

The Potential Application of Large Language Models in Pharmaceutical Supply Chain Management

David Aguero, PharmD and Scott D. Nelson, PharmD

KEYWORDS Large Language Models; Artificial Intelligence; Informatics; Supply Chain; Pharmacy

J Pediatr Pharmacol Ther 2024;29(2):200–205

DOI: 10.5863/1551-6776-29.2.200

The pharmaceutical supply chain is growing increasingly intricate, with various challenges arising in drug inventory management, procurement, distribution, and dispensing processes. Issues such as shortage mitigation, inconsistency in regulatory compliance, cost control, cold-chain storage, adaptation to technologic advances, and secure information sharing all pose significant difficulties to pharmaceutical supply chain management.¹ Consequently, this adds immense pressure on health system pharmacies to enhance their arsenal of available tools.²

Pediatric hospitals in particular encounter unique challenges relating to pharmaceutical and supply shortages, which can cause delays in vital patient procedures, alterations in care protocols, and even unexpected changes in care locations.

Historically, supply chain operations have been improved by using physical automation technologies and methodologies such as robotic process automation. Efficient operations use business intelligence to drive informed decisions regarding sourcing, pricing, and patient care, which can be facilitated either by internal teams or through vendor-provided solutions.³ While this approach does mitigate several problems, it rarely results in holistic problem resolution—like a chronic illness, these issues wax and wane.

Addressing these supply chain challenges is crucial to ensure that patients receive the most optimal care available.³ Emerging technologies like large language models (LLMs) and generative artificial intelligence (AI) present new opportunities to address these supply chain challenges. LLMs can be used for knowledge management, data analysis, and process automation, contributing to the ongoing digitalization of supply chains, in ways that are difficult to predict, but worthwhile considering.⁴

Generative AI and Large Language Models

LLMs are a type of deep-learning artificial neural network that can perform various natural language processing (NLP) tasks, such as recognizing, translating, predicting, and generating text. However, unlike traditional NLP models, LLMs can capture long-range

dependencies and contextual information within the entire prompt by using what are called self-attention mechanisms, instead of word by word. Self-attention allows the model to focus on the most relevant parts of the input and output, and how they relate to each other.⁵ This self-attention is based on a mathematical model called a transformer, which uses many layers of neural networks to process the input and output text sequences. The transformer consists of an encoder that converts the input text into numerical representations called tokens, and a decoder to convert the results back into text. These LLM transformers are trained to predict the next best word, or missing words, based on many examples of natural language from massive amounts of data, often from diverse sources and domains, to learn the patterns and structures of natural language. LLM transformers have billions of parameters, which are the mathematical weights the model learns during training and uses to make predictions during inference. Hence, an already trained transformer model that is used to generate text is called a generative pretrained transformer. LLMs have achieved remarkable results in various NLP benchmarks and applications, such as question answering, text summarization, text generation, and sentiment analysis. Training a transformer requires massive amounts of compute power,⁶ so these pretrained transformers can also be adapted to specific tasks or domains by fine-tuning, which is the process of updating the parameters of a pretrained model by using a smaller and more relevant dataset or even using reinforcement learning from human feedback of results.⁷

LLMs have proven to be a disruptive technology, and there is a lot of interest in leveraging LLMs across many domains to perform different tasks, and as such, they have the potential to affect health care in many ways. For example, LLMs can help pharmacists and clinicians with drug information retrieval, drug interaction checking, adverse drug event detection, medication adherence monitoring, patient education, and clinical decision support.⁸ However, LLMs also pose some challenges and limitations, such as data quality, data privacy, model reliability, AI hallucinations, model explainability, and ethical implications. Therefore, it is

important to evaluate the performance and validity of LLMs in health care settings and to ensure their safe and responsible use.

In this commentary, we will explore several ways LLMs could be applied to pharmaceutical supply chain management, and some areas where there are potentially major limitations to their use.

The Authors Believe LLMs Are Likely Useful in the Following Areas of the Pharmaceutical Supply Chain

Systems Integration. LLMs can likely help integrate various inventory optimization tools and disparate databases by using natural language to generate queries and responses, which could reduce the need for manual data entry and extraction across different systems. For example, an LLM could help extract relevant information from unstructured or semi-structured data sources, such as free-text or voice recordings, and then structure the relevant information in a way that can be used by other systems (e.g., inventory management systems based on upstream prescriber orders). Additionally, LLMs perform well at code generation, which could reduce the effort needed to convert data from multiple systems into a standard data model, or help with translating the data from one system to another so it can then be used in the other system, thereby providing an end-to-end integration and data flow by translating between systems. The limitation here is that the model will not be able to extract data that do not exist or are not available to the LLM from the external systems.

