Look Forward:
What Generative AI Means For Healthcare

Written & Edited By

Nate Shames, Head of Healthcare

Mario Martone, Senior Machine Learning Engineer

Marjorie Schillo, Machine Learning Researcher

Contents

Introduction

This whitepaper explores the transformative potential of applying Natural Language Processing (NLP) and Generative AI to non-clinical aspects of healthcare.

Harnessed properly, this technology can achieve significant results by automating administrative tasks, appointment scheduling, optimizing resource allocation, and enhancing operational efficiency within healthcare systems. Through the power of quality, real artificial intelligence, institutions can streamline appointment scheduling, automate billing processes, and refine inventory management.

Let's examine the history of NLP and Generative AI, and the machine learning techniques used by EliseAI specifically. As healthcare evolves, embracing Generative AI stands as a key to survival and financial success as demand on facilities heightens and operations become more costly.
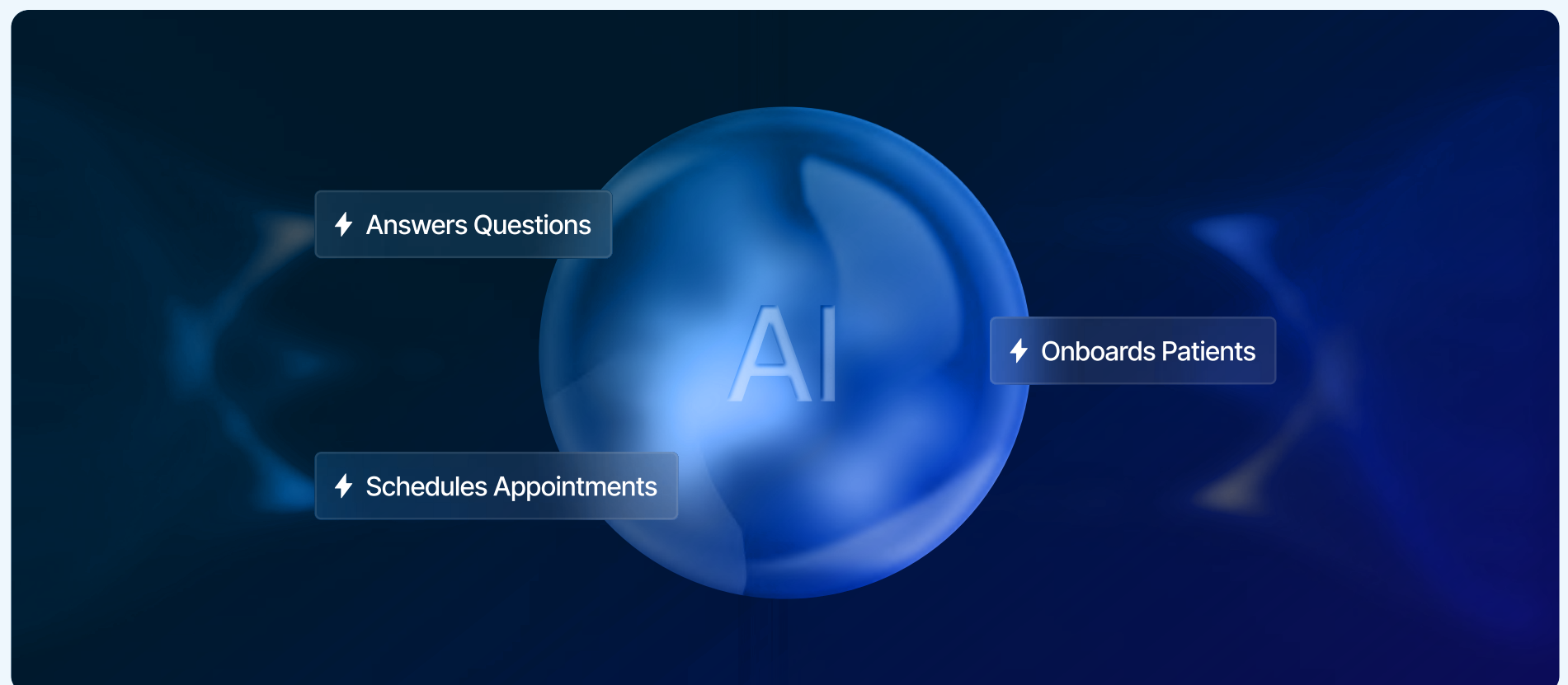
The History of NLP

Before examining the application of current models to your present use case, it's important to know the history of NLP and generative AI, illuminating the true potential of the technology and the timeliness in adopting it now.

