How Large Language Models Translate Raw Data into Expert Rules

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

ABSTRACT

The research highlights how large language models (LLM) can transform technical knowledge into a formal ontology, then score the output as an expert system. Traditional ontology-builders support expert systems that humans can audit and explain the rules for confident predictions. A shortcoming of building ontologies from scratch involves the initial knowledge transfer from experts and curating the generation of conditional rules as decision trees. The research experimentally examines whether converting unstructured data into structured candidates accelerates rule extraction. The study extends conversion beyond entity extraction for people, persons, organizations, or intent prediction for motivations or risk management. Our work highlights the historically significant DoD challenges spearheaded by investments in massive symbolic artificial intelligence projects like Cyc (similar to enCYClopedia). The original ontology commitment of labor alone exceeded 1000- 3000 person-years of effort "to describe how the world works." The latest LLMs (e.g., OpenAI's GPT-3, Google's PaLM) typically encode 40 terabytes of the world's (internet) knowledge and provide convenient question-and-answer and application programming interfaces (APIs) that can export ontology-ready rules and semantic relationships to support deterministic expert systems. The research examines the scalability of this approach for examples taken from classic military training problems, such as building a predictive cyber threat actor model for nation-states. We design advanced decision aids for fusing situational and threat awareness and analytical dashboards. We evaluate these case studies for the LLMs and human questioners who extract artifacts for building expert systems. This process potentially solves the lack of reliability from existing LLMs and human feedback to solve the otherwise intractable needs for predictable medical or combat decision-makers. If the LLMs distill the world knowledge, human inquisitors distill the expert artifacts in reliable and testable ways that remove indeterminism from existing dialog generators.

ABOUT THE AUTHORS

David Noever has 27 years of research experience with NASA and the Department of Defense in machine learning and data mining. He received his Ph.D. from Oxford University, as a Rhodes Scholar, in theoretical physics and his B.Sc. from Princeton University, summa cum laude, and Phi Beta Kappa. While at NASA, he was named 1998 Discover Magazine's "Inventor of the Year" for the novel development of computational biology software and internet search robots, culminating in co-founding the startup company cited by Nature Biotechnology as first in its technology class. He has authored more than 100 peer-reviewed scientific research articles and book chapters. He also received the Silver Medal of the Royal Society, London, and is a former Chevron Scholar, San Francisco. His primary research centers on machine learning, algorithms, and data mining for analytics, intelligence, and novel metric generation.

J. Wesley Regian has 32 years of experience in cognitive performance modeling and knowledge-based software technology development, primarily for military applications with AFRL, AFOSR, and DARPA. His work has supported over 50 fielded systems. He has published over 100 papers on intelligence analysis, human terrain modeling, knowledge representation, knowledge management, human learning and memory, individual and developmental differences in human cognition, spatial ability and spatial information processing, cognitive modeling, skill acquisition, componential analysis of spatial tasks, cognitive automaticity, psychometrics, artificial intelligence, hypertext, hypermedia, training, computer-based training, intelligent computer-based training, virtual reality, and multi-source intelligence fusion. Dr. Regian was a National Research Council research adviser for ten years and Senior Scientist for Knowledge-Based Systems at the US Air Force Armstrong Research Laboratory.

How Large Language Models Translate Raw Data into Expert Rules

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

INTRODUCTION

Expert systems, which mimic human decision-making and provide authoritative answers, have become critical across various fields. In military training, these systems function as intelligent tutors, providing personalized learning experiences based on a student's learning style and pace (Alexopoulos et al., 2023). Within cybersecurity, military analysts use expert systems for intrusion detection, malware identification, and threat analysis (Boury-Brisset, 2004; Council on Foreign Relations, 2023). These systems are vital in equipment maintenance, optimizing production processes, improving quality, and reducing downtime (Bowman, Lopez, & Tecuci, 2001).

Doctors and field medics use expert systems for diagnosis, treatment planning, and drug prescription (Hu, Pan, Li, & Yang, 2023). These applications highlight the need for informed decisions and reliable outcome forecasts, mainly when dealing with complex decisions, outliers, or "edge cases" (Bubeck et al., 2023).

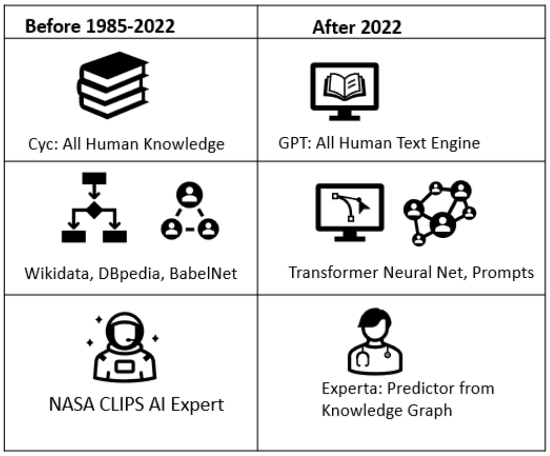

Before 1985-2022
After 2022
Cyc: All Human Knowledge
GPT: All Human Text Engine
Wikidata, DBpedia, BabelNet
Transformer Neural Net, Prompts
NASA CLIPS AI Expert
Experta: Predictor from Knowledge Graph

Figure 1. Evolution of Knowledge-Based AI to LLM Expert Systems

More general machine intelligence represents a new frontier in AI research which seemed decades away as recently as 2022. However, the latest Large Language Models (LLMs), artificial intelligence models trained on extensive text data, can translate raw data into expert rules, enhancing rule-based systems (Chang, 2023). By sifting through vast data sets, LLMs extract relevant information and codify it into actionable rules, potentially boosting system robustness, adaptability, and predictive capability. However, these approaches also introduce new challenges, including system design complexity ("black box" methods), the risk of biases, and the need to maintain interpretability (Bubeck et al., 2023; Radford et al., 2019). Failed applications of LLMs focus on their known weaknesses in math and computation, privacy, life-changing advice in consequential legal or health matters, trust, and reliability. Interestingly, these known LLM weaknesses combine symbiotically with the strengths of expert systems, notably if their reasoning engine offers transparency and auditable validation of its knowledge base and explainable rules.