Human-Data Interface and Data Analysis. Similar to the systems integration, LLMs could help with the human-computer interface and facilitate data analysis.⁹ For example, using their ability to translate natural language to code could be useful for extracting and analyzing data from a database. A person could ask a question of the data, such as “Which medications do we use the most each week, and what is their total cost?” The LLM would then convert the question into a structured query, extract the data, then write python code to analyze those data to determine which medications are used the most and how much they have cost. Given the chatbot nature of the LLMs, the LLM could even ask clarifying questions to better formulate the question, such as asking how far back the user would like to look for summing the total cost of each medication. This natural conversational format could facilitate easier use of the computer systems. However, anyone who has worked in reporting know the challenges to shaping questions and understanding the true intent behind a question to provide valuable information. The LLMs could provide the user with the information based on what they asked, but that may not be what they really wanted, or could even be wrong. It is important to test and verify results to make

sure the statistical analysis chosen is the appropriate approach, and that the databases used have reliable data to answer the questions.

Process Automation. LLMs could help facilitate the automation of complex hierarchical and planning tasks, such as automatically performing a task when certain conditions are met or helping to triage messages to the correct recipients. LLMs have the capability to report even when there is missing or incoherent data present, they are less affected by spelling mistakes or grammatical errors, and could even make mass changes to data. However, having LLMs make mass changes to data to automatically carry out tasks could cause problems, and there should always be a human in the loop to review and approve the action to be taken.

Simulation and Training. LLMs can theoretically be used to help improve supply chain resilience by simulating disruption scenarios to help generate and improve risk-management plans. This is achieved through the analysis of archival data and external influencers to highlight weak points. The ability of these AI models to digest vast amounts of data makes it possible to detect risk-associated patterns and forecast possible disturbances.¹⁰ A trained LLM can evaluate factors such as historical supplier disruptions, meteorologic patterns, and market seasonality, and early results suggest this has been successful in other industries.¹¹ Upon identifying an elevated risk by the generative AI model involved in the supply chain planning, it could simulate the effect on supply routes. This prognostic capacity could enable pharmacy operations to establish stronger emergency plans. Additionally, LLMs can quickly generate synthetic datasets for training and onboarding new staff by walking through various scenarios, asking questions, and responding to questions.⁹ However, LLMs are not able to predict unusual or rare events, such as war, natural disasters, cyberattacks, or pandemics (i.e., black swan events) that could very well affect the supply chain.

Knowledge Management. Knowledge management is a challenge in any industry or company. LLMs could be trained on internal, unstructured data, such as emails, meeting minutes, training materials, transaction histories, notes, best-practice documents, policy and standard operating procedure documents, and notifications received from outside systems. This could help facilitate search of this valuable source of institution-specific recommendations, answer questions, and identify how similar issues were handled in the past.⁹ However, policies and practices tend to change over time, and the LLM could return outdated materials, or require retraining. The performance of the model is only going to be as good as the information it trains on.

Text Summarization. One key feature of LLMs is their ability to summarize large volumes of text, which

could be used to help chronic or document decisions made as well as the rationale. LLMs could be used to summarize notifications of potential shortages, then combined with internal data, put those notifications in institution-specific context for human review. While text summarization is helpful, there is always the risk of missing or leaving out key information.

Optimizing Travel Routes and Recommending Methods to Reduce Environmental Impact. One potential area could be to use LLMs to optimize travel or delivery routes. One could imagine putting the delivery data and concepts into textual sentences and then do analysis to optimize the routes. However, there are already sophisticated algorithms and software used to optimize travel routes.

Where LLMs Would Struggle to Produce Value

While LLMs hold potential in many prediction-based applications, they also face significant hurdles in other aspects of pharmaceutical supply chain. The authors believe LLMs would struggle to produce value in the following scenarios.

