. Results of reasoning when answering the question

The experiment uses OWL classes and object properties as defined in Figure 1 (with added MEDICATION class) but the SWRL coding in Figure 4 infers a different object property. This time inference is between individuals of PATIENT and MEDICATION classes.

The result of this reasoning, i.e. the answer to the competency question above, is given in Figure 5.

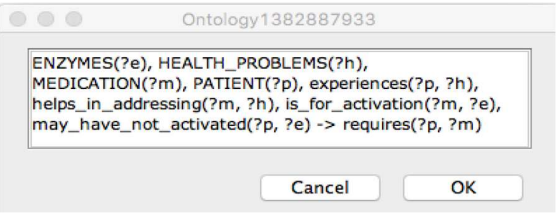

Figure 4: Screenshot of SWRL Rule for the question from experiment 2

The message from experiment 2 is clear: SWRL rule from Figure 4 did not return the name of any medication suitable for the patient. In other words there is no information in the model which can create semantic overlapping between PATIENT and MEDICATION. However, Figure 5 does mention that TG_2 enzyme might not have been activated for the patient, but there is no indication that the process can recommend the medication. Information available in the answer from Figure 5 refers to a chemo-biological, process within the brain tissue.

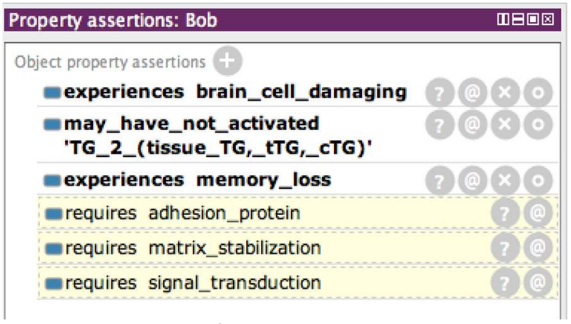

Figure 5: Results of reasoning as an answer to the question from experiment 2

The lack of semantic overlapping in this example would indicate that either

- there is NO enzyme which could be activated/inhibited for addressing the disease, or
- there is no medication discovered for activated/inhibited enzyme (TG_2 in our case).

The lack of results is as important as having the result, because it may trigger more research in the field where semantic overlapping is missing and create initiatives for drug discoveries in our example..

4. ChatGPT for Biomedical Knowledge Discovery

In this section we take competency questions from section 3 and use them in ChatGPT.

We start with Table 1 and question 1, which corresponds to competency question from experiment 1 in section 3. The question is placed in the left column of Table 1. The answer to question 1 in Table 1 is placed in the right column of the table.

We expect that the ChatGPT would answer something in line with

"the patient would benefit from the activation/ inhibition of enzyme TG_2_(tissue_TG_tTG_cTG)'.

Which means that ChatGPT should identify TG_2 enzyme in the answer to question 1.

The lack of identifying TG_2 enzyme requires additional prompting which is dictated by both: the ChatGPT answer to previous questions and biomedical scientist's knowledge and goals: how do I find answer to my competency question 1, if ChatGPT does not deliver? Therefore question 2 in Table 1 is reasonable:

"Can you find any other enzyme which can be associated with Alzheimer's disease?"

Table 1. Answers to questions 1,2 and 3

Questions 1,2,3	ChatGPT answer Excerpts
Question 1 Which enzyme should be activated for a person who might have Alzheimer (we are not sure yet) but he experienced memory loss and suspected brain cell damaging after examining the patient's scan. The activated enzyme could help in the controlling of the progress of the disease	For a person suspected of having Alzheimer's disease, one enzyme that has been identified as potentially beneficial to activate is sirtuin 1 (**SIRT1**). Benefits are Reducing Amyloid-beta Production, Enhancing Neuronal Survival, Improving Cognitive Function and Reducing Inflammation and Oxidative Stress
Question 2 Can you find any other enzyme which can be associated with Alzheimer's disease?	Certainly! Another enzyme might be considered for therapeutic intervention is beta-secretase 1 **(BACE1)**. It is involved in the production of amyloid-beta peptides. Other Enzymes Associated with Alzheimer's Disease are **Gamma-Secretase, Acetylcholinesterase (AChE), Glycogen Synthase Kinase-3 (GSK-3), Caspases.**
Question 3 What do SIRT1, BACE, gamma-secretase, acetylcholinesterase, GSK-3, and caspases have in common	SIRT1, BACE1, gamma-secretase, acetylcholinesterase, GSK-3, and caspases have several commonalities Involvement in Alzheimer's Disease Pathophysiology, Potential Therapeutic Targets, Regulation of Key Cellular Processes Complex Regulation and Multifunctionality Each enzyme has a complex regulatory mechanism and multiple roles in different biological pathways, reflecting their multifunctionality in cellular physiology.