See the evolutionary trajectory of **Natural Language Processing (NLP)** and **Generative AI**, spanning from the very first chatbot of the 1960s to the cutting-edge advancements witnessed in the present day.

1960s – 2000s

Pattern-matching, statistical approaches and rules

The first chat-bots (most notably Eliza, devised by Joseph Weizenbaum in 1966) were able to attempt the Turing test. **Rudimentary pattern-matching techniques** were used by these chatbots to initiate experimentation. More r**esearch emerges, focused on statistical approaches**, wherein information theory, probabilistic models, and data-driven methodologies. enable a modeling of language patterns from extensive datasets (see e.g. IBM's Candide.) A machine translation system **SYSTRAN** is developed to focus on language manipulation through predefined grammatical and syntactical rules. While somewhat effective in specific domains, these types of **rule-based and statistical models** failed to produce truly natural conversation.

2010-2018

Rise of Machine Learning

The resurgence of interest in NLP during the 2010s was fueled by the emergence of machine learning. Artificial intelligence models known as **neural networks had increased computational power** and could use larger text datasets than before, making them more conversationally sophisticated. Early architectures such as **recurrent neural networks (RNNs)** and **long short-term memory (LSTM) networks** demonstrated impressive capabilities in language translation and sentiment analysis. **The transformer architecture** revolutionized the field by introducing attention mechanisms in the seminal "Attention is All You Need" paper by Vaswani et al. in 2017.

2018 – Present
Natural Language Understanding (NLU) and the GPT Era

Now AI doesn't just process natural language but understands it deeply, from human-like NLP to human-like NLU. A single NLU model can compute and complete diverse, open-ended, and complex tasks. The early front-runner in the development of this new breed of generative models is undoubtedly **OpenAI**.

2019   **GPT-2** demonstrated unparalleled language generation capabilities.

2020   **GPT-3** further expanded AI's boundaries and skills, able to complete multifaceted language tasks, from translation to code generation.

2023   **GPT-4** improved on the advancements of GPT-3 and sparked a wave of competition among infrastructure providers.

The rest of 2023 saw the release of **Google's Gemini, Meta's LLaMA 2**, and **Anthropic's Claude 2.1.**

Future
What's Next For Artificial Intelligence

The trajectory of NLP and Generative AI promises continual innovation. Ongoing advancements in models, AI training techniques, and their ethical frameworks will shape the landscape. The integration of multimodal capabilities, combining language understanding with visual and auditory inputs, holds immense potential for the development of comprehensive AI systems.

Main NLP Architectures & Techniques

Several machine learning techniques are used to enable NLP comprehend, interpret and generate human-like language.

Architecture

The revolutionary **transformer architecture**, introduced in the "Attention is All You Need" paper, redefined NLP. Transformers capture contextual relationships between words in their sequence, like in a sentence or paragraph, enabling deep processing and enhanced modeling.