Predicting Medication Shortages. Medication shortages are generally triggered by acute events such as increases in demand, raw and/or active product ingredient shortages, or complications in manufacturing processes. Because LLMs are trained on historical data and may not have real-time access to such rapidly changing information, they are challenged to predict theoretical shortages. In fact, relying on these models to predict a shortage could cause a triggering event that leads to a medication shortage. Further, much of the data needed to predict medication shortages is private, meaning only specific entities (Group Purchasing Organization, wholesalers, or specific organizations like the United States Pharmacopeia) have access to the right set of information to predict a theoretical shortage with any accuracy. Black swan events, such as the COVID-19 pandemic, cause additional complications for LLMs, which may lack a full contextual understanding necessary to interpret the nuanced effects of geopolitical events, economic trends, natural disasters, and other factors affecting supply chain operations.

Enhanced Data Quality. The importance of standardization and structured syntax and semantics in enhancing supply chain operations cannot be overstated. Despite the existence of common data standards such as the Global Trade Item Number and the Global Location Number from GS1, their adoption is not universal. It is possible for 2 facilities within the same health system to use different systems, or different versions of the same system. Furthermore, different major wholesalers provide varying information for identical products. This means Wholesaler A and Wholesaler B may categorize the therapeutic class of

a new drug differently, leading to inconsistencies and posing substantial challenges for LLM.

Even as GS1 US and other organizations promote data standardization, limited standardization persists in the development, configuration, and data storage methods of supply chain systems. This can impede LLMs' ability to derive meaningful insights from unstructured data, exemplifying the "Garbage In, Garbage Out" concept. Particularly for pediatric hospitals, the lack of standardization may lead to medication administration errors and other crucial downstream processes. For instance, if an LLM approves distribution of a product containing preservatives, it may negatively affect a sensitive pediatric patient population. Furthermore, the lack of comprehensive integration and interoperability among various health systems today can hinder data and insights sharing across different sections of the supply chain.

Nonetheless, LLMs can assist in detecting anomalies and effecting repairs in mismatched databases. Clinical content vendors can use trained models to scrutinize the content (typically sent monthly as part of a subscription) for missing or miscoded fields, based on preexisting data. Such data eventually finds its way into both electronic health record and third-party-supported inventory systems that affect patient care. Leveraging historical data, LLMs could also assess trends and detect anomalies within the system, such as miscoded revenue fields or quantity conversion factors between units of issue (like vials) and units of sale (like cartons containing 5 sleeves of 6 vials).

Inventory Management. While powerful in making real-time decisions based on historical information, placing high-acuity orders based on present day workflows (consignment, 340B accumulation) suggest LLMs are a stronger copilot than a replacement for humans in current state. These tasks require a human touch, or a "thoughtful, warm body," to make decisions based on nuanced factors that a language model may not fully comprehend (such as substitution of a compounded preparation in place of a commercially used preparation, or considering loan/borrow activity for patient-specific doses within a geographic area, and factoring these scenarios in appropriately for inventory use and forecasting).

While LLMs can augment intelligence with predictive analytics and give personalized product alternative recommendations, they may not always provide accurate local forecasting based on historical usage. This could lead to incorrect predictions and potential inventory imbalances. Furthermore, while they can automate tasks like document processing through intelligent automation or digital workers that combine conversational AI with robotic process automation, there are already more specialized tools available for such tasks. These tools are designed specifically for inventory management and can provide more accurate

and efficient results, making them a better choice for managing multiple ordering systems and adjusting inventory levels regularly.

Discussion and Limitations on the Application of LLMs

Keeping LLMs Up-to-Date Can Be Difficult. Training LLMs requires massive amounts of compute power and data to train on, which could pose a significant limitation to their application in supply chain. Pharmaceutical supply chain management is a very dynamic and complex domain that involves multiple stakeholders, regulations, and uncertainties. The LLMs could be making recommendations based on historical or out-of-date information. Keeping LLMs up-to-date with the latest information and trends could prove to be very difficult and costly. Moreover, retraining LLMs can break the reinforcement learning from human feedback process,¹² which is a technique that uses the feedback from human users and their preferences to optimize the LLM's output performance. Therefore, retraining the model can degrade its performance and alignment with human values and preferences. One possible way to overcome these challenges is by using an application programming interface (API) to improve the prompt. In this method, the LLM uses the prompt given by the user, then calls an API to pull the most up-to-date information, then adds that information back into the prompt to improve the results. An API could provide relevant and updated information from various sources, such as databases, web search results, or news articles, and incorporate that information into the new prompt fed back to the LLM. Using an API can also enable more fine-grained control over the LLM's behavior, such as specifying the tone, style, or length of the output. Given the LLM's ability to write code, it could even help write its own API interfaces.

