IMO Intelligent Medical Objects

ChatGPT and generative AI in healthcare

Based on the Melax Tech webinar, *ChatGPT and Generative AI: Paradigm Shift in Healthcare and Biopharma Innovations*

ChatGPT and other generative artificial intelligence (AI) models are already disrupting how we write and retrieve information across industries, prompting great speculation about how they will improve – and potentially pose new risks – in healthcare.

IMO data scientists and biomedical informaticists are immersed in this world, exploring and exploiting how generative AI is shifting the paradigm in natural language processing (NLP) research and applications within healthcare and the biopharmaceutical realm. Based on views and insights they shared in the webinar, *ChatGPT and Generative AI: Paradigm Shift in Healthcare and Biopharma Innovations**, this paper reviews the evolution of large language models (LLMs) including ChatGPT, how they compare to traditional NLP, and opportunities and potential limitations for ChatGPT and similar tools in the healthcare industry.

ChatGPT 101

The ChatGPT AI chat bot released on November 13, 2022, is one of the fastest growing applications ever, gaining more than one million users in five days and 100 million active monthly users just two months after launch[1]. Developed by OpenAI, ChatGPT is a generative pre-trained transformer (GPT) model trained to understand natural language and code.

To level set, a few definitions may prove helpful:

- ChatGPT is an NLP tool, part of a category referred to as large language or foundation models.
- NLP is a type of AI that helps enable computers to understand words as humans do.
- Lastly, generative AI uses a blend of supervised and unsupervised machine learning models to create new content that includes images, videos, audio, text, and even programming codes.

ChatGPT uses data from across the internet to create text outputs in response to inputs, or "prompts." Training or programming a GPT model is done by creating prompts that provide instructions or examples of how to successfully complete a task.

What sets ChatGPT apart is that it follows a process called reinforcement learning from human feedback (RLHF), which incorporates human responses into training. In addition to a more traditional unsupervised pre-training process, with prompts that are a natural language description of the task, ChatGPT includes human review of model output and integrates feedback that further helps the model understand the human intention. As a result, ChatGPT can understand and answer almost any human question. And you don't have to be trained as a prompt engineer to use it.

The evolution of NLP and generative AI

NLP is nothing new. Four mainstream paradigms in NLP reveal how this capability evolved to become the powerful tool it is today.

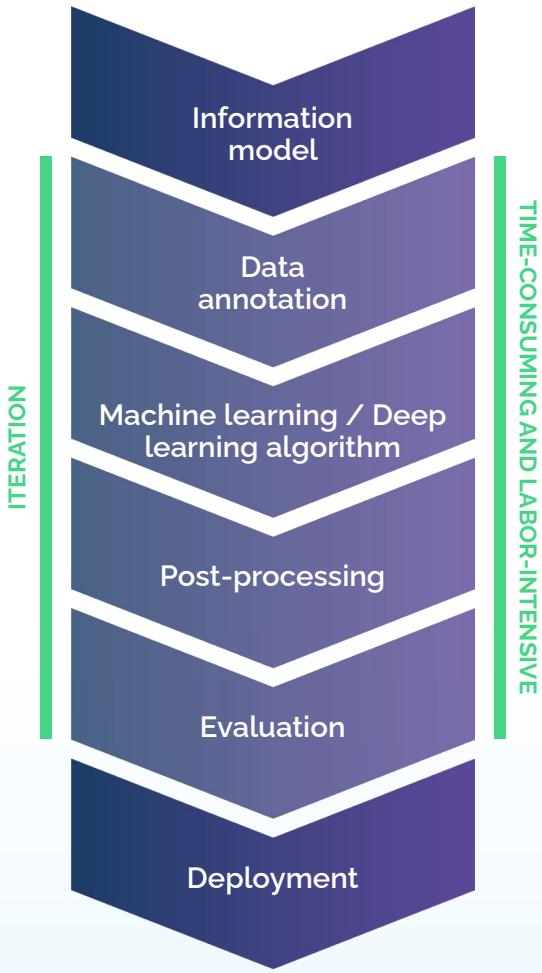
- **Symbolic NLP**: Also called rule-based NLP, it was developed in the 1980s and used a comprehensive set of rules and keyboard-based features to identify predetermined patterns in text.
- **Neural networks**: Around 2,000 statistical machine learning models such as Random Forest (RF) Support Vector machines (SVM) were applied to solve NLP tasks. Since training neural networks requires high computing power, this phase picked up when the processing power of graphics processing unit (GPU) chips came to high-performance computers for resource-intensive tasks.
- **Transformer models**: These were introduced by Google in 2017 in the paper, "Attention is All you Need," and became the NLP approach of choice. With transformer models, an attention mechanism adds weighting for different words in the input and natural language for specific NLP tasks, creating an approach that effectively captures long-term dependencies while significantly improving performance with less training data.
- **Generative NLP**: The new NLP paradigm is often called generative NLP and started with the birth of GPT models.