Prompts and their answers listed in tables 1 and 2 show how difficult is for ChatGPT to give the expected answer. Most surprising is the answer to question 5 where ChatGPT had an opportunity to list all enzymes relevant to the Alzheimer's disease pathology and TG_2 has still not been identified.

Table 2. Answers questions 4,5,6

Question 4,5,6	ChatGPT answers Excerpts
Question 4 Are you sure that you mentioned all enzymes involved with Alzheimer's disease?	The enzymes mentioned—SIRT1, BACE1, gamma-secretase, acetylcholinesterase, GSK-3, and caspases—are some of the primary enzymes involved in Alzheimer's disease pathology. Here are some additional enzymes and proteins involved in Alzheimer's disease: Alpha-Secretase (ADAM10), Neprilysin, Insulin-Degrading Enzyme (IDE), Matrix Metalloproteinases (MMPs),. Cathepsins, Pin1 (Peptidyl-Prolyl Isomerase). Presenilins (PSEN1 and PSEN2), Lysosomal Enzymes, Apolipoprotein E (ApoE), Tau Kinases (e.g., CDK5, MARK)
Question 5 Could you list ALL enzymes involved in the Alzheimer's disease pathology	Certainly! Here is a comprehensive list of enzymes involved in Alzheimer's disease pathology, categorized by their primary roles and functions within the disease mechanism: **Enzymes Involved in Amyloid-beta Production and Clearance, in Tau Pathology in Neurotransmitter Regulation in Apoptosis and Neuroinflammation, in Cellular Metabolism and Stress Response** Continues in Table 4.
Question 6 Which enzyme is associated with Aβ deposits and lesion-associated astrocytes	Glutaminyl Cyclase (QC) plays a significant role in the pathology of Alzheimer's disease by modifying amyloid-beta peptides and in lesion-associated astrocytes, which are reactive glial cells found near amyloid plaques, contributing to the disease's progression and the inflammatory response

Only answer to question 7 in Table 3 identifies TG_2. The question is formulated by the biomedical scientist to test options for getting TG_2 included in any answer. Also, it is taken from the common

biomedical knowledge on enzymes and brain physiology and thus possible co-relation to the Alzheimer's disease pathology.

There are two remarks. First: ChatGPT answers in Tables 1 and 2 are not incorrect. They are perfectly formulated and descriptive, but not precise and completed. Second: answer to question 8 (which is identical to question 5) still does not include TG_2, even after ChatGPT's inclusion of TG_2 in the answer to question 7. Therefore question 9 which asks "have you forgotten to mention TG_2" is unavoidable.

Table 3. Answers to questions 7,8,9

Question 7,8,9	ChatGPT answers Excerpts
Question 7 Which enzyme can covalently cross-link neurofilament proteins into insoluble polymers which block neurotransmitter signals	The enzyme **transglutaminase 2 (TG2)** can covalently cross-link neurofilament proteins into insoluble polymers, which can block neurotransmitter signals. This activity of TG2 is associated with the pathogenesis of Alzheimer's disease and other neurodegenerative disorders, where it contributes to the formation of protein aggregates and neuronal dysfunction. TG2 contributes to the formation of protein aggregates and neuronal dysfunction.
Question 8 (it is the same as question 5 – repeated) Could you list all enzymes involved in the Alzheimer's disease pathology	**Answer is the same as the answer to question 5. TG2 has NOT been mentioned.**
Question 9 Have you forgotten to mention transglutaminase TG2?	No, I did mention **Transglutaminase 2 (TG2)** in the previous list. To ensure clarity and completeness, here is the section specifically related to TG2: The section/list is available in Table 5

4.1. Comments on Manual Prompting

The manual prompting from Tables 1 and 2 are managed by the biomedical scientist who avoided mentioning TG_2 in prompts, hoping for ChatGPT's "reaction" and inclusion of TG_2 in its answers.

