Artificial Intelligence in the Pharmaceutical Industry: Understanding Transformers and Large Language Model Agents for Hong Kong Pharmacists

Lam, Nate
ElendiLabs Ltd, Unit 1209, 12/F, Winfield Commercial Building, 6-8A Prat Ave, Tsim Sha Tsui

INTRODUCTION

Artificial Intelligence (AI) has fundamentally transformed numerous industries, and the pharmaceutical sector is no exception. As Hong Kong pharmacists navigate an increasingly complex regulatory landscape, understanding AI technologies—particularly large language models (LLMs) and their underlying transformer architecture—becomes crucial for staying competitive and compliant in the pharmaceutical industry[1]. This article provides an accessible introduction to transformer mechanisms in LLMs and explores their revolutionary applications as autonomous agents in the pharmaceutical industry.

UNDERSTANDING LARGE LANGUAGE MODELS AND THE TRANSFORMER ARCHITECTURE

What Are Large Language Models?

LLMs are sophisticated AI systems designed to understand and generate human-like text by learning from vast amounts of textual data[2]. These models, such as GPT-4 and Claude, have demonstrated remarkable capabilities in processing complex language tasks, making them invaluable tools for pharmaceutical applications where precision in documentation and communication is paramount.

The Foundation: Transformer Neural Networks

At the heart of modern LLMs lies the transformer architecture, a revolutionary neural network design that has transformed how AI systems process information[3]. Unlike traditional neural networks that process information sequentially, transformers can analyze entire sequences simultaneously, making them exceptionally efficient for language-related tasks.

The Self-Attention Mechanism: The Core Innovation

The transformer's most significant innovation is the **self-attention mechanism**[3]. Think of attention as a spotlight that helps the model determine which parts of the input text are most relevant when processing a specific word or phrase. For example, when analyzing the sentence "The drug showed efficacy in patients with diabetes," the attention mechanism helps the model understand that "drug" and "efficacy" are closely related, even though they appear at different positions in the sentence[3]. The self-attention mechanism works through three key components:

1. **Query (Q)**: Represents what the model is currently focusing on
2. **Key (K)**: Represents all available information in the sequence
3. **Value (V)**: Contains the actual content to be processed

The model calculates attention scores by comparing queries with keys, determining how much attention to pay to each part of the input when processing specific elements[3].

Positional Encoding: Understanding Context

Since transformers process all input simultaneously rather than sequentially, they require a mechanism to understand word order and position. Positional encoding provides this capability by adding position-specific information to each word, ensuring the model understands that "The patient received the drug" conveys a different meaning than "The drug received the patient"[3].

HOW TRANSFORMERS PROCESS PHARMACEUTICAL DOCUMENTATION

When a transformer-based LLM processes a document, it follows these key steps:

1. **Tokenization**: Breaking down text into manageable units (words, subwords, or characters)
2. **Embedding**: Converting tokens into numerical representations
3. **Attention Computation**: Identifying relationships between different parts of the document
4. **Layer Processing**: Multiple transformer layers progressively refine understanding
5. **Output Generation**: Producing relevant responses or analyses

This architecture makes transformers particularly effective for pharmaceutical applications because they can:

- Simultaneously consider multiple sections of lengthy documents
- Identify complex relationships between safety data, efficacy results and requirements
- Maintain context across hundreds of pages of documentation

LLM AGENTS: THE NEXT EVOLUTION IN THE PHARMACEUTICAL INDUSTRY

From Passive Tools to Active Agents

While traditional AI systems require explicit instructions for each task, **LLM agents** represent a paradigm shift toward autonomous operation[4][5]. These agents can reason, plan, and execute complex workflows independently, making them ideal for the multifaceted challenges in the pharmaceutical industry.
An LLM agent differs from a standard LLM in several key ways:

- **Autonomy**: Agents can operate independently with minimal human intervention
- **Tool Integration**: They can access and utilize external databases, APIs, and software systems
- **Multi-step Reasoning**: Agents can break down complex tasks into manageable subtasks
- **Adaptive Learning**: They can adjust their approach based on feedback and changing conditions

APPLICATIONS IN THE PHARMACEUTICAL INDUSTRY

1. Regulatory Submission Preparation and Review

LLM agents can revolutionize the preparation of regulatory submissions by:

Cross-Reference Verification: These agents can automatically verify that information remains consistent across different sections of a submission, identifying discrepancies that might otherwise go unnoticed[6].

Regulatory Intelligence: Agents can continuously monitor regulatory guidance updates from agencies like the FDA, EMA, and Hong Kong's Department of Health, automatically flagging changes that might impact ongoing or planned submissions**[7]**.