The most flexible and adaptable models use **transformers** in a sequence-to-sequence architecture. These models take input sequences and generate corresponding output sequences, making them versatile for tasks requiring a one-to-one or many-to-many mapping.
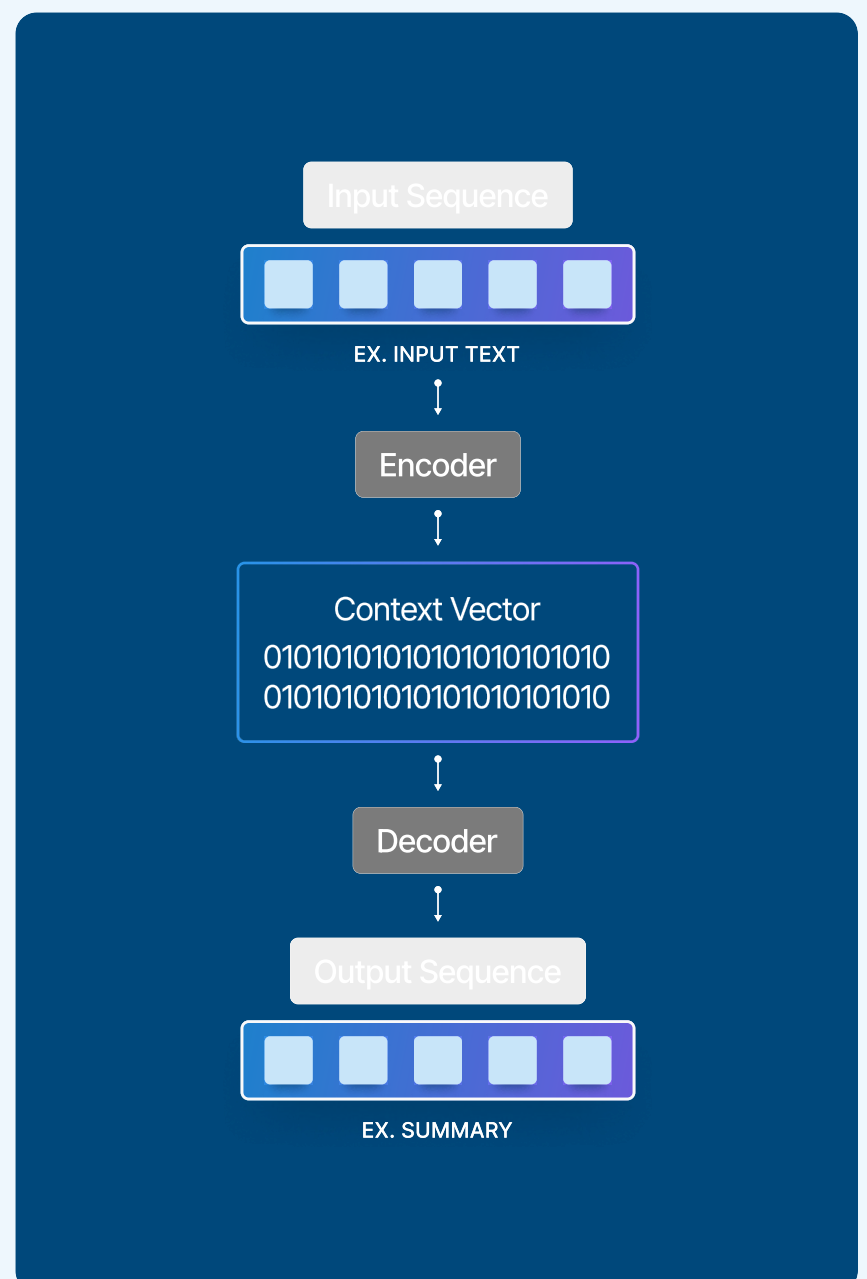
Training Techniques For Models

Another powerful class of algorithms in the realm of Generative AI are **Generative Adversarial Networks (GANs)**. GANs consist of a generator and a discriminator, engaged in a competitive process. The generator creates synthetic data, while the discriminator distinguishes between real and generated samples. Through iterative training, GANs enhance the generator's ability to produce realistic data that can be indistinguishable from the training set.

Four different training techniques used to produce contextual, conversational AI.

Supervised Learning

✓ Models trained on labeled datasets, and learning patterns to make predictions.

✓ Smaller, faster models specialize in a specific task, and predict limited outcomes.

Unsupervised Learning

✓ Models trained on unlabeled data, allowing AI models to independently identify patterns and structures.

✓ Models learn from essentially all of human-kind's collected texts.

✓ Models generate meaningful conversations, capture relationships and contextual nuances without guidance.

Transfer Learning

✓ Unsupervised pre-training models first on vast unlabeled datasets so they get a broad understanding of language.

✓ Fine-tuning them using smaller labeled datasets.

✓ Improves efficiency and performance.

Reinforcement Learning

✓ Model is given a reward function and learns to optimize its output to maximize rewards.

✓ Incorporates human feedback and response rate as part of reward function.

✓ Notably used to train ChatGPT to produce human-quality language.

How To Evaluate An NLP Model

The release of GPT-4 constitutes the first milestone in achieving general human-level comprehension of language.

