hyro

Generative AI in Healthcare: How to Not Be Left Behind

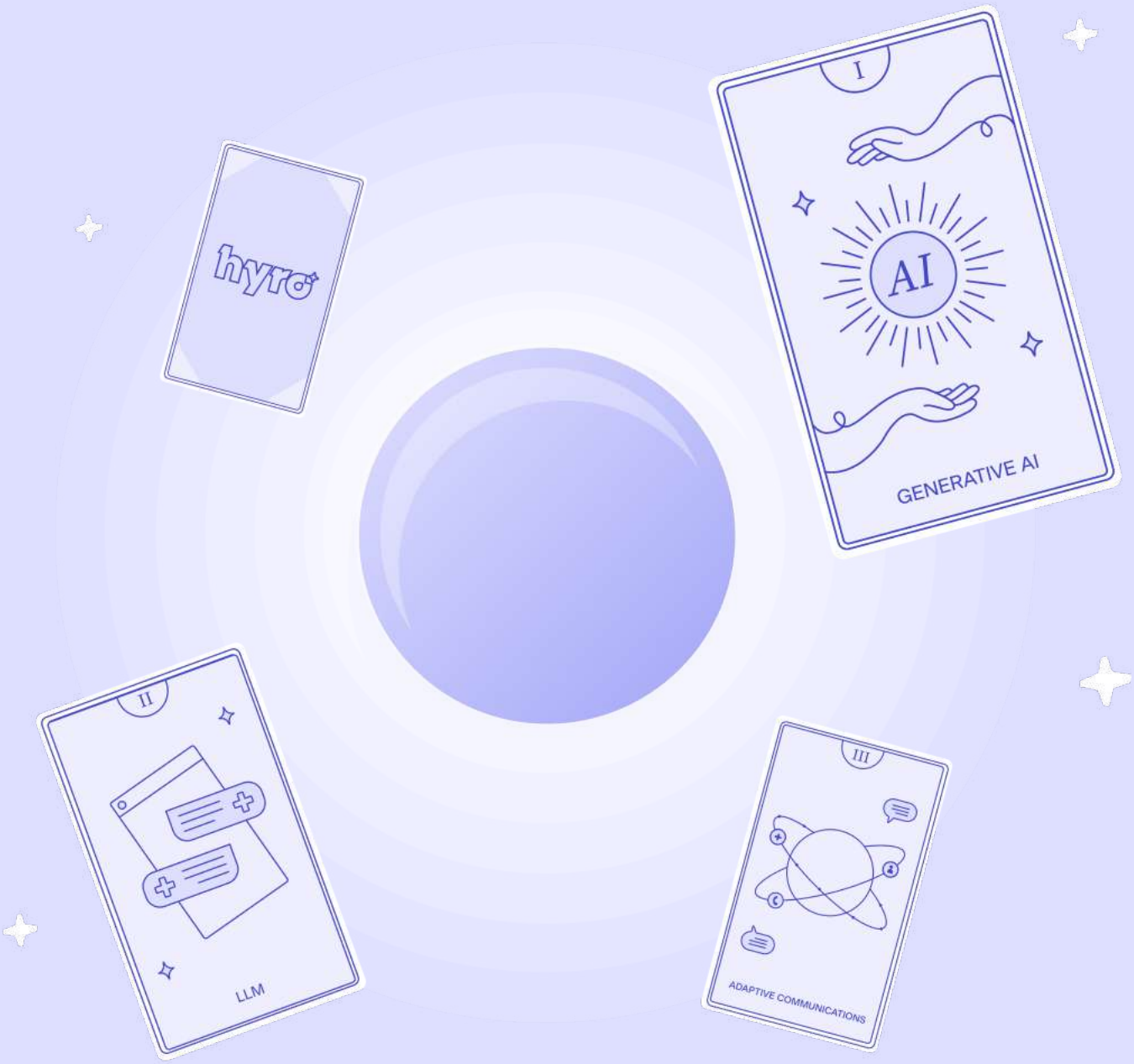
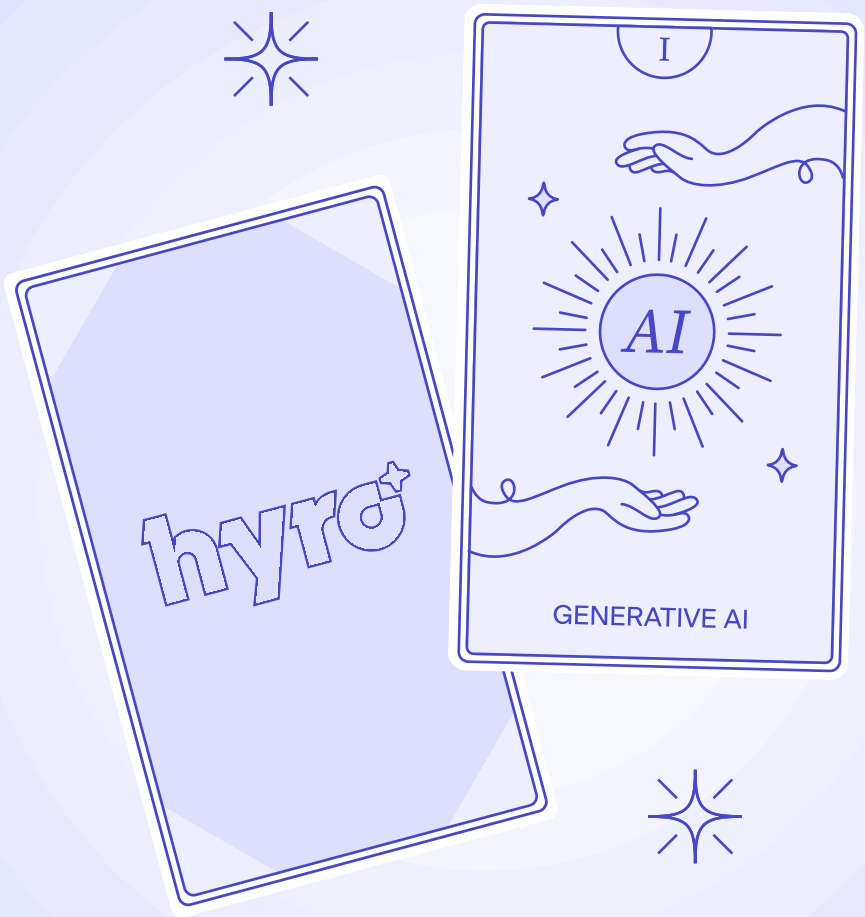