How can Large Language Models (LLMs) translate raw data into expert rules, thus enriching rule-based systems? This question is at the crux of a new frontier in AI research, straddling the divide between data-driven and knowledge-based approaches. LLMs, as AI models typically built on transformer architectures (Vaswani et al., 2017), can predict words based on sentence context (Radford et al., 2019). They can extract and codify information from vast data into rules for expert systems, enhancing their robustness and adaptability (Trajanoska et al., 2023). Figure 1 schematically illustrates the evolution of expert systems before November 2022 breakthroughs in LLM models. Indeed, the process of leveraging LLMs to enhance expert systems, as demonstrated in the paper's use cases and examples, can be distilled into three key steps:

1. **Input Data**: The source material for training and applying the LLM is crucial. Whether it's medical literature, military manuals, or historical documents, the input data's quality, and diversity play a significant role in the resulting expertise of the model (Radford et al., 2019). Figure 2 shows a knowledge graph describing entities and notable facts about their properties and tendencies.

2. **Prompt Engineering**: To effectively leverage the power of LLMs, it is essential to design and tune prompts adeptly. Phrasing the input involves shaping queries to guide the model toward the desired responses (Chang, 2023). Prompt engineering is a critical and iterative skill in harnessing LLMs for specific tasks.
3. **Output Interpretation**: The output from an LLM, whether it forms expert rules for an intelligent tutoring system, cyber threat detection model, or disease diagnostic system, must be in a helpful format to the downstream application. Formatting the desired knowledge representation could range from structured rule sets to more human-readable formats like markdown or decision-tree representations (Trajanoska et al., 2023; Bubeck et al., 2023). Figure 3 gives an example of describing the relationships between characters in the Napoleonic war saga, Tolstoy's half-million-word tome, War and Peace.

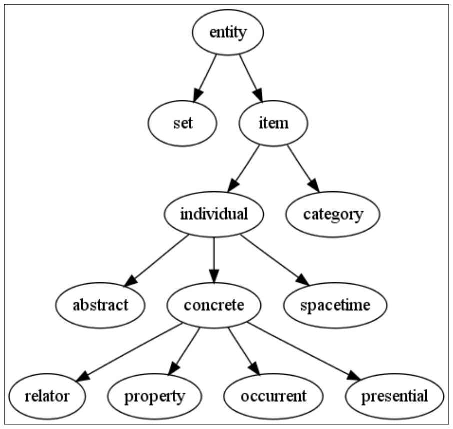


Figure 2. Ontology To Represent Knowledge as Properties and Fact Collections, connected to a Reasoning Engine to Make Predictions

These three input, prompt, and output steps are fundamental across applications and highlight the commonalities in how LLMs enhance rule-based systems.

Previous attempts to encode all human knowledge relied on the tedious assembly of all human encyclopedic knowledge ("Cyc" project, 1985), specialized encodings like Wikidata, and prediction engines like NASA CLIPS (Wygrant, 1989). As a primary AI data artifact, Cyc is an ambitious artificial intelligence project that aims to assemble a comprehensive ontology and database of human knowledge, common sense, and reasoning capabilities. Initiated by Douglas Lenat in 1984, the Cyc project underscores the endeavor to build a foundation of 'encyclopedic' machine knowledge that can be employed to make human-like decisions and inferences (Lenat et al., 1985). It represents a long-term attempt to create an AI that understands the world at a human level, with all the nuance, context, and common sense that such an understanding entails. Cyc aims to equip machines with a depth of knowledge and reasoning to interact meaningfully with the world and human users. Sarjant, Legg, Robinson, and Medelyan (2009) documented feeding Wikipedia, a vast unstructured knowledge source, into Cyc.

Further research by Matuszek, Witbrock, Cabral, and DeOliveira (2006) introduced the syntax and content of Cyc, shedding light on the complexities of large-scale knowledge base management. Unlike other AI systems, Cyc does not learn from data. Instead, human experts meticulously encode knowledge into Cyc through logical assertions, leading to a set of rules the system uses to infer new knowledge. The assertions encompass everything from basic facts and common sense to intricate, domain-specific expertise across various subjects. Despite the immense effort and progress, codifying all human knowledge is vast and complex. There are countless domains of knowledge to cover, each with unique and often nuanced concepts, facts, and rules. Much of human knowledge is contextual and varies with culture, time, and individual interpretation, making creating a universally applicable system challenging. There is also the issue of commonsense understanding, much of which humans take for granted but is

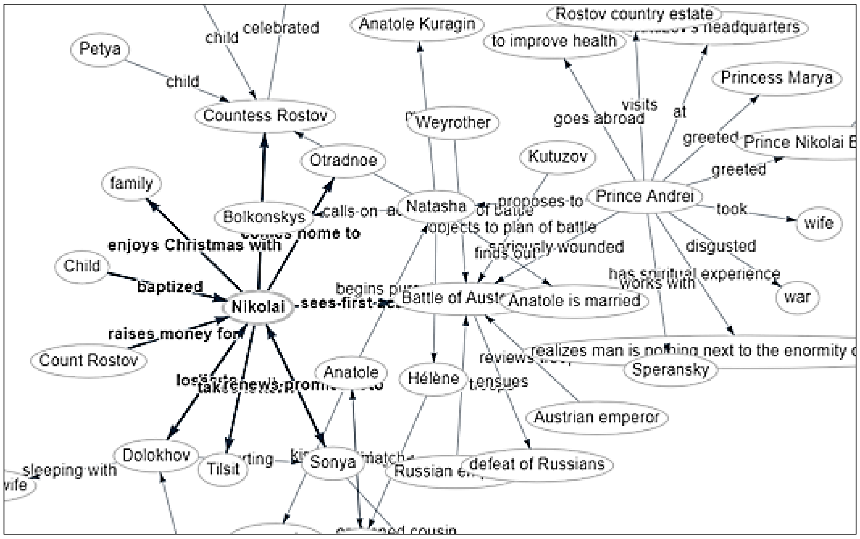


Figure 3. Example knowledge graph of complex entities and relationships that LLMs extract from 1225 pages of Tolstoy's tome, War and Peace, on the Napoleonic Wars

remarkably difficult to articulate and encode systematically. Representing physical commonsense challenges the builders of knowledge systems: conveying that one set of units is more significant than or less than a comparative metric, real-world responses like solids that float on water, or time elements like Wednesday last week appear on a calendar before Tuesday this week.