The Generative Model

💬	👤	⚡
Converses Fluently	Has super-human expertise across multiple fields.	Accomplishes tasks with machine-optimized speed.

GPT-4 can pass a mock coding interview for a software engineering position at Amazon with a perfect score and matches the average human performance on the Multistate Bar Exam. It takes 4 minutes to complete the 2 hour coding exam. However, the same model will also fail to complete grade-school level arithmetic problems. And perhaps most troublesome are hallucinations, where the AI will present completely non-factual information with confidence.

The widely publicized successes of GPT-4 and ChatGPT have spurred development of other models with human-level capabilities across many tasks. In an attempt to make evaluation more rigorous, developers have established benchmark datasets with corresponding metrics. For example: **Google's PALM2** outperforms GPT4 on advanced common sense reasoning, as measured by the WinoGrande benchmark. **Anthropic's Claude2** has been developed to specifically excel in honesty and to exclude toxic content. Models such as **LLaMA**, **Vicuna**, and **Orca** all claim to be comparable to GPT-4 while being open-sourced, cheaper, faster, or smaller.

Given the growing menagerie of highly capable and flexible models, evaluating model performance in real-world applications is a critical and challenging task. Benchmark datasets usually fail to capture the context and the evolving character of realistic enterprise data.

EliseAI has developed an effective performance evaluation pipeline that overcomes these difficulties.

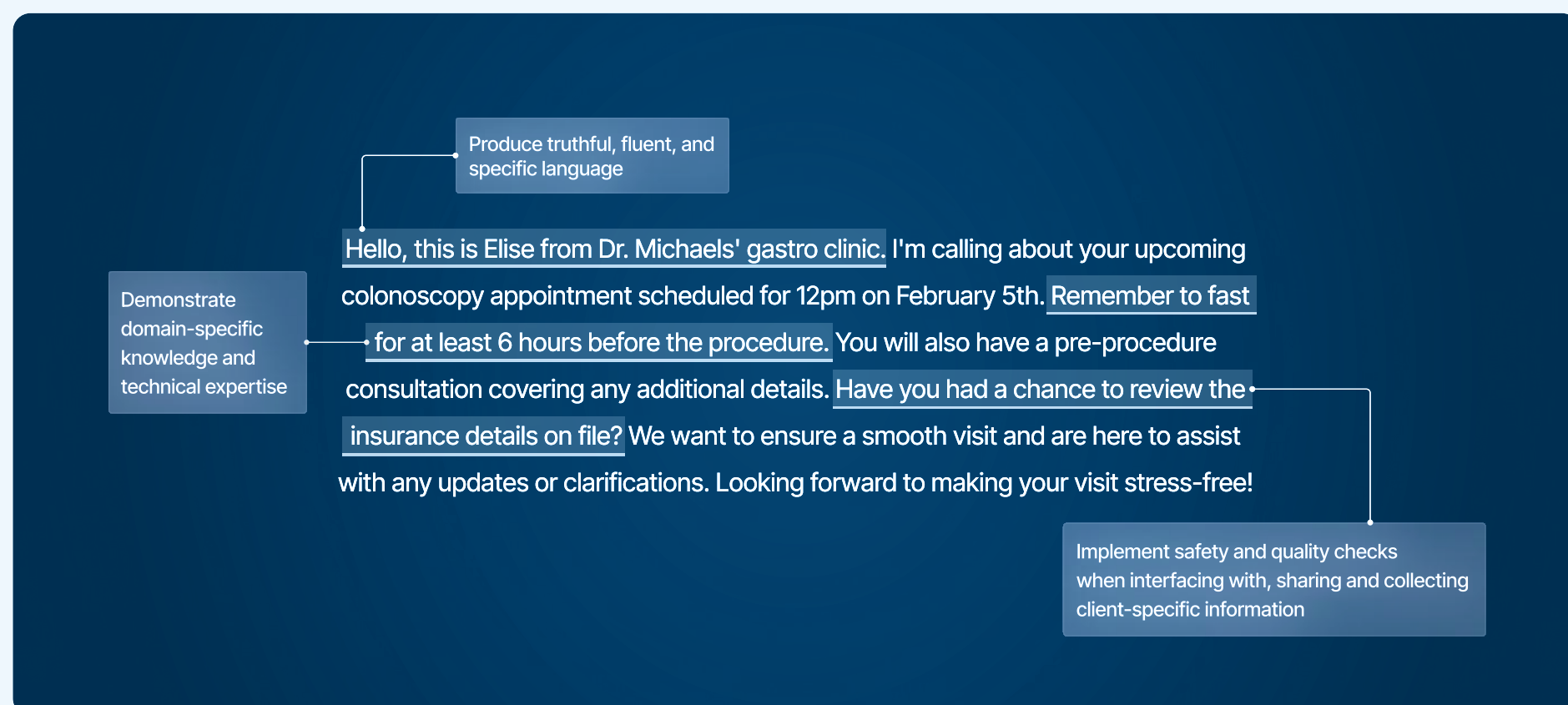|		
▤	📄	↺
Careful metric selection and dataset creation.	Extensive testing of AI by domain experts in business-relevant context.	An iterative approach considering user feedback.