TABLE OF CONTENTS

01

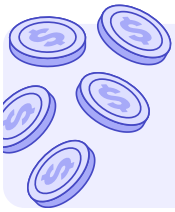
Introduction: The Rise of Generative AI



1.1

Is Your Health System Prepared for the Rise of Generative AI?

Generative artificial intelligence is an innovation whose time has come—and one that healthcare organizations cannot afford to ignore. Within five years, generative AI is expected to become a predominant component of all aspects of patient communication and engagement. It will transform workflows, improve patient and clinician experiences, and lead to historic cost-cutting across the entire sector. In fact, a recent report from McKinsey and Harvard found that the US healthcare industry could save up to $360 billion annually with wider adoption of generative AI.



$360B annual projected cost savings for the US healthcare industry with wider adoption of generative AI.

Your introduction to the world of *generative AI*

This handbook offers a comprehensive healthcare-oriented introduction to generative AI, covering its scope, limitations, and why hospitals and health systems must adopt it now.

We will explore pending issues and areas of concern and delve into what it means to use generative AI tools, including conversational AI, **responsibly**. Finally, we'll paint a clear picture of an effective solution in action—one that drastically improves patient access, patient engagement, and operational efficiencies within health systems while remaining HIPAA-compliant and safe for patient-facing use.

What are you going to learn?