2. Compliance Monitoring and Quality Assurance

Real-time Compliance Tracking: LLM agents can monitor pharmaceutical operations continuously, identifying potential compliance issues before they become violations[8]. They can track changes in manufacturing processes, supplier qualifications, and quality control procedures against regulatory requirements.

Automated Adverse Event Reporting: Agents can process safety data from multiple sources, automatically identifying reportable adverse events and generating preliminary safety reports in compliance with pharmacovigilance requirements[6].

3. Stakeholder Communication and Training

Automated Correspondence: LLM agents can draft responses to regulatory queries, prepare meeting minutes, and generate status reports for internal and external stakeholders.
Training Material Development: Agents can create customized training materials for different teams, ensuring staff remain current with evolving regulatory requirements and best practices[8][6].

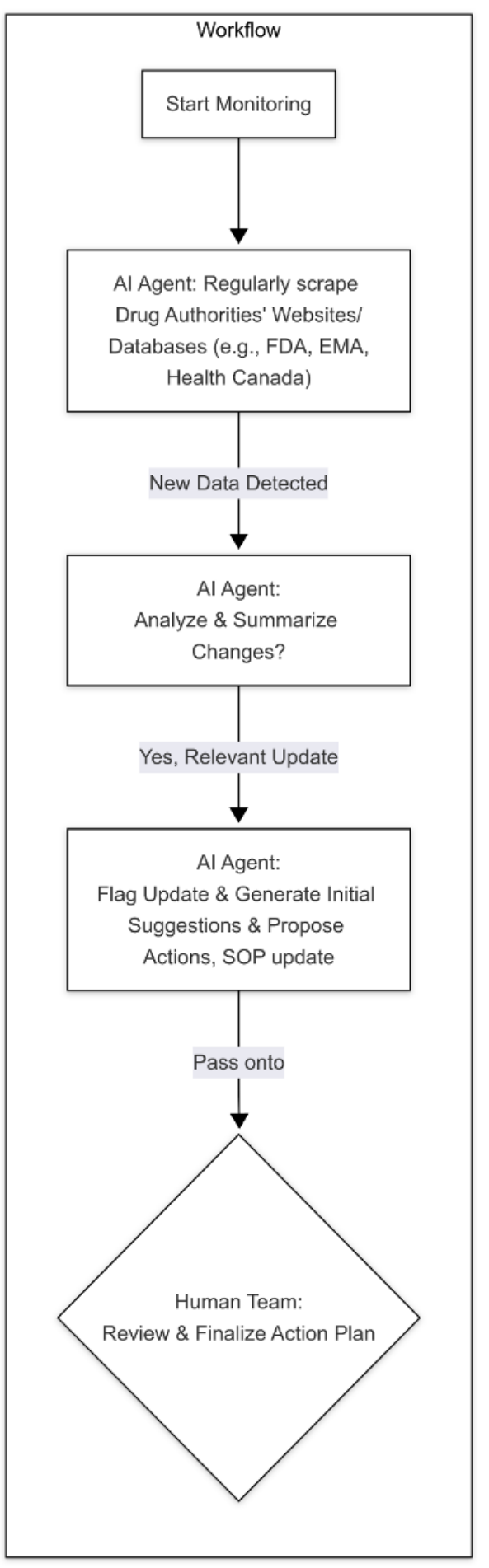

***Figure 1**: Possible agentic workflow in a pharmaceutical company*

Training Material Development: Agents can create customized training materials for different teams, ensuring staff remain current with evolving regulatory requirements and best practices[8][6].

GLOBAL EXPERIENCES: LLMs IN REGULATORY SUBMISSIONS

The use of LLMs in preparing pharmaceutical regulatory submissions overseas is moving rapidly from the experimental to the actual implementation phase, driven by the need for greater efficiency in managing vast amounts of regulatory documentation[9].

Document Drafting and Generation

Companies are using commercial LLMs (like specialized versions of GPT-4 or Gemini) to generate draft sections of regulatory submissions, such as initial versions of risk analyses, standard operating procedures (SOPs), and components of the Common Technical Document (CTD), particularly in non-clinical or administrative sections[9]. This includes drafting high-quality language for reports and compliance documents, which human experts then review and refine the "human-in-the-loop" approach[9][10].