EliseAI's Approach To Artificial Intelligence

While there are now several models that can achieve near-human to super-human performance in language-based tasks, no single model has yet come even close to **Artificial General Intelligence (AGI)**. Meaning, models can't reliably extrapolate knowledge to perform simple tasks, despite being highly proficient at more complex tasks.

EliseAI has designed a system that incorporates a team of model '**agents**'. EliseAI is able to take advantage of the tremendous power of these models within a secure and reliable framework.

This system is modular and able to evolve and improve over time. EliseAI can conduct domain specific conversations and demonstrate technical expertise. Each model can implement safety and quality checks as well as interface with client-specific knowledge. With the models acting in a framework, their planned construction and division of labor ensures that tasks can be performed in parallel, keeping latency at a minimum. The resulting conversations achieve human-quality language and interactions while minimizing the risk of model-hallucination.

Each model agent can be trusted to...

Produce truthful, fluent, and specific language

Demonstrate domain-specific knowledge and technical expertise

Hello, this is Elise from Dr. Michaels' gastro clinic. I'm calling about your upcoming colonoscopy appointment scheduled for 12pm on February 5th. Remember to fast for at least 6 hours before the procedure. You will also have a pre-procedure consultation covering any additional details. Have you had a chance to review the insurance details on file? We want to ensure a smooth visit and are here to assist with any updates or clarifications. Looking forward to making your visit stress-free!

Implement safety and quality checks when interfacing with, sharing and collecting client-specific information

Harnessing AI
For Enterprise

This technical overview outlines key considerations for integrating Generative AI seamlessly into non-clinical healthcare processes, with a particular emphasis on integrations and process automation.

1. Understanding Non-Clinical Challenges

The initial step in deploying Generative AI in non-clinical operations is to comprehend its unique challenges. Challenges can revolve around administrative complexities, such as specific appointment scheduling based on the doctor's schedule and speciality or processing claims and billing with contextual awareness. Generative AI can be leveraged to automate these non-clinical tasks, freeing up resources for more value-added activities.

2. Customization for Workflow Alignment

Tailoring Generative AI solutions to align with existing non-clinical workflows is imperative. The technology should seamlessly integrate into daily operations without causing disruption. Customization ensures that administrative processes are enhanced rather than overhauled, promoting smooth adoption among healthcare administrative professionals.

3. Identification of Use Cases

Clearly define the non-clinical use cases for Generative AI. Consider automating appointment scheduling, optimizing billing procedures, or facilitating insurance claim processing. Identifying specific areas where AI can alleviate administrative burdens allows for a targeted and effective deployment strategy.

4. Data Integration and Security

Data security remains a priority, requiring adherence to healthcare regulations such as HIPAA. Implementing robust encryption, access controls, and audit trails ensures the protection of sensitive administrative information.

5. Process Automation in Administrative Tasks

Generative AI can play a pivotal role in automating various administrative tasks. From appointment scheduling to billing and claims processing, the implementation of AI-driven natural language understanding facilitates the automation of routine, time-consuming tasks. This results in increased operational efficiency and allows administrative staff to focus on more strategic and patient-centric activities.