ChatGPT and the traditional NLP development cycle

ChatGPT and generative AI dramatically change the approach to NLP. Historically, different methods were required for different NLP tasks, such as name entity recognition, sentiment analysis, and text summarization. And each NLP task required specific training data, testing data, validation data, and the creation of different models. With ChatGPT, one model fits all.

In addition, ChatGPT reduces the time-consuming and labor-intensive NLP step of fine tuning. A typical NLP task starts with an information model, which defines the elements and what one wants to extract from the textual data. The steps to develop the model include data annotation, training of the machine learning deep learning algorithm, and post processing to add rule dictionaries on top of the machine learning results. Algorithm performance is then evaluated by a domain expert who checks the results.

FIG 1: **NLP Development Cycle: Fine-Tuning**



Fine-tuning is an iterative process, repeating each of these steps several times until satisfying performance and results are achieved.

Other NLP practices shorten the development cycle with varying impacts on the results. ChatGPT offers different levels of "shot learning," which dictate the number of examples provided to train the model. With **few-shot** learning one can give a model two to three examples, far fewer examples than with the iterative process of fine tuning. With **one-shot** learning the model is provided with one example. And with **zero-shot** learning one simply provides a plain English description of the task, without any examples, and the model will execute the prediction automatically. These approaches are faster and make reasonably good predictions, but typically can't match the performance achieved with fine tuning.

NLP practitioners always strive for few-shot or zero-shot learning because it can dramatically shorten the NLP development work cycle. However, for medical writing, information retrieval, and healthcare NLP applications that involve a high degree of clinical and biomedical nuance, the process of fine tuning and the additional effort for human intervention and expertise are necessary to ensure the process is steered correctly.

Generative AI in clinical and biomedical applications

Free text, unstructured data across healthcare – in clinical notes, real-world evidence, medical literature, pathology notes, clinical trial safety reports, and more – hold critical information, yet require sophisticated use of NLP to efficiently extract and produce insights that are meaningful and reliable.

For clinical notes in electronic health records (EHRs), NLP can extract important information accurately and at scale, allowing for downstream analysis. It can create real-world evidence journeys to track patients' diseases, symptoms, conditions, and treatments across multiple health systems and millions of patients. And systematic literature reviews (SLRs) can be automated to extract and summarize information from different articles, saving time and effort. GPT technology can also help optimize clinical trials, allowing for the easy extraction of variables and endpoint measures, clinical trial design, cohort definition, and protocol writing. Other clinical and biomedical applications of NLP include:

- Clinical named-entity extraction
- Clinical relation extraction
- De-identification of protected health information (PHI)
- Clinical text summarization
- Automatic literature review
- Biomedical literature tables and OCR processing
- Sentiment analysis and misinformation detection of social media and news
- Post-marketing safety analysis and report analysis
- Research protocol writing

Limitations of ChatGPT and generative AI

Current ChatGPT and generative AI models present several challenges to producing sound results, especially when findings and

recommendations impact health, treatments, and potentially life-or-death care decisions. Five of these challenges are discussed below.

Misinformation: Models built on limited and/or incomplete medical training data will not be optimized in the biomedical domain, therefore limiting knowledge. Even when humans add expertise by creating and refining prompts, there is variability in results based on the level of domain expertise involved.

Hallucinations: This one is a bit alarming: these models will sometimes make things up. IMO data scientists have encountered several examples where ChatGPT produced information derived from completely fabricated references.

Transparency: It is not clear how AI models reach their conclusions. Since the models are not open to the public, we also don't know what data is embedded in each model. For example, models trained on data with innate biases can produce results that are also biased.