Information Retrieval and Summarization

LLMs are used for rapidly summarizing extensive regulatory guidance, literature reviews, and clinical trial data. Specialized LLM-based tools can parse health authority guidelines to answer targeted compliance queries, significantly reducing research time for regulatory affairs professionals[10]. The EMA itself has introduced an AI-enabled knowledge mining tool, Scientific Explorer, for European Union (EU) regulators to facilitate quick, focused searching of regulatory scientific information[11].

Benefits

LLMs excel at processing and generating human-like text, which drastically accelerates the preparation of vast documentation required for submissions. For instance, they can draft high-quality initial sections of the CTD, such as administrative summaries, risk analyses, and SOPs. Furthermore, their capability for rapid summarization and targeted information retrieval from extensive regulatory guidelines and scientific literature cuts down the time regulatory affairs professionals spend on research, allowing them to focus on strategic review and critical analysis[12]. This is evident in tools developed by regulatory bodies like the EMA's Scientific Explorer, which leverages AI for knowledge mining to support faster regulatory decision-making[11].

Shortfalls

Despite the notable benefits, several critical shortfalls or challenges must be addressed for the safe and compliant use of LLMs in this highly regulated domain[13]. The primary concern is LLM "hallucination," where the models generate factually incorrect or nonsensical, yet plausible-sounding, information. This poses a significant risk if included in a submission concerning patient safety or product efficacy. Another major shortfall involves data quality and integration. LLMs are trained on enormous, diverse datasets of varying quality and provenance, making it difficult to guarantee their output is grounded in accurate, up-to-date, and jurisdiction-specific regulatory intelligence[13]. Closely related are issues of privacy and security, as general-purpose LLMs risk non-compliance with stringent data protection laws by potentially learning or exposing sensitive corporate or patient data during interaction. Finally, regulatory validation is complex because the proprietary nature of LLMs makes it challenging to provide regulatory agencies with the necessary transparency into how the model was built, trained, and verified, hindering the ability to perform robust quality assurance[14].

Resolutions

To address these shortfalls, regulatory bodies and the pharmaceutical industry are implementing targeted resolutions. The most critical resolution is the mandatory "human-in-the-loop" approach, where all LLM-generated content must be critically reviewed, cross-checked, and validated by qualified regulatory affairs professionals before submission[15]. To mitigate data quality and provenance issues, companies are moving towards developing and using curated, domain-specific LLMs trained exclusively on verified, high-quality, and proprietary internal data alongside certified regulatory texts[14]. Privacy and security concerns are addressed by using secure, enterprise-grade LLM platforms that offer robust data encryption, granular access controls, and strict adherence to data governance policies, ensuring sensitive input data is not used for model retraining. Furthermore, jurisdictions are developing risk-based regulatory frameworks to assess and validate AI/LLM tools, focusing on criteria like transparency, the quality of the training data, and continuous post-market monitoring, rather than demanding full model transparency[16].

HOW HONG KONG CAN LEARN FROM OTHER REGULATORY JURISDICTIONS

As Hong Kong moves towards establishing its own Centre for Medical Products Regulation (CMPR) and a "primary evaluation" approval pathway, it can learn from the experience of jurisdictions like the EMA and FDA.

Prioritize Regulatory Capacity Building

Following the lead of the EMA and FDA, Hong Kong must invest heavily in upskilling its regulatory staff in AI and LLMs to effectively review and assess AI-driven submissions. This includes establishing specialized AI-focused teams and continuous training[17].

Adopt a Risk-Based, Phased Approach

As the EMA's initial LLM guidance focuses on internal administrative tasks and the EU's AI Act is being phased in, Hong Kong should start by defining a clear, low-risk scope for LLM use (e.g., initial drafting, summarization) and then

gradually expand to higher-risk areas, always maintaining human oversight and a strong focus on validation.

Develop Clear Guiding Principles

The CMPR should quickly develop and publish guiding principles for both industry and its internal staff on the safe, responsible, and ethical use of LLMs, explicitly addressing issues like hallucination mitigation, data provenance, and the requirement for critical cross-checking of all LLM outputs (EMA's guiding principles are an excellent reference)[9].

Emphasize Data Standards and Interoperability

Hong Kong's shift towards a data-driven regulatory model (moving away from solely document-driven submissions) should align with international standardization efforts, such as the eCTD and the International Council for Harmonisation (ICH) frameworks. This will ensure data is machine-readable and ready for future AI/LLM applications, facilitating international recognition of Hong Kong's regulatory processes[18].

Focus on Regulatory Sandboxes

Implement regulatory "sandboxes" or pilot programs to test and refine governance frameworks for AI/LLM tools in a real-world, controlled setting before broad application[19].