Question 6 was our first attempt to come closer to the role of TG_2 in prompting, considering that

a) question 5 produced the answer which mentions "**Enzymes Involved in Neurotransmitter Regulation"** (see table 4) and
b) TG_2 is an enzyme which can block neuro transmission, which is well documented in the literature (Odi et al., 2014),

Asking about "Aβ deposits and lesion-associated astrocytes" was a natural step forward in prompting. However, its answer still did not include TG_2.

Finally in question 7 the biomedicals scientist used **the wording available in one of our publications** (and in any similar publication available on the Internet), which says that

"TG_2 enzyme can covalently cross-link neurofilament proteins into insoluble polymers and block neurotransmitter signals"

and converted the statement above into question 7. We finally received the answer we expected.

In question 8 we wanted to check if ChatGPT has learned anything from our prompting and include TG_2 whenever we asked for the list of enzymes relevant for the management of Alzheimer's disease. It appears that this was not the case and thus the question:

Could you list ALL enzymes involved in the Alzheimer's disease pathology

It was when we asked question 9 "Have you forgotten to mention transglutaminase TG2?" that the Question 1 in Table 1 was answered by ChatGPT answer.

Table 4. Continuation of the answer to Question 5

Certainly! Here is a comprehensive list of enzymes involved in Alzheimer's disease pathology, categorized by their primary roles and functions within the disease mechanism: **Enzymes Involved in Amyloid-beta Production and Clearance are** Beta-Secretase 1 (BACE1), Gamma-Secretase Complex, Alpha-Secretase (ADAM10),Neprilysin, Insulin-Degrading Enzyme (IDE), Matrix Metalloproteinases (MMPs), Cathepsins, **Enzymes Involved in Tau Pathology are** Glycogen Synthase Kinase-3 (GSK-3) Cyclin-Dependent Kinase 5 (CDK5), Microtubule Affinity-Regulating Kinase (MARK), Protein Phosphatase 2A (PP2A), Pin1 (Peptidyl-Prolyl Isomerase), **Enzymes Involved in Neurotransmitter Regulation are** Acetylcholinesterase (AChE), Butyrylcholinesterase (BuChE), **Enzymes Involved in Apoptosis and Neuroinflammation are** Caspases (e.g., Caspase-3, Caspase-9)NADPH Oxidase Inflammasomes (e.g., NLRP3) **Enzymes Involved in Cellular Metabolism and Stress Response** Sirtuin 1 (SIRT1), AMP-Activated Protein Kinase (AMPK) Lysosomal and Proteolytic Enzymes Lysosomal Enzymes (e.g., Cathepsins) , Ubiquitin-Proteasome

5. Debate and Conclusions.

This paper attempts to bring more light into the world of GenAI by looking at one of its more lucrative outcomes: generating new text which could be interpreted as knowledge discovery. There is no doubt

that GenAI and ChatGPT can do that, but the experiments from this paper left the authors disappointed. ChatGPT did not deliver expected answers to the questions and did not contribute towards knowledge discovery the way it was expected.

One of the goals in section 2.1 is to find out if we can achieve knowledge discovery and dissemination by using ChatGPT only (i.e. avoiding the tested logic reasoning). In this particular example the answer is NO because it is still disappointing that the well-researched enzyme TG_2, available across the vast biomedical literature and subject of research for decades, has not been identified by ChatGPT as enzyme associated with neurobiological brain conditions.

However, our answer to the goal to find out if *ChatGPT can cut preparation time in knowledge discoveries, answer ad-hoc questions we may have, and thus secure knowledge sharing* is probably YES. It took just 15-20 minutes and biomedical expertise knowledge to think about prompting in Tables 1,2 and 3, which is much faster compared to the months spent on modelling and running the reasoning process in Figure 1.

Unfortunately, we do not have a definitive answer to our next goal form section 2.1: *Can ChatGPT deliver fast. i.e. instant results which are similar to running computations with FOL/SWT reasoning in knowledge discoveries?* If we are looking for "similar results" then YES, but if we wish to have a piece of exact information in biomedical science, then NO. Shall we in future try to use GenAI first and then check results with the FOL/SWT reasoning models? This might depend on many factors, but if time is NOT an issue then the answer is YES.

5.1. Lessons learned and future work

It is difficult to evaluate GenAI technology in the world of text manipulation with the constant updating, tuning and training of its data. It is also difficult to find a non-biased way to measure values of GenAI outcome. Our approach was dictated by our previous knowledge, expertise and persistence to get answers from ChatGPT which we expected. This is biased. Therefore, comparing plausible answers given by ChatGPT with facts discovered through logic reasoning, might not be the best possible way of "judging ChatGPT".