In contrast, the most recent expert systems automate the knowledge base (Bubeck et al., 2023), encoding human knowledge as a complex equation or weights (Radford et al., 2019) and inferring using scoring engines like Experta modules (Perez, 2023) to make predictions and forecasts. Massive symbolic artificial intelligence projects like Cyc (similar to enCYClopedia) committed the DoD and other federal agencies to a four-decade commitment of labor alone exceeding 1000- 3000 person-years of effort "to describe how the world works." Alternatively, the OpenAI development of its knowledge base (currently Generative Pre-trained Transformer-4) encoded almost 45 terabytes (TB) of the internet as written human text. OpenAI's approach essentially automates the aggregation of every word humans have written down since the days of papyrus (Sumerians around 3400 BC, Getty Center, 2021). Remarkably since November 29, 2022, the OpenAI text generator became the fastest-growing software in history (five times faster than TikTok) and now generates the entire human written historical record (45 TB) every 14 days and growing (Polymer, 2023).

The present work offers the following structure, starting with an in-depth introduction. The paper sets the context and states the research purpose. We follow this opening section with a literature review examining prior works on LLMs in expert systems and rule-based reasoning in AI. We then describe the methodology employed in the study, including using LLM, rule-based reasoning methodology, and generating and validating knowledge graphs. The paper outlines use cases in diverse fields to illustrate how it can support informed decisions with probabilities and consequences. Each use case defines a new problem statement, outlines the implementation of LLM, and reports results and critical analysis. The study also highlights the limitations of the current performance and suggests potential improvements. The work concludes with a comparative analysis, discussing the applicability and scalability of the method across various domains, supported by figurative representation of data and trends.

Background and LLM Context of the Study

Since 1956 when Alan Turing first posed the question, "Can a machine think?" the last seven decades have witnessed rapid advances followed by plateaus in research understanding. Researchers describe the conundrum as requiring an intersection of data, hardware, and the right software, with Harvard Business Review answering the AI question of "Why now?" as relying on 90% data, 9% hardware, and 1% software advances (Brynjolfsson and McAfee, 2017). The authors assert the correct data, hardware, and algorithms can "combine to generate improvements of up to a millionfold

Table 1. Experimental Research Questions and Hypotheses

Hypothesis	**Research Question**	**Potential Factors/Answers**
Large Language Models (LLMs) can enhance the explainability of rule-based expert systems.	How effectively can LLMs be integrated into rule-based expert systems to improve their explainability?	Evaluation metrics for explainability, domain-specific considerations, and the complexity of rules
LLMs can contribute to the development of anticipatory analytics in expert systems.	How can LLMs be leveraged to generate predictive insights in expert systems?	The efficiency of the model in prediction, relevance, and novelty of generated insights
Integrating LLMs into rule-based expert systems can be effective across different domains.	How does the effectiveness of LLMs in enhancing rule-based expert systems vary across different use cases, such as medical diagnosis, cyber threat attribution, and decision support systems?	Variance in performance across domains, context-specific requirements, and constraints
The insights generated by LLMs in rule-based expert systems can provide actionable and valuable information.	What are the practical utility and potential impact of the insights generated by LLM-enhanced expert systems in real-world scenarios?	Real-world applicability, decision-making impact, user feedback

in applications such as the pedestrian-detection vision systems used in self-driving cars." Table 1 summarizes the research program to merge LLMs with expert systems.

LLMs like GPT have shown promising capabilities in many areas. Alexopoulos et al. (2023) examined the use of ChatGPT in open government data. Lande and Strashnoy (2023) used ChatGPT for concept networking. In knowledge management, Lin et al. (2023) showed how ChatGPT could help with semantic database management. Kollapally et al. (2023) proposed a health determinant ontology. Nirenburg, Krishnaswamy, and McShane (2023) integrated machine learning with knowledge base systems. Radford et al. (2019) demonstrated LLMs could be used for unsupervised multitask learning. There are concerns with deploying LLMs in real-world scenarios. Au Yeung et al. (2023) questioned the readiness of AI chatbots, including LLMs, for clinical use. Chang (2023) used the Socratic method for more complex dialogues with LLMs. Deng (2023) proposed using NLP for web interactions. Auer et al. (2023) evaluated the scientific question-answering capabilities of LLMs.

Prior Research on Rule-Based Reasoning in AI

Rule-based reasoning in AI, foundational to expert systems, has a robust research tradition. For instance, Bowman, Lopez, and Tecuci (2001) delve into ontology development for military applications, illustrating early applications of rule-based reasoning in complex settings. Boury-Brisset (2004) showcases an ontological approach to military knowledge modeling, while Song, Ryu, and Kim (2010) advance ontology-based decision support in military systems.

Work on broader ontology building is also plentiful. Fensel (2001), Fernández-López and Gómez-Pérez (2002), and Semy, Pulvermacher, and Obrst (2004) offer comprehensive overviews of building ontologies, highlighting challenges and opportunities. Recently, Trajanoska, Stojanov, and Trajanov (2023) proposed improving knowledge graph construction with LLMs, marrying traditional rule-based reasoning and modern machine learning.

Our methodology leverages NASA's CLIPS engine (Wygant, 1989), a rule-based language for expert systems initially developed in the 1980s. Notable features include pattern matching, inferencing, procedural and fact-oriented programming, and uncertainty and conflict resolution handling. Despite its age, its powerful inference engine, portability, and supportive community make it a popular choice for expert systems. We utilize the Python version, Experta (Perez, 2023), for its convenience in AI rules engine applications for expert system development.

Previous Attempts at Using LLM For Anticipatory Analytics

Anticipatory analytics is an approach to data analysis that focuses on predicting future events, behaviors, and outcomes based on historical and real-time data. The goal is to identify trends, patterns, and correlations in existing data that may indicate or forecast future possibilities. The underlying principle of anticipatory analytics is to move beyond mere description and diagnostics of past and current events to provide actionable predictions. Predictions could range from predicting customer behavior for personalized marketing, identifying potential system failures in technology, forecasting economic trends, anticipating public health issues, or predicting geopolitical changes. Figure 3 shows an example decision tree for predicting mission success based on weather and terrain inputs to a model for command support.