RECOMMENDATIONS FOR HONG KONG PHARMACISTS

Education and Training: Hong Kong pharmacists should proactively engage in specialized AI literacy programs. These programs should not only cover the foundational understanding of transformer-based LLMs and their pharmaceutical applications but also critically address their inherent limitations, such as the potential for 'hallucinations' and algorithmic biases, which are paramount for patient safety in medication management. Training should leverage local resources like the Hong Kong Academy of Medicine (HKAM) AI Portal, which offers insights into ethical and legal aspects of AI in healthcare. Emphasis should be placed on developing the technical fluency to interpret AI-generated drug information, adverse event reports, and medication guidance, ensuring that human critical judgment always validates AI outputs before clinical application. This is crucial for pharmacists in their direct patient care roles, where precision and accuracy are non-negotiable.

1. **Pilot Programs**: Pharmacists in Hong Kong should advocate for and participate in small-scale pilot programs within their practice settings (e.g., community pharmacies, hospital pharmacies, or clinical units). These pilots could explore LLM applications in areas directly impacting patient care, such as generating initial responses to common drug information queries, assisting with medication reconciliation, or flagging potential drug interactions for pharmacist review. Given the current absence of specific AI guidelines for pharmaceutical products from the Hong Kong Department of Health or the Pharmacy and Poisons Board, these pilots must strictly adhere to a 'human-in-the-loop' design, where AI outputs are always reviewed and validated by a qualified pharmacist to ensure patient safety and compliance with existing regulations.

2. **Collaboration**: Hong Kong pharmacists should actively seek collaborations with local AI technology providers, academic institutions (such as the University of Hong Kong and Chinese University of Hong Kong), which are already engaged in medical AI discussions, and regulatory consultants. This collaboration is vital to ensure that AI tools are developed and implemented with a deep understanding of Hong Kong's specific healthcare context, patient needs, and the nuances of local pharmaceutical practice. Engaging with these partners can help tailor AI solutions that are not only technologically advanced but also ethically sound and culturally appropriate for the Hong Kong population.

3. **Regulatory Engagement**: Given the current absence of dedicated AI-specific legislation for pharmaceutical products in Hong Kong, pharmacists must maintain an active and proactive dialogue with key local health authorities. This includes the Pharmacy and Poisons Board (PPB), which oversees traditional drug registration. Pharmacists, with their direct involvement in medication use and patient safety, are uniquely positioned to provide invaluable practical insights and feedback to help shape future, specific guidelines for AI and LLM use in the pharmaceutical industry, ensuring they are robust, practical, and patient-centric. Adherence to general ethical AI frameworks and data privacy guidelines from the Digital Policy Office and Office of the Privacy Commissioner for Personal Data (PCPD) remains crucial during this evolving period.

CONCLUSION

The integration of LLM agents in the pharmaceutical industry represents a transformative opportunity. By understanding the underlying transformer architecture and self-attention mechanisms, pharmacists can better appreciate how these sophisticated AI systems can enhance regulatory efficiency, accuracy, and compliance.

As the pharmaceutical landscape continues to evolve, those who embrace and effectively implement these AI technologies will gain significant competitive advantages in drug development, regulatory approval, and market access. The key to success lies in thoughtful implementation that combines the power of AI automation with the irreplaceable value of human expertise and oversight.

For Hong Kong pharmacists, now is the time to begin exploring these technologies, understanding their capabilities and limitations, and preparing for a future where AI agents serve as invaluable partners in ensuring safe, effective medicines reach patients as efficiently as possible.

Author's background

Lam, Nate *is currently the Director at ElendiLabs Ltd in Hong Kong. He is also a registered pharmacist in Hong Kong. He holds a Bachelor's degree in Pharmacy and a Master's degree in Computer Science. His email address is nate.lam@elendilabs.com.*