6. User Training and Adoption:

Conduct comprehensive training sessions for administrative staff to ensure smooth adoption of Generative AI tools. Addressing any concerns or skepticism about AI in administrative workflows is crucial. Training programs should highlight the practical benefits of the technology, showcasing how it streamlines tasks and contributes to overall workflow efficiency.

8. Scalability and Future Readiness

As healthcare technologies evolve, Generative AI solutions should be adaptable to new integrations, updates, and emerging non-clinical use cases. This future-ready approach ensures that the healthcare enterprise remains at the forefront of technological innovation in administrative operations.

7. Collaboration with IT and AI Experts

Close collaboration between administrative staff, IT specialists, and AI experts is essential for successful deployment. This collaborative approach ensures that Generative AI solutions are not only technically robust but also tailored to the unique needs of healthcare administrative workflows. Regular communication channels should be established.

Technological Barriers To VoiceAI

Building a generative AI application that harnesses voice technology comes with its unique set of challenges, blending the complexities of natural language processing (NLP) and the intricacies of capturing and understanding human speech.

CHALLENGE 1

Achieving a high level of accuracy in speech recognition

- ✓ Accents, dialects, and variations in speech patterns pose challenges.

- ✓ VoiceAI adapts to diverse linguistic nuances by learning vast and diverse datasets to ensure reliability across different user demographics.
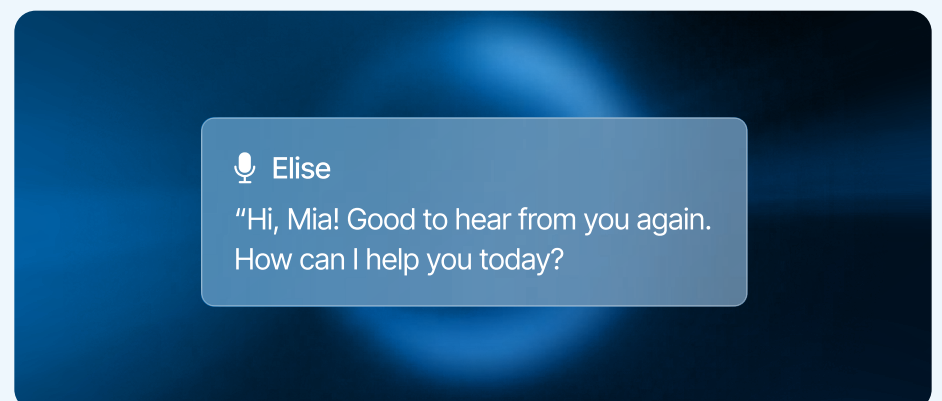
CHALLENGE 2

Ensuring privacy and security for patients

- ✓ Robust encryption and authentication mechanisms safeguard sensitive information.

- ✓ Strikes the right balance between personalization and privacy.

CHALLENGE 3

Generating coherent and contextually relevant responses

- ✓ Transcribes the words accurately, comprehends meaning, and provides contextual responses.

- ✓ Requires sophisticated algorithms and continuous refinement through user feedback.

🎙 Elise
"Hi, Mia! Good to hear from you again. How can I help you today?

CONCERN

Hi, Elise. I've been experiencing some digestive issues. Can you help me schedule an endoscopy appointment?

APPOINTMENT

Conclusion

Machine learning, Natural Language Processing (NLP), and Generative AI have unveiled a realm of possibilities and innovations for healthcare.

The historical journey from the early days of rule-based systems to the contemporary era of advanced generative models reflects the relentless evolution of technology which is continuing at its fastest ever pace. The infusion of machine learning techniques, including supervised and unsupervised learning, neural networks, and the transformer architecture has elevated the capabilities of NLP and paved the way for Generative AI to emerge as a potent force.

Our exploration of NLP and Generative AI models has revealed their immense potential to redefine non-clinical operations in healthcare. From automating administrative tasks to enhancing communication through voice generation, these models hold the promise of streamlining workflows, improving efficiency, and ultimately contributing to a more patient-centric healthcare ecosystem.

The specific requirements outlined for implementing NLP and Generative AI in an enterprise underscore the need for a holistic approach. From robust data integration and security measures to tailored model customization aligned with existing workflows, the key lies in meticulous planning, collaboration, and a clear understanding of your network's unique needs.