Applying LLMs for anticipatory analytics is a relatively nascent field, but several pioneering studies have made noteworthy contributions. Starting with the work by Alexopoulos et al. (2023), they proposed a research agenda for applying ChatGPT, an LLM, vis-à-vis Open Government Data (OGD). Their work hints at the potential of using LLMs for predictive analytics based on publicly available data. Regarding engineering, Hu et al. (2023) discussed the opportunities and challenges of using ChatGPT for design knowledge management. Their work suggests that LLMs can foresee design problems and propose solutions. Lande and Strashnoy (2023) proposed an interesting approach using ChatGPT for concept networking. By extracting and visualizing concepts from LLM outputs, it might be possible to anticipate trends and changes in various domains.

Figure 3 highlights a model decision support task to predict whether logistical operations can succeed with current weather, terrain, and transport. Finally, Sima and de Farias (2023) work delves into the potential of AI chatbots for data exploration of federated bioinformatics knowledge graphs. This work indicates the possibility of using LLMs for anticipating insights from complex, interconnected datasets. These studies collectively illustrate the exciting potential and ongoing challenges of leveraging LLMs for anticipatory analytics. Anticipatory analytics can provide decision-

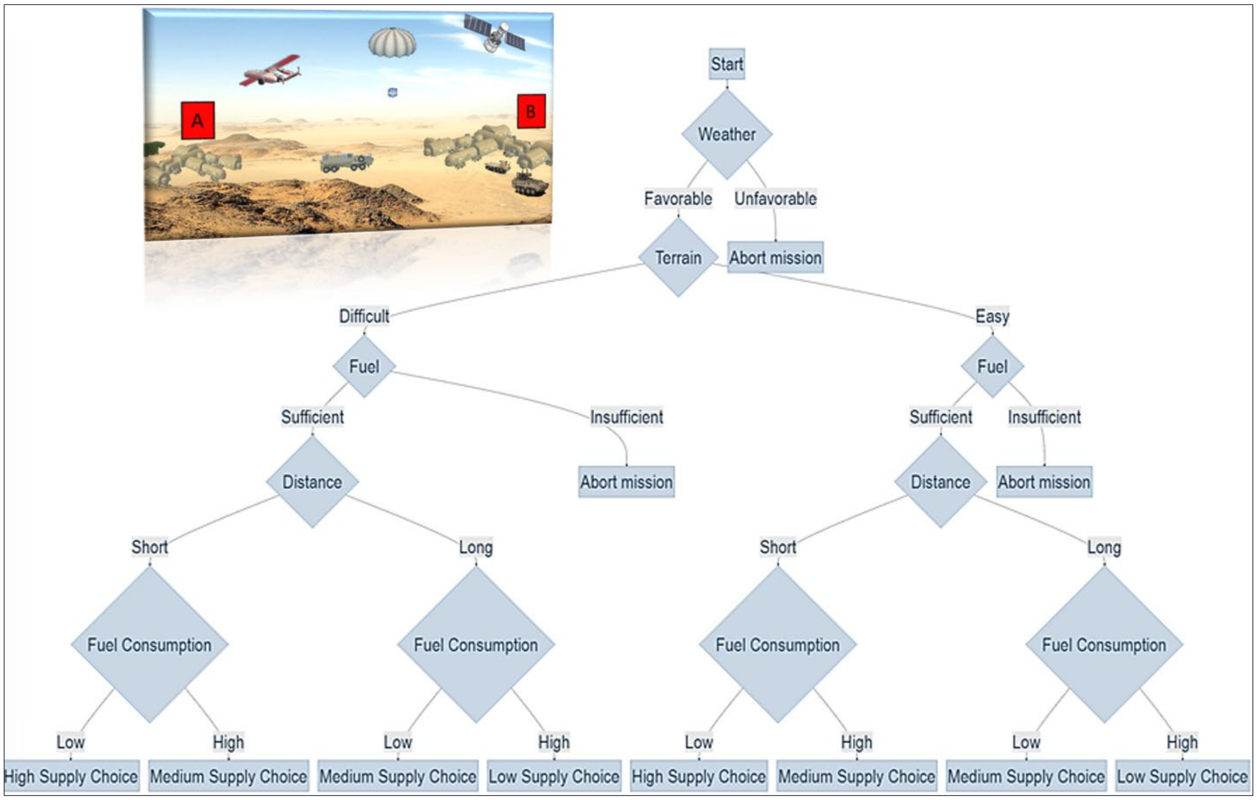


Figure 3. Example Decision Support Based on AI Models of Logistics Challenges Moving Supplies from Base A To B Under Weather, Terrain, And Fuel Constraints. LLMs Provide the Starting Basis for The Extracted Parameters and Choices to Guide the Knowledge-Based Predictions. In A Full Knowledge Representation System Like Cyc, You Would Also Include Relationships Between Entities, Constraints, Rules, And Other Types of Knowledge.

makers with crucial foresight by leveraging various analytic techniques, including statistical models, machine learning algorithms, and AI. This foresight can inform proactive strategies and interventions, provide a competitive advantage, and support risk management. In essence, anticipatory analytics tries to answer the question, "What will happen?" based on the patterns seen in data. However, it's important to note that the predictions are probabilistic and do not guarantee future events or outcomes. They represent a likelihood based on the available data and the sophistication of the predictive models.

METHODS AND RESULTS

Several use cases feature different LLM elements. The decision support example uses a typical Cyc case, where a commander relies on a prediction model for an estimated probability of success. To encode all human knowledge in formats that support reasoning engines, we introduce the Python Experta module, which draws inspiration from the NASA CLIPS engine to support AI reasoning. A military commander might ask the question: Given these variables for weather, terrain, and vehicle availability, can my mission succeed in transporting supplies from Base A to B?