Table 5. Continuation of the answer to question 9

Including **Transglutaminase 2 (TG2)**, here is a final comprehensive list of enzymes involved in Alzheimer's disease pathology:

1. Beta-Secretase 1 (BACE1)
2. Gamma-Secretase Complex
3. Alpha-Secretase (ADAM10)
4. Neprilysin
5. Insulin-Degrading Enzyme (IDE)
6. Matrix Metalloproteinases (MMPs)
7. Cathepsins
8. Plasmin
9. Glycogen Synthase Kinase-3 (GSK-3)
10. Cyclin-Dependent Kinase 5 (CDK5)
11. Microtubule Affinity-Regulating Kinase (MARK)
12. Protein Phosphatase 2A (PP2A)
13. Pin1 (Peptidyl-Prolyl Isomerase)
14. Acetylcholinesterase (AChE)
15. Butyrylcholinesterase (BuChE)
16. Caspases (e.g., Caspase-3, Caspase-9)
17. NADPH Oxidase
18. Inflammasomes (e.g., NLRP3)
19. Sirtuin 1 (SIRT1)
20. AMP-Activated Protein Kinase (AMPK)
21. Lysosomal Enzymes (e.g., Cathepsins)
22. Ubiquitin-Proteasome System (UPS)
23. Transglutaminase 2 (TG2)
24. Glutaminyl Cyclase (QC)
25. Presenilins (PSEN1 and PSEN2)
26. Tau Kinases (e.g., Tau-tubulin kinase 1 - TTBK1)

This list includes all key enzymes and relevant proteins that play a significant role in the pathology of Alzheimer's disease, encompassing various aspects such as amyloid-beta production and clearance, tau pathology, neurotransmitter regulation, apoptosis, neuroinflammation, cellular metabolism, stress responses, protein degradation, and

In addition, it is very difficult to challenge logic reasoning and its results. The example from this paper shows that, in the case of enzymes and associated human pathology, ChatGPT answers are not complete. They are not wrong, and they make the point, but they have missing information within their answers. We talk here about well-known research on TG_2 and its role in human pathology, which can be found in biomedical literature all over the Internet. For some reason it has not been easy for ChatGPT to identify the enzyme.

There are numerous ways of remedying the deficiencies of GenAI. They are listed below:

- We may have planned our prompting carefully and developed a special mechanism of asking questions, which could be used for ChatGPT learning. Prompt engineering may be the way forward in this case.
- We could have used RAG in association with prompt engineering. There is a real chance that we could have found databases with enough information on TG_2 and its relationship with human pathology, which would be used in answering ChatGPT questions.

- We could set up a set of GenAI agents which could be used (and trained) on a specific data related to the "role of enzymes in the management of neurological diseases". There are tools and interactive development environments which enable the creation of GenAI agents.
- Considering the size of ChatGPT and its complicated and unsustainable tuning, we might re-think small language models, instead of sticking to LLM. We could create an environment where small language model will thrive and cover all semantic of the knowledge in biomedical science, answer all our questions and thus enable knowledge discovery as we expected.
- Finally, we should look at the latest initiative from META and their brave departures form LLM (LeCun, Y. 2024). Their vision is to make AI systems learn and reason like animals and humans. Would this mean that we should come closer to logic reasoning in managing the content from GenAI? What should be used for reasoning?

In section 4 manual promoting is underpinned with our understanding of the role of enzymes in the treatment of diseases. The manual generation of prompts could have been automated by software, and this type of automation could use FOL and logic reasoning, to determine which prompt could be used, when and why (Juric et al., 2023).

However, one question remains unanswered: How important is it to receive an instant, complete and correct answer by ChatGPT? If we wish to talk about important and urgent knowledge discoveries in emergencies and human catastrophes, if we are in urgent need of discoveries of drug/vaccines, if we must define new treatments of complicated and devastating diseases, then YES, we will need a very accurate and instant answer from GenAI.

In section 3 we illustrated an example of logic reasoning which gave us instant answers when we needed them. The computational model and reasoning in FOL took a long time to develop and implement with SWT, but with its low maintenance the software solution based on logic reasoning would have a very long shelf life and will be seen as a reusable software model (Juric, 2017). When using ChatGPT for asking questions, as in section 4, we needed no preparation and ChatGPT was ready to deliver the best/most plausible answer, which might not be the answer we expected, but with prompting, we may get closer to it.