It Is All About the Prompt. The usefulness of LLMs and their results depend largely on the quality and design of the prompts, which influence the LLM's behavior and output. However, there are some limitations to prompts, such as context and token limits. Context limits are when the prompts may not provide enough context or information for the model to generate relevant and accurate texts. For example, if the prompt were "Write a report on the inventory status of the pharmacy," the LLM may not know which pharmacy, what time period, or what level of detail is required. Part of this is also restricted by token limits because most LLMs have a maximum number of tokens (words or characters) that they can handle as input or output. For example, if the prompt is "Write a summary of the latest research on COVID-19 vaccines in one sentence," the input prompt would not be large enough for the LLM to process all the latest research, and the model may not be able to fit all the necessary information in such a short output.

Therefore, the importance of good prompts cannot be overstated. Good prompts are clear, specific, and concise. They provide enough context and information for the model to generate high-quality and relevant responses. They also use appropriate constraints and formats to guide the model's output. For example, a good prompt for an inventory report could be:

"Write a report on the inventory status of ABC Pharmacy for the month of October 2023. Include the following information:

- The total number of items in stock, broken down by category (e.g., prescription drugs, over-the-counter drugs, medical supplies).
- The average turnover rate and shelf life of each category.
- The current and projected demand for each category.
- The recommended actions to optimize the inventory management and reduce waste.
- Limit your report to 500 words and use bullet points and tables where possible."

To produce the most effective prompts, some organizations have even started hiring prompt engineers to get the most out of the LLMs. Prompt engineering is the process of creating effective prompts for LLMs. It is a skill that requires creativity, attention to detail, and experimentation. Prompt engineering is an emerging field, and by learning and applying prompt engineering techniques, users could leverage the power of LLMs to automate tasks and improve productivity.

Privacy and Security Concerns. As with any technology, LLMs also pose certain security risks for organizations. For example, LLMs may consume sensitive data such as patient records, prescriptions, inventory levels, or supplier contracts, and use that private data for further training or inference. This could lead to data leakage, unauthorized access, or privacy violations if the LLMs are not properly secured, audited, or regulated. Moreover, LLMs may generate inaccurate, harmful, or malicious content based on the patterns they learned during training, which could compromise the quality, safety, or integrity of the pharmaceutical supply chain.

Some of the common vulnerabilities that affect LLM applications are as follows¹³:

- **Prompt injections:** When someone bypasses filters or manipulates the LLM by using crafted prompts that make the model ignore previous instructions or perform unintended actions.
- **Data leakage:** When the LLM reveals sensitive or private information through the LLM's responses, either intentionally or unintentionally.
- **Inadequate sandboxing:** This results from failing to isolate the LLM from the underlying system or network, thereby allowing the LLM to access or modify resources that it should not.

- **Unauthorized code execution:** When someone exploits the LLM prompt to have it execute arbitrary code or commands on the system or network, potentially compromising its security or functionality.
- **Overreliance on LLM-generated content:** When someone trusts the LLM's output without verification, validation, or moderation, which could result in errors, biases, or misinformation. This is also referred to as automation bias.¹⁴
- **Inadequate AI alignment:** When the organization fails to align the LLM's objectives, values, or incentives with the developer's intent or legal requirements, which could lead to undesired or harmful outcomes.
- **Training data poisoning:** When someone compromises the integrity of the training data, either by injecting malicious data or by modifying existing data, which could affect the LLM's performance or behavior.
- **Adversarial attacks:** When someone crafts inputs that are designed to fool or mislead the LLM, either by causing it to produce incorrect or misleading outputs, or by reducing its confidence or accuracy.
- **Model stealing:** When someone extracts or copies the LLM's parameters, architecture, or functionality, either by querying the LLM repeatedly or by analyzing its responses, which could violate intellectual property rights, enable unauthorized use, or expose sensitive information.
- **Model inversion:** This happens when someone infers or reconstructs the training data or the LLM's internal state, either by observing the LLM's outputs or by exploiting its vulnerabilities, which could breach data privacy or confidentiality.