References

1. Serrano, D. R., Luciano, F. C., Anaya, B. J., Ongoren, B., Kara, A., Molina, G., Ramirez, B. I., Sánchez-Guirales, S. A., Simon, J. A., Tomietto, G., Rapti, C., Ruiz, H. K., Rawat, S., Kumar, D., & Lalatsa, A. (2024). Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine. Pharmaceutics, 16(10), 1328. https://doi.org/10.3390/pharmaceutics16101328
2. Ildiz, M.E., Huang, Y., Li, Y., Rawat, A.S. & Oymak, S.. (2024). From Self-Attention to Markov Models: Unveiling the Dynamics of Generative Transformers. Proceedings of the 41st International Conference on Machine Learning, in Proceedings of Machine Learning Research 235:20955-20982 Available from https://proceedings.mlr.press/v235/ildiz24a.html.
3. 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.
4. Lin, Ryan & Ojha, Siddhartha & Cai, Kevin & Chen, Maxwell. (2024). Strategic Collusion of LLM Agents: Market Division in Multi-Commodity Competitions. 10.48550/arXiv.2410.00031.
5. Buscemi, A., Proverbio, D., Bova, P., Balabanova, N., Bashir, A., Cimpeanu, T., Correia da Fonseca, H., Duong, M. H., Domingos, E. F., Fernandes, A. M., Krellner, M., Ogbo, N. B., Powers, S. T., Santos, F. P., Shamszaman, Z. U., Song, Z., Di Stefano, A., & Han, T. A. (2025). Do LLMs trust AI regulation? Emerging behaviour of game-theoretic LLM agents. arXiv. https://arxiv.org/abs/2504.08640
6. Madalaimuthu, M. M. (2025). Regulatory in the Pharmaceutical Industry. Journal of Social Science and Humanities, 7(3), 37–42. https://doi.org/10.53469/jssh.2025.7(03).08
7. Han, Y., & Guo, Z. (2024). Regulator-Manufacturer AI Agents Modeling: Mathematical Feedback-Driven Multi-Agent LLM Framework. arXiv. https://arxiv.org/abs/2411.15356
8. Ahmed, S., & Bhati, S. (2022). Role of Regulatory Affairs in Pharmaceutical Industry. International Journal of Scientific Research in Science and Technology, 9(4), 376–380. https://doi.org/10.32628/IJSRST2293140
9. European Medicines Agency. (2024). Artificial intelligence: Guiding principles on the use of large language models in regulatory science and for medicines regulatory activities. Retrieved from https://www.ema.europa.eu/en/documents/other/guiding-principles-use-large-language-models-regulatory-science-medicines-regulatory-activities_en.pdf
10. IntuitionLabs. (2025, June 24). Artificial intelligence and LLMs in regulatory affairs. https://intuitionlabs.ai/articles/ai-llms-regulatory-affairs
11. European Medicines Agency. (2024). Scientific Explorer: Frequently asked questions (EMA/126785/2024). https://www.ema.europa.eu/system/files/documents/other/scientific-explorer-frequently-asked-questions_en.pdf
12. McKinsey & Company. (2024, January 9). Generative AI in the pharmaceutical industry: Moving from hype to reality. https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality
13. Balasubramanian, V. (2024, Januari). Large language models: Extracting and summarizing regulatory intelligence from health authority guidance documents. Global Forum, 16-19. https://globalforum.diaglobal.org/issue/january-2024/large-language-models-extracting-and-summarizing-regulatory-intelligence-from-health-authority-guidance-documents/
14. Ong, J. C. L., Chang, S. Y.-H., William, W., Butte, A. J., Shah, N. H., Chew, L. S. T., Liu, N., Doshi-Velez, F., Lu, W., Savulescu, J., & Ting, D. S. W. (2024). Ethical and regulatory challenges of large language models in medicine. The Lancet Digital Health, 6(6), e428–e432. https://doi.org/10.1016/S2589-7500(24)00061-X
15. D. Al-Turki et al., "Human-in-the-Loop Learning With LLMs for Efficient RASE Tagging in Building Compliance Regulations," in IEEE Access, vol. 12, pp. 185291-185306, 2024, doi: 10.1109/ACCESS.2024.3512434.
16. European Commission. (2024). Regulatory framework for AI. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
17. Singh, R., Paxton, M. & Auclair, J. Regulating the AI-enabled ecosystem for human therapeutics. Commun Med 5, 181 (2025). https://doi.org/10.1038/s43856-025-00910-x
18. Macdonald, J. C., Isom, D. C., Evans, D. D., & Page, K. J. (2021). Digital Innovation in Medicinal Product Regulatory Submission, Review, and Approvals to Create a Dynamic Regulatory Ecosystem-Are We Ready for a Revolution?. Frontiers in medicine, 8, 660808. https://doi.org/10.3389/fmed.2021.660808
19. Ong, Jasmine & Ning, Yilin & Liu, Mingxuan & Ma, Yian & Liang, Zhao & Singh, Kuldev & Chang, Robert & Vogel, Silke & Lim, John & Tan, Iris & Freyer, Oscar & Gilbert, Stephen & Bitterman, Danielle & Liu, Xiaoxuan & Denniston, Alastair & Liu, Nan. (2025). Regulatory Science Innovation for Generative AI and Large Language Models in Health and Medicine: A Global Call for Action. 10.48550/arXiv.2502.07794.