7. References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., & Bambrick, J. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 1-3.

Ahmed, A., Hou, M., Xi, R., Zeng, X., & Shah, S. A. (2024, May). Prompt-Eng: Healthcare Prompt Engineering: Revolutionizing Healthcare Applications with Precision Prompts. In Companion Proceedings of the ACM on Web Conference 2024 (pp. 1329-1337).

Almami, E., Ahmed, M.Z. Juric, R. (2016) Software Applications Built Upon SWRL Enabled OWL Ontologies, 21st SDPS 2016 Int. Conf. 2016.

Almami, I., Almami, E., & Juric, R. (2016a) Knowledge Dissemination In Biomedical Science: Using Ontological Reasoning For The Analyses Of Activated Enzymes In Various Diseases In Proceedings of the 21th International Conference on SDPS '16, Florida

Almami, Eiman (2017) A Reference Computational Model to Manage the Semantics of Learning Environments using SWRL Enabled OWL Ontologies, PhD Thesis, University of Plymouth, UK. Available at https://pure.plymouth.ac.uk/ws/portalfiles/portal/38449798/2017almami10425118phd.pdf

Almeida, J.B., Frade, M.J., Pinto, J.S., Melo de Sousa, S. (2011). First-Order Logic. In: Rigorous Software Development. Undergraduate Topics in Computer Science. Springer, London.

Beltran, M. A., Ruiz Mondragon, M. I., & Han, S. H. (2024, June). Comparative analysis of generative ai risks in the public sector. In Proceedings of the 25th Annual International Conference on Digital Government Research (pp. 610-617).

Bi, Z., Dip, S. A., Hajialigol, D., Kommu, S., Liu, H., Lu, M., & Wang, X. (2024). AI for Biomedicine in the Era of Large Language Models. arXiv preprint arXiv:2403.15673.

Breit, A., Waltersdorfer, L., Ekaputra, F. J., Sabou, M., Ekelhart, A., Iana, A.,Teije, A. t. (2023). Combining machine learning and semantic web: A systematic mapping study. ACM Computing Surveys, 55(14s), 1-41.

Chaikinn P. (2023) The Unpredictable Abilities Emerging From Large AI Models, article in Quanta Magazine, available at https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/

d'Amato, C. (2020). Machine learning for the semantic web: Lessons learnt and next research directions. *Semantic Web*, *11*(1), 195-203.

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A. (2022) Empowering Biomedical Discovery with AI Agents, preprint available at https://arxiv.org/pdf/2402.06196v2

Genkina, D. (2024). AI Prompt Engineering Is Dead Long Live AI Prompt Engineering. Spectrum IEEE. https://spectrum.ieee.org/prompt-engineering-is-dead

He, Y., Chen, J., Dong, H., & Horrocks, I. (2023). Exploring large language models for ontology alignment. arXiv preprint arXiv:2309.07172.

Juric, R., Almami, E., & Almami, I. (2018). Semantic Overlapping of OWL Models in Biomedicine. proceedings of the SDPS 2018 Workshop on Accountability in AI.

Juric, R. (2016). Could Semantic Web Technologies Create New Computational Models Outside Semantic Web, 21st SDPS 2016 Conference.

Juric, Radmila; Almami, Eiman; Almami, Btesam; and Ahmed, Zaki, "Semantic Computational Models for

Polypharmacology: Applications in Drug Repurposing" (2023). *Hawaii International Conference on System Sciences 2023 (HICSS-56)*. 3.

Juric, R., Williams, E., & Kim, I. (2020). Semantic Overlapping in Translational Bioinformatics Applied to the Matching between Clinical Trial Eligibility Criteria and Patient Needs. *Bioinformatics*.

Juric, R., & Almami, E. (2019). Proposing Drug Repositioning through Semantic Reasoning. proceedings of the 24th Conference on SDPS.

Jin, Q., Leaman, R., & Lu, Z. (2024). PubMed and beyond: biomedical literature search in the age of artificial intelligence. Ebiomedicine, 100.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Jungwirth, D., Haluza, D. (2023) Artificial Intelligence and Public Health: An Exploratory Study **International Journal of Environmental Research and Public Health.*

Kido, H. (2023). A Simple Generative Model of Logical Reasoning and Sta Learning. arXiv:2305.11098.