- Why health systems must adopt generative AI now
- Risks and promises of large language models (LLMs)
- What a safe gen AI solution for healthcare looks like
- The architecture behind Hyro's gen-AI powered platform

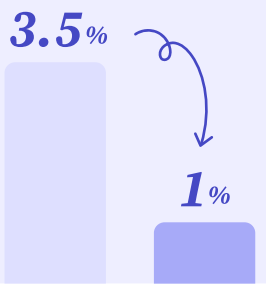
1.2

High Expenses, Razor-Thin Margins

The World Health Organization (WHO) projects a global ***shortfall*** of 10 million healthcare workers by 2030. Meanwhile, hospital labor ***expenses*** have increased by 20% since 2020.

Hospitals are still struggling financially post-pandemic, with healthcare consulting firm Kaufman Hall reporting that hospitals' operating margins dropped from 3.5% in August 2023 to 1% in September 2023. To navigate these choppy waters, hospitals are directing efforts toward cost reduction and efficiency enhancements, closely analyzing patient engagement and experience, and scrutinizing both the net patient revenue impact and service quality delivered by their call center support teams.

Figure 1: Declining Operating Margins



According to Hyro's State of Call Centers Report, 74% of surveyed call center leaders feel pressured by their system's executives to demonstrate that their call center could become a profit generator rather than a cost center. Additionally, 39% of respondents identified staff burnout and turnover as the primary culprits behind inefficiencies in their call centers.

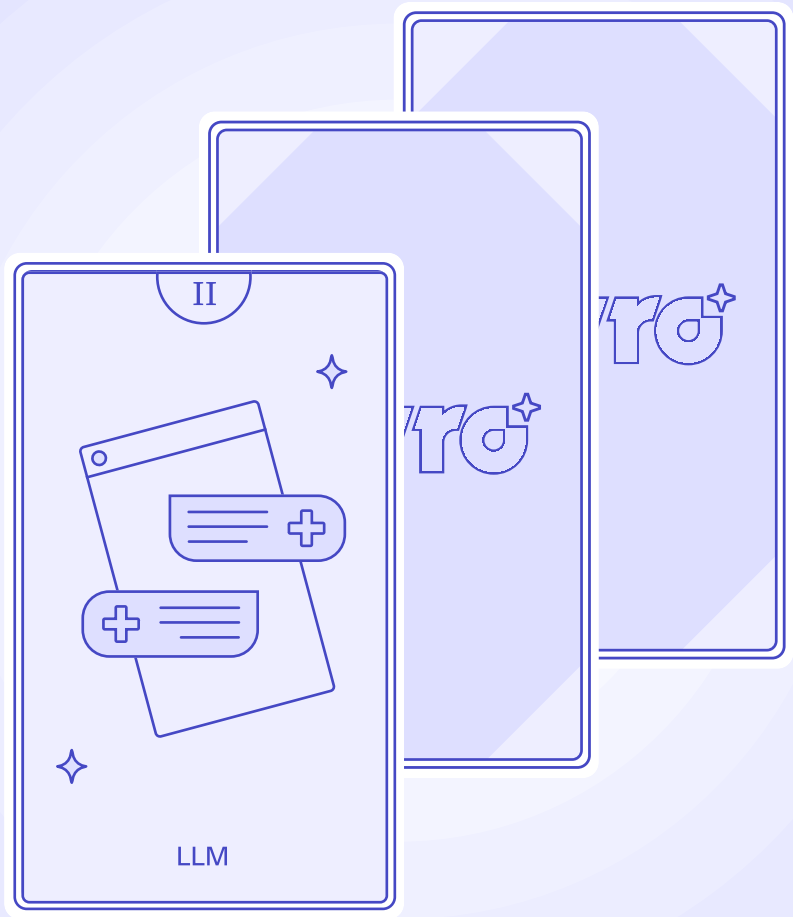
Other points of patient contact, such as mobile apps, SMS messaging, and websites, also tend to fall short because of limitations in their ability to reliably provide accurate and actionable responses to patient questions and requests.