Originality: Protecting and differentiating content written by a real person is a high-profile concern already being debated in mainstream media and discussed in academic journals. Authorship, plagiarism, and copyrights are issues that will require extensive consideration.

Recency: Most large language models (LLMs) are not constantly being trained. They learn and are built based on a finite training time and data set and are at risk of becoming out of date quickly. In healthcare, where the body of knowledge is growing and changing rapidly, ensuring accurate and current information is of particular concern.

Using ChatGPT in healthcare today

As experts in clinical and biomedical NLP, IMO has perspective on how organizations can leverage ChatGPT and LLMs in healthcare today.

FIG 2: **Current utilities of ChatGPT and LLMs in healthcare NLP tasks**

Information extraction	
Powerful across multiple IE tasks & domains	**Go**
Terminology support is weak	**Use with confidence**
Knowledge generation	
Very knowledgeable across multiple domains	**Go**
Hallucinations happened with inaccurate references	**Use with caution**
Decision making	
Great potential in automating many tasks	**Use with greater caution, especially on critical functions**

For NLP tasks involved in **information extraction**, we feel organizations can proceed with confidence. There are some weaknesses based on the level of expertise used in the prompts, however one strength of ChatGPT is that it can quickly extract and summarize information.

For uses that require **knowledge generation**, we recommend using ChatGPT with caution. The current models are very knowledgeable across multiple domains, however, they are still prone to misinformation and hallucinations.

Finally, for any project or task undertaken with the goal of **supporting a decision**, our recommendation is to use with even greater caution. As in the other categories, ChatGPT has great potential to automate and speed up many of the tasks of NLP. However, when

the results are being used to inform decisions and the stakes are higher, we believe there is significant risk and more work is required to ensure results are trustworthy, especially in the realm of healthcare.

Conclusion

ChatGPT is a breakthrough for NLP and for digital innovation. And there will undoubtably be rapid and exponential growth in these technologies, especially with GPT-4 now available, and a Google healthcare-specific LLM, Med-PaLM 2, in development.

Generative AI and LLMs such as ChatGPT are disrupting the traditional approach to medical writing and information retrieval, shifting the paradigm in NLP research and projects. While they have limitations, they bring efficiency to solving NLP-based problems and offer a low barrier to entry for users who perform multiple NLP tasks.

In healthcare, where the body of knowledge is growing and changing rapidly, and when medical training data is often not optimized in the biomedical domain, ensuring accurate and current domain expertise is required for reliable results. If we can inject the process with clinical and biomedical knowledge, we can inform prompt design, ensure fine tuning, and use the technology properly to get to a cheaper, better system with results we can trust.

To learn more, watch the on-demand webinar, *ChatGPT and Generative AI: Paradigm Shift in Healthcare and Biopharma Innovations*.

To discover how IMO solutions apply NLP to improve the value and usability of unstructured clinical data, visit imohealth.com/nlp.

IMO Intelligent Medical Objects

About Intelligent Medical Objects

Intelligent Medical Objects is a healthcare data enablement company that ensures clinical data integrity and quality—making patient information fit-for-purpose across the healthcare ecosystem, from hospitals to health information exchanges to payers, and beyond.

IMO's vast footprint in EHRs powers our ability to capture and preserve clinical intent at the highest level of specificity. Our secure technology platform and products then help our clients to transform and extract the greatest value from their data. In short, IMO is the catalyst that enables accurate documentation, precise population cohorting, optimized reimbursements, robust analytics, and better care decisions to optimize patient outcomes.

* The webinar was facilitated by Mitch Higashi, PhD. Participants included:

- Dr. Jingcheng Du, Vice President, Life Science Solutions at IMO, and Adjunct Assistant Professor at UT Houston
- Dr. Xiaoyan Wang, Senior Vice President, Life Science Solutions at IMO
- Dr. Yanshan Wang, Assistant Professor and Vice Chair of Research in the Department of Health Information Management at University of Pittsburgh

[1] Antipandey, J., Open AIs ChatGPT Breaks User Adoption Rates to 1 Million, Analytics Insight, January 30, 2023, https://www.analyticsinsight.net/openais-chatgpt-breaks-user-adoption-rates-to-1-million/