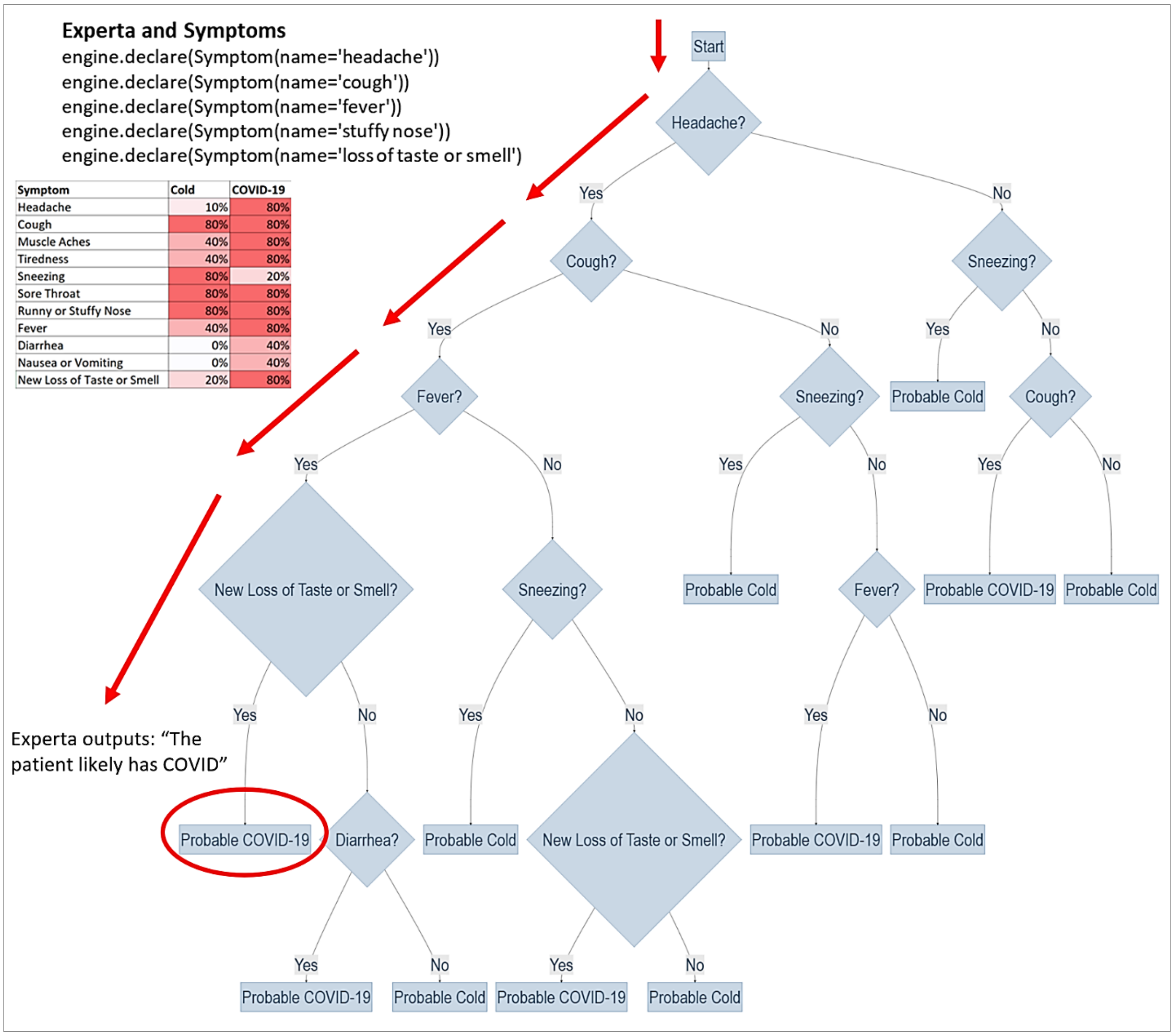
The second medical domain example demonstrates a definitive diagnostic decision based on patient symptoms. To make the example concrete in Figure 4, we use LLMs to generate a differential diagnosis (common cold vs. COVID-19) based on reported patient symptoms. The LLM role centers on extracting facts from patient records or study materials, assigning a diagnostic criterion of choices or "if-then" conditional cases. For example, if the patient reports sudden loss of taste and smell, the LLM generates text decisions applying a higher probability of COVID-19 than the common cold.

Example 1. Decision Support for Logistics Problem

We model a command decision derived from facts about logistic feasibility. In Cyc, such a problem could be represented as a collection of facts and rules about the locations of bases, the types, and capacities of available transport vehicles, the types of terrain between sites, the current weather conditions, etc. To solve this classic military challenge, we employ LLMs to extract a viable Cyc class structure, formulate expert rules compatible with the Experta module (Perez, 2023) and make predictions with visualization to support the final decision (Figure 3).

Our model logistics challenge offers the prediction, "The operation is feasible." because the quantity of supplies to be moved (900.0) is less than the transport capacity (1000.0) given the current weather, terrain, transport choice for trucks, fuel consumption, and base locations (Figure 3). This information would be used to determine which transport mode(s) could accomplish the task and what the potential difficulties or delays might be. We are considering moving supplies (900 units) by truck with fuel consumption in the same units (10 units/hour) between Base A and Base B. The commander must weigh the feasibility of success based on the distance between the two locations (A and B), the weather along the route (represented as conditions at A and B only for simplicity), and the terrain type.

For context, one might compare this choice to several historical logistics failures that changed the course of human events, such as Napoleon's 1812 decision to march on Moscow in winter or the second Lebanon War (2006) when



Experta and Symptoms

```
engine.declare(Symptom(name='headache'))
engine.declare(Symptom(name='cough'))
engine.declare(Symptom(name='fever'))
engine.declare(Symptom(name='stuffy nose'))
engine.declare(Symptom(name='loss of taste or smell'))
```

Symptom	Cold	COVID-19
Headache	10%	80%
Cough	80%	80%
Muscle Aches	40%	80%
Tiredness	40%	80%
Sneezing	80%	20%
Sore Throat	80%	80%
Runny or Stuffy Nose	80%	80%
Fever	40%	80%
Diarrhea	0%	40%
Nausea or Vomiting	0%	40%
New Loss of Taste or Smell	20%	80%

Experta outputs: "The patient likely has COVID"

Figure 4 Knowledge-Based AI Expert System Automated from Medical Literature Using LLM Encodings And Prompts

Israeli Defense Forces failed to plan for rapid combat advances that overran logistics. The latter challenge led to misdirected Israeli airdrops of food and water that Hezbollah diverted for their troops.