Today's patients demand convenience, easy access to care, frictionless scheduling and prescription refills, quick answers to their questions, and other timely and appropriate responses. Aligning with these expectations and thwarting patient migration to competitors demands that healthcare organizations adopt new technologies. Luckily, game-changing help is available.

74% of call center leaders are under mounting internal pressure to prove their call center's value.

02

The Solution



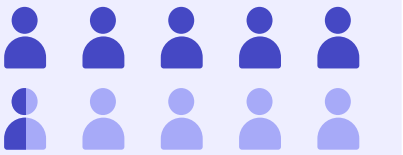
2.1

Generative Artificial Intelligence (With Caveats)

Generative AI in healthcare can be harnessed to alleviate repetitive tasks bogging down health systems. Used correctly and responsibly, it can help minimize workload (and burnout) driven by routine, patient-facing use cases such as answering frequently asked questions, billing, scheduling, prescription refills, preauthorization, password resets, and IT support. Potential administrative use cases for generative AI include analytics, care coordination and continuity, and clinician training.

Health IT leaders face internal and external pressure to adopt AI and demonstrate ROI, but uncertainty holds many back. While 46% of healthcare organizations plan to integrate gen AI solutions within the next 12 months to combat contact center labor shortages, 55% of clinicians have reservations about the safety and reliability of this technology.

Figure 2: Stances on Generative AI Adoption



46% of healthcare organizations plan to integrate generative AI solutions within the next 12 months.

The emergence of large language models (LLMs) should make conversational AI better, more advanced, faster, and easier to train and upskill. However, the rapid, widespread adoption of this relatively unfamiliar technology in healthcare has triggered confusion and apprehension. In the next chapter we'll quickly examine the basics to clear some of this fog.

Patient-facing Use Cases

- FAQ resolution
- Appointment scheduling
- Patient registration
- Prescription refills

Administrative Use Cases

- Care coordination
- Clinician training
- Billing code generation
- Improving clinical documentation

2.2

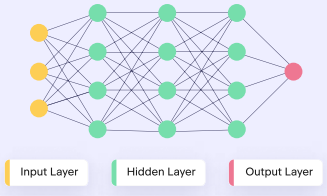
How Large Language Models Work

What is a *Large Language Model* (LLM)?

A **language model** is a computer system programmed to statistically predict the next most appropriate word to fill in a blank space in a sentence or phrase based on the context of the given text. A **large language model (LLM)** is a machine learning model that's trained on vast amounts of text corpus and billions of datasets from publicly available sources.

Language Model
(LM)

A computer system designed to statistically predict the next most appropriate word to fill in a blank.



Large Language Model
(LLM)

A machine learning model that's trained on vast amounts of text corpus and billions of datasets from publicly available sources.

More technically, LLMs use neural networks based on transformers, a type of natural language architecture invented by Google that consists of an encoder and a decoder. (The GPT in ChatGPT stands for Generative Pre-trained Transformer.) What makes these transformers so effective is their ability to scale and process large amounts of diverse datasets.

These LLMs use three steps to predict the next sequence of words:

1. The model breaks the prompt into words, phrases, or even individual characters called "tokens," passing them through its "encoder," which provides each unit with its own numerical representation.
2. The model processes these tokens through a neural processing unit (NPU) for the precise context of words and phrases, identifying patterns and relationships so it constructs a message that makes sense.
3. Finally, the model decodes the numerical tokens in "regular" human language, supplying its predicted next word (such as "Once upon a time there was ..."). The transformer works on auto-repeat, processing and predicting input token after input token until it has either completed its function or exhausted the inserted number of tokens.

The transformer employs natural language understanding (NLU)—a sub-branch of natural language processing (NLP), only far more intuitive and accurate—to arrive at the most likely next word to answer your query and to give you the most relevant response to your input, which is why these models sound so authentic and can hold human-like conversations.

Since