Li, Y., Huang, Y., Ildiz, M. E., Rawat, A. S., & Oymak, S. (2024). Mechanics of next token prediction with self-attention. International Conference on Artificial Intelligence and Statistics,

LeCun, Y. (2022, February 23). Yann LeCun on a vision to make AI systems learn and reason like animals and humans. AI at Meta. Retrieved September 01, 2024, from https://ai.meta.com/blog/yann-lecun-advances-in-ai-research/

Lin, Z., Akin, H., Rao1, R., Hie1, B., Zhu, Z., Lu, W., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., (2022) Language models of protein sequences at the scale of evolution enable accurate structure prediction, preprint at biorxiv https://www.biorxiv.org/content/10.1101/2022.07.20.500902v1.full.pdf

Mostert, B. (2024, January 23). Announcing the OCI Generative AI Agents RAG service. Oracle Blogs. Retrieved June 01, 2024, from https://blogs.oracle.com/ai-and-datascience/post/oci-generative-ai-agents-rag-service

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., Gao, J. (2024) Large Language Models: A Survey, preprint available at https://arxiv.org/pdf/2402.06196v2

Odii, B. O., & Coussons, P. (2014). Biological functionalities of transglutaminase 2 and the possibility of its compensation by other members of the transglutaminase family. The scientific world journal, 2014(1), 714561.

Poon, H., Naumann, T., Zhang, S., & González Hernández, J. (2023). Precision Health in the Age of Large Language Models. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,

Preiß, Jennifer; von Brackel-Schmidt, Constantin; and Leible, Stephan, "Student Perspectives on Generative Artificial Intelligence: Exploring Pre-Framing and Risks in Higher Education" (2024). ECIS 2024 Proceedings. 15.

Raiola, R. (2023) Future Tense ChatGPT, Can You Tell Me a Story? An exercise in challenging the true creativity of generative, CACM May 2023

Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. (2024). Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies and Limitations. IEEE Access

Shanahan, M. (2024). Talking about large language models. Communications of the ACM, 67(2), 68-79.

Shojanoori R., Juric, R., Lohi, M (2012) Computationally Significant Semantics in Pervasive Healthcare, Journal of Integrated Design and Process Science, IOS Press, Volume 16, Number 1 – 2012, pp 43-62.

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S.R., Cole-Lewis, H.J., Neal, D., Schaekermann, M., Wang, A., Amin, M., Lachgar, S., Mansfield, P.A., Prakash, S., Green, B., Dominowska, E., Arcas, B.A., Tomašev, N., Liu, Y., Wong, R.C., Semturs, C., Mahdavi, S.S., Barral, J.K., Webster, D.R., Corrado, G.S., Matias, Y., Azizi, S., Karthikesalingam, A., & Natarajan, V. (2023). Towards Expert-Level Medical Question Answering with Large Language Models. *ArXiv, abs/2305.09617*.

Tian, S., Jin, Q., Yeganova, L., Lai, P., Zhu, Q., Chen, X., Yang, Y., Chen, Q., Kim, W., Comeau, D.C., Islamaj, R., Kapoor, A., Gao, X., & Lu, Z. (2024). Opportunities and Challenges for ChatGPT and Large Language Models in Biomedicine and Health. *Briefings in Bioinformatics*, Volume 25, Issue 1, January 2024, bbad493

ACM-USTPC, (2023, June 27). Principles For The Development, Deployment, And Use Of Generative AI Technologies, In: ACM-USTPC GenAI Principles, available at https://www.acm.org/binaries/content/assets/public-policy/ustpc-approved-generative-ai-principles

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 systems*, 30.

Wooldridge, M. (2023) The future of generative AI, Royal Institute Turing Lectures, London. UK https://www.youtube.com/playlist?list=PLuD_SqLtxSdVDcrCYIHayTL91DapuIHrO

W3C, (2004) World Wide Web Consortium: Semantic Web "Layer Cake", https://www.w3.org/2004/Talks/0412

W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview (2012), http://www.w3.org/TR/owl2-overview/

Zhang, K., Yu, J., Yan, Z., Liu, Y., Adhikarla, E., Fu, S., Chen, X., Chen, C., Zhou, Y., Li, X., He, L., Davison, B.D., Li, Q., Chen, Y., Liu, H., & Sun, L. (2023). BiomedGPT: A Unified and Generalist Biomedical Generative Pre-trained Transformer for Vision, Language, and Multimodal Tasks. *ArXiv, abs/2305.17100*