Example 2. Differential Medical Diagnosis

We model a physician's or field medic's decision as a choice between diagnosing a common cold or COVID-19 based on a patient's symptoms. We use LLMs to extract the symptom description ("loss of taste") from the medical literature (Mayo Clinic, 2023), assign reasonable probabilities for either disease to show that symptom, then generate a visualization of the medical choices.

Figure 4 shows an abbreviated knowledge graph and the input prompt to the LLM. For instance, a medical article describes the input to the LLM for extraction of key concepts in narrative format (Mayo Clinic, 2023):

> *1) Loss of taste or smell is a distinct symptom commonly reported in COVID-19 cases but not usually in cases of a common cold; 2) Symptoms such as fever and dry cough are also more frequently associated with COVID-19. On the other hand, symptoms like sneezing or a runny or stuffy nose are often associated with a common cold but can also occur in cases of COVID-19.*

Once the LLM provides a diagnosis, those expert rules translate into a predictive infection assignment using the probabilities of each symptom's association with the underlying causes. The resulting Experta module with encoded rules yields a supportable decision on a most likely diagnosis and its final report.

A fever and new loss of taste or smell are particularly distinctive of COVID-19, as these are symptoms less commonly seen in a cold. As shown in Figure 4, strong indicators lean towards a diagnosis of COVID-19 over a common cold. The remaining symptoms (headache, cough, and runny or stuffy nose) are common to both illnesses but occur in COVID-19 with a higher likelihood.

Given these inputs, the expert system determines that these symptoms are more indicative of COVID-19 due to the combination and severity of the symptoms. This underlines our approach's capability to diagnose illnesses based on input symptoms differentially, providing valuable assistance in clinical diagnosis and decision-making processes.

Example 3. Cybersecurity Strategies for Countering Advanced Persistent Threats

The cybersecurity investigation centers on the APT30 knowledge graph for a military Chinese cyber unit (Council on Foreign Relations, 2023) known as Naikon, PLA Unit 78020, and Lotus Panda (Table 3). This threat actor uses email spear-spear fishing to target (mainly) Southeast Asian competitors. The group is known to wedge its trading relationships with Nepal as part of a global influence campaign to weaken India as a long-term strategic rival (FireEye, 2015). Table 3 shows a prompt asked of ChatGPT and complex entity extraction from a press release (Nepal Government, 2023) that might inform the knowledge graph as a typical artifact.

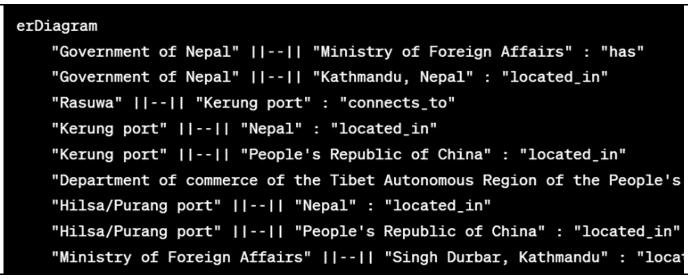
Table 3. LLM and Knowledge Graph Analysis for APT 30

Prompt to LLM (ChatGPT MAY 12, 2023 Version)	Response in Markdown Text Formats for Entity Extraction
PROMPT: Use bold markdown to **highlight the entities** for persons, organizations, places, and dates for the following press release with entity types—Mark after each entity (PERSON) for a person, (ORG) for an organization, etc. PERSON - People, including fictional ones. NORP - Nationalities or religious or political groups.	**Government of Nepal**(ORG) **Ministry of Foreign Affairs**(ORG) **Kathmandu, Nepal**(GPE) **Rasuwa**(GPE) and **Kerung Port** (FAC) between **Nepal**(GPE) and the **People's Republic of China**(GPE) have officially resumed its operation for two-way trade from today. The **Department of Commerce of the Tibet Autonomous Region of the People's Republic of China**(ORG) organized an official ceremony in **Kerung**(GPE) today to observe the opening of the port. Similarly, **Hilsa/Purang port**(FAC) has also been opened for one-way

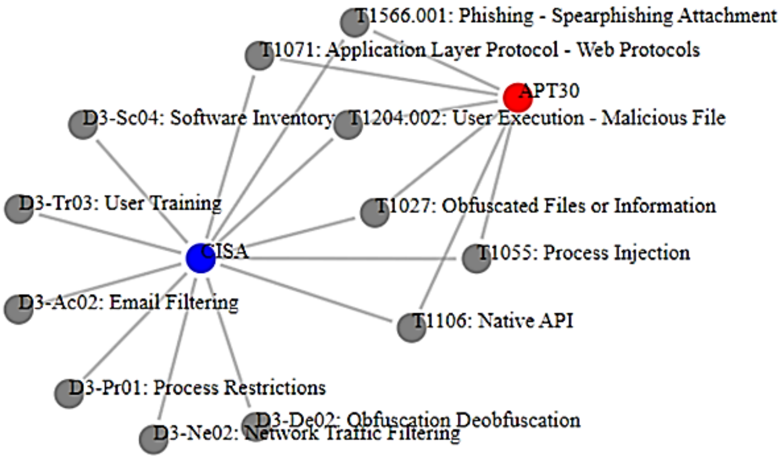
Table 3. LLM and Knowledge Graph Analysis for APT 30

Prompt to LLM (ChatGPT MAY 12, 2023 Version)	Response in Markdown Text Formats for Entity Extraction
FAC - Buildings, airports, highways, bridges, etc. ORG - Companies, agencies, institutions, etc. GPE - Countries, cities, states. LOC - Non-GPE locations, mountain ranges, bodies of water. PRODUCT - Objects, vehicles, foods, etc. (Not services.) EVENT - Named hurricanes, battles, wars, sports events, etc. (etc.)	trade from **December 26, 2022**(DATE). The resumption of the ports is expected to augment bilateral trade between **Nepal**(GPE) and **China**(GPE). **Ministry of Foreign Affairs**(ORG) **Singh Durbar, Kathmandu**(FAC) **December 28,2022**(DATE)