AI Hallucinations. LLMs are trained to predict the next best word, based on statistical patterns from large amounts of data, and as such, this can sometimes result in errors or misinformation. An "AI hallucination" occurs when LLMs generate output that is not logically consistent with the training data, but this is done in such a way as to convey confidence in the incorrect answer, which may still sound plausible to the end user. Hallucination can cause confusion, misunderstanding, or false beliefs if the user is not aware of the limitations and uncertainties of LLMs. Therefore, it is important to view LLMs as *augmented intelligence*, rather than artificially replacing human intelligence. *Augmented intelligence* means that LLMs are designed to assist and complement human intelligence, not to replace or surpass it. Additionally, even though LLMs can answer questions, they are still only statistical models and not capable of truly *understanding* the meaning, context, or domain of the training data or prompt.

In summary, LLMs represent a promising technology that could greatly enhance some processes within

pharmaceutical supply chain management. Nevertheless, their application comes with certain constraints and risks: Clinicians and leaders will likely encounter hurdles when trying to leverage LLMs to boost supply chain efficiency. The authors recommend using LLMs for clerical tasks that depend on historical and readily accessible data, and it is crucial to verify the output before allowing LLMs to operate autonomously. The use of LLMs should be approached with prudence, transparency, and responsibility, and should always be complemented by human expertise and discernment.

Article Information

Affiliations. Department of Biomedical Informatics (SDN), Vanderbilt University Medical Center, Nashville, TN; Department of Pharmacy and Pharmaceutical Sciences (DA), St. Jude Children's Research Hospital, TN.

Correspondence. David Aguero, PharmD; david.aguero@stjude.org

Disclosures. The authors declare no conflicts or financial interest in any product or service mentioned in the manuscript, including grants, equipment, medication, employment, gifts, and honoraria. Scott Nelson is on the advisory board for Merative Micromedex and Baxter Health and David Aguero is an elected member of GS1 US Healthcare Executive Committee

Ethical Approval and Informed Consent. Not applicable.

Submitted. December 27, 2023

Accepted. January 4, 2024

Copyright. Pediatric Pharmacy Association. All rights reserved. For permissions, email: membership@pediatricpharmacy.org

References

1. Huss G, Barak S, Reali L, et al. Drug shortages in pediatrics in Europe: the position of the European Pediatric Societies. *J Pediatr*. 2023;261:113472.
2. Wosińska ME, Mattingly II TJ, Conti RM. A framework for prioritizing pharmaceutical supply chain interventions. *Health Affairs Forefront*. 2023.
3. DiPiro JT, Nesbit TW, Reuland C, et al. ASHP Foundation Pharmacy Forecast 2023: Strategic Planning Guidance for Pharmacy Departments in Hospitals and Health Systems. *Am J Health Syst Pharm*. 2023;80(2):10–35.
4. Seifert, RW, Markoff, R. How will Large Language Models impact supply chains? IDM. 2023. <https://www.imd.org/ibyimd/innovation/large-language-model-impacts-on-supply-chain/> Accessed November 22, 2023.
5. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 2017;30.
6. Zhang M, Li JT. A commentary of GPT-3 in MIT Technology Review 2021. *Fundam Res*. 2021;1(6):831–833.
7. Liu GK-M. *Transforming Human Interactions With AI Via Reinforcement Learning With Human Feedback (RLHF)*. MIT Schwarzman College of Computing, Envisioning

- the Future of Computing Prize 2023; 2023. <https://computing.mit.edu/cross-cutting/social-and-ethical-responsibilities-of-computing/envisioning-the-future-of-computing-prize-2023/>, Accessed November 22, 2023.
8. Liu S, Wright AP, Patterson BL, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc.* 2023;30(7):1237–1245.
 9. Ralf Seifert RM. How will Large Language Models impact supply chains? 2023. Accessed November 22, 2023. <https://www.imd.org/ibyimd/supply-chain/large-language-model-impacts-on-supply-chain/>
 10. GS1. Generative AI in the supply chain. 2023. Accessed December 26, 2023. <https://www.gs1us.org/articles/generative-ai-in-the-supply-chain>
 11. Li B, Mellou K, Zhang B, et al. Large language models for supply chain optimization. Preprint. Posted online Month day, 2023. arXiv preprint arXiv:230703875.
 12. Casper S, Davies X, Shi C, et al. Open problems and fundamental limitations of reinforcement learning from human feedback. Preprint. Posted online July, 23, 2023. arXiv preprint arXiv:230715217.
 13. S M. 10 LLM vulnerabilities and how to establish LLM security. 2023. <https://www.hackerone.com/vulnerability-management/owasp-llm-vulnerabilities>, Accessed December 10, 2023.
 14. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc.* 2012;19(1):121–127.