PROMPT: Convert the marked entities into an **entity relation diagram** using Mermaid.js format that can be rendered into a knowledge graph (https://mermaid.live/)



```
erDiagram
    "Government of Nepal" ||--|| "Ministry of Foreign Affairs" : "has"
    "Government of Nepal" ||--|| "Kathmandu, Nepal" : "located_in"
    "Rasuwa" ||--|| "Kerung port" : "connects_to"
    "Kerung port" ||--|| "Nepal" : "located_in"
    "Kerung port" ||--|| "People's Republic of China" : "located_in"
    "Department of commerce of the Tibet Autonomous Region of the People's
    "Hilsa/Purang port" ||--|| "Nepal" : "located_in"
    "Hilsa/Purang port" ||--|| "People's Republic of China" : "located_in"
    "Ministry of Foreign Affairs" ||--|| "Singh Durbar, Kathmandu" : "loca
```

PROMPT: How can the MITRE ATT&CK and D3FEND pairs for APT30 which uses the spear-phishing technique to download malware, and how to defend against it step-by-step about the framework's tactic and technique nomenclature? The knowledge graph shows the cyber defenders (CISA) and the attackers (APT30, attributed to China). Attached to the nodes are techniques defenders and attackers use to reach their goals and are labeled by the MITRE framework.	 T1566.001: Phishing - Spearphishing Attachment T1071: Application Layer Protocol - Web Protocols APT30 D3-Sc04: Software Inventory T1204.002: User Execution - Malicious File D3-Tr03: User Training T1027: Obfuscated Files or Information CISA T1055: Process Injection D3-Ac02: Email Filtering T1106: Native API D3-Pr01: Process Restrictions D3-De02: Obfuscation Deobfuscation D3-Ne02: Network Traffic Filtering

The system's reasoning capabilities were demonstrated as it identified the threat actor's tactics, techniques, and procedures (TTPs), which included spear-phishing campaigns and malware deployment. Recognizing these techniques, the expert system suggested appropriate countermeasures such as enhancing phishing awareness among staff, implementing robust email security filters, and ensuring up-to-date antivirus software to detect and neutralize malware. Furthermore, by analyzing the attack tree, the expert system could forecast future actions by APT30, enabling proactive measures. Attributing the attack to a specific threat group and predicting their next move is a critical advantage in cybersecurity defense, providing valuable lead time to harden defenses and protect valuable assets. The primary result for this use case proved to be adaptable entity extractions that spawned plausible attack trees that could match an existing ontology or categorization framework (MITRE ATT&CK, Strom et al., 2018).

DISCUSSION

The ability to anticipate future scenarios or outcomes, known as anticipatory analytics, is paramount in expert systems and LLMs. The capacity to predict future actions or conditions allows these systems to make proactive recommendations or decisions, a highly sought-after feature in many applications ranging from medical diagnosis to threat intelligence and decision support systems.

LLMs can be effectively integrated into rule-based expert systems to improve their explainability. The inherent ability of LLMs to comprehend and generate natural language makes them ideal for articulating the reasoning process of expert systems, improving their transparency. The LLM's linguistic capabilities can be harnessed to map complex rule chains into coherent narratives, elucidating the system's logic to human users. LLMs can parse vast information, identify patterns and correlations, and predict likely outcomes or future events. For instance, in the presented examples, LLMs identified symptoms of diseases and potential future actions of a cyber threat group. The effectiveness of LLMs in enhancing rule-based expert systems is shown to vary across different use cases. In the medical diagnosis example, LLMs enabled the system to interpret symptoms and make accurate diagnoses based on symptom probabilities. In the cyber threat attribution example, LLMs helped attribute attacks to specific threat groups and suggested suitable countermeasures. In the decision support system, the LLM was beneficial in interpreting complex logistics parameters and providing optimal supply routes. Their ability to articulate complex decision-making processes in understandable terms improves transparency, increasing user trust and facilitating adopting of AI solutions in various sectors.

While there have been significant strides in expert systems and LLMs, a substantial gap exists in integrating these two areas effectively. Expert systems have been valuable in many domains due to their robust rule-based reasoning capabilities, encapsulating the tacit knowledge of human experts. On the other hand, LLMs have demonstrated impressive abilities in understanding and generating human-like text based on vast quantities of data. However, existing research has often treated these two as distinct paths to AI development. There is a dearth of studies that thoroughly investigate, implement and evaluate the blending of these systems' unique strengths.

The second gap lies in the inability of the current methodologies to fully leverage the anticipatory and predictive capabilities of LLMs within expert systems. While there are promising attempts, as evidenced by some early-stage research on anticipatory analytics, there is still much room for exploration. Existing LLM applications largely remain descriptive or reactive rather than anticipatory. They excel at providing answers based on pre-existing knowledge but fall short in making robust, reliable predictions about future events or trends. Similarly, traditional expert systems tend to operate within their predefined rule sets and cannot extrapolate or adapt beyond their encoded knowledge. Therefore, developing frameworks that merge the adaptive learning of LLMs with the rule-based reasoning of expert systems while enhancing their predictive power is a largely uncharted territory that needs further exploration.

CONCLUSIONS

The advances in AI and NLP have led to LLMs that comprehend, generate, and engage with human language. Underpinned by transformer-based machine learning architectures, LLMs are trained on extensive text data, enabling them to understand context, generate effective responses, and display a form of commonsense reasoning. When integrated with expert systems for emulating human expert decision-making, these LLMs can synergistically enhance both domains. In these blended systems, LLMs facilitate the formation of human-readable knowledge and reasoning, offering an element of explainability often absent in opaque AI models.

This study contributes to AI and rule-based expert systems in several ways. Firstly, it illustrates the application of LLMs in expert systems across multiple domains, emphasizing their versatility. Secondly, it proposes a methodology to generate and validate knowledge graphs from LLM, demonstrating their efficacy in improving decision-making processes. Thirdly, through the diverse use cases examined - medical diagnosis with COVID symptoms, cyber threat attribution for APT30, and a decision support system - the study provides a practical demonstration of the applicability of LLM and the proposed methodology. This breadth of application is a crucial contribution, suggesting potential universal applicability. Finally, the paper provides a comprehensive comparative analysis across these various use cases, which underpins the robustness and scalability of the approach and presents a roadmap for future research in the area.

ACKNOWLEDGEMENTS

The authors thank the PeopleTec Technical Fellows program for encouragement and project assistance. This research benefited from encouragement from US Army Space and Missile Defense Command. The author thanks the researchers at Open AI for developing LLMs and allowing public access to ChatGPT.

REFERENCES

Alexopoulos, C., Saxena, S., Loukis, E. N., Rizun, N., Ioanna Maratsi, M., & Ali, M. (2023). ChatGPT Application Vis-A-Vis Open Government Data (OGD): a Research Agenda. *Available at SSRN 4378771*.

Au Yeung, J., Kraljevic, Z., Luintel, A., Balston, A., Idowu, E., Dobson, R. J., & Teo, J. T. (2023). AI chatbots not yet ready for clinical use. *Frontiers in Digital Health*, *5*, 60.

Auer, S., Barone, D. A., Bartz, C., Cortes, E. G., Jaradeh, M. Y., Karras, O., ... & Tsalapati, E. (2023). The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge. *Scientific Reports*, *13*(1), 7240.

Boury-Brisset, A. C. (2004). *Ontological approach to military knowledge modeling and management*. DEFENCE RESEARCH AND DEVELOPMENT CANADAVALCARTIER (QUEBEC).

Bowman, M., Lopez, A., & Tecuci, G. (2001, March). Ontology development for military applications. In *Proceedings of the Thirty-ninth Annual ACM Southeast Conference, Athens, GA, ACM Press: New York, NY*.

Brynjolfsson, E., & McAfee, A. (2017). What's driving the machine learning explosion. *Harvard Business Review*, *18*(3), 118.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Chang, E. Y. (2023, March). Prompting large language models with the socratic method. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0351-0360). IEEE.

Council on Foreign Relations, (2023) APT 30, Cyber Operations, https://www.cfr.org/cyber-operations/apt-30

Deng, X. (2023, April). A More Accessible Web with Natural Language Interface. In 20th International Web for All Conference (pp. 153-155).

Fensel, D., & Fensel, D. (2001). *Ontologies* (pp. 11-18). Springer Berlin Heidelberg.

Fernández-López, M., & Gómez-Pérez, A. (2002). Overview and analysis of methodologies for building ontologies. *The knowledge engineering review*, *17*(2), 129-156.

FireEye, (2015), Special Report: APT30 And The Mechanics Of A Long-Running Cyber Espionage Operation, https://www2.fireeye.com/rs/fireye/images/rpt-apt30.pdf

Getty Center, (2021), S. Brown, Where Did Writing Come From? https://www.getty.edu/news/where-did-writing-come-from/

Hu, M., Pan, S., Li, Y., & Yang, X. (2023). Advancing Medical Imaging with Language Models: A Journey from N-grams to ChatGPT. *arXiv preprint arXiv:2304.04920*.

Kollapally, N. M., Keloth, V. K., Xu, J., & Geller, J. (2023). Integrating Commercial and Social Determinants of Health: A Unified Ontology for Non-Clinical Determinants of Health. *arXiv preprint arXiv:2304.01446*.

Lande, D., & Strashnoy, L. (2023). Concept Networking Methods Based on ChatGPT & Gephi. *Available at SSRN 4420452*.

Lenat, D. B., Prakash, M., & Shepherd, M. (1985). CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI magazine*, *6*(4), 65-65.

Lin, W., Babyn, P., & Zhang, W. (2023). Context-based Ontology Modelling for Database: Enabling ChatGPT for Semantic Database Management. arXiv preprint arXiv:2303.07351.

Matuszek, C., Witbrock, M., Cabral, J., & DeOliveira, J. (2006). An introduction to the syntax and content of Cyc. *UMBC Computer Science and Electrical Engineering Department Collection*.

Mayo Clinic, (2023), COVID-19, cold, allergies and the flu: What are the differences?, https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/covid-19-cold-flu-and-allergies-differences/art-20503981

Nepal Government (2023), Press Release: Opening of Two-Way Trade, https://cn.nepalembassy.gov.np/wp-content/uploads/2023/01/Press-Release-Opening-of-Two-way-trade.pdf

Nirenburg, S., Krishnaswamy, N., & McShane, M. (2023). Hybrid Machine Learning/Knowledge Base Systems Learning through Natural Language Dialogue with Deep Learning Models.

Perez, R.A.M, (2023), Experta: Python Library for Building Expert Systems Inspired by CLIPS, https://github.com/nilp0inter/experta

Polymer, (2023), 40+ Important ChatGPT Statistics to Know in April 2023, https://www.polymersearch.com/blog/chatgpt-statistics

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

Reed, S. L., & Lenat, D. B. (2002, July). Mapping ontologies into Cyc. In *AAAI 2002 Conference workshop on ontologies for the semantic Web* (pp. 1-6).

Sarjant, S., Legg, C., Robinson, M., & Medelyan, O. (2009, September). " All You Can Eat" Ontology-Building: Feeding Wikipedia to Cyc. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 341-348). IEEE.

Semy, S. K., Pulvermacher, M. K., & Obrst, L. J. (2004). *Toward the use of an upper ontology for US government and US military domains: An evaluation*. MITRE CORP BEDFORD MA.

Sima, A. C., & de Farias, T. M. (2023). On the Potential of Artificial Intelligence Chatbots for Data Exploration of Federated Bioinformatics Knowledge Graphs. *arXiv preprint arXiv:2304.10427*.

Song, S., Ryu, K., & Kim, M. (2010, May). Ontology-based decision support for military information systems. In *2010 IEEE Long Island Systems, Applications and Technology Conference* (pp. 1-5). IEEE.

Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., & Thomas, C. B. (2018). Mitre att&ck: Design and philosophy. In *Technical report*. The MITRE Corporation.

Turing, A. M. (1956). Can a machine think. The world of mathematics, 4, 2099-2123.

Trajanoska, M., Stojanov, R., & Trajanov, D. (2023). Enhancing Knowledge Graph Construction Using Large Language Models. *arXiv preprint arXiv:2305.04676*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wygant, R. M. (1989). CLIPS—A powerful development and delivery expert system tool. *Computers & industrial engineering*, *17*(1-4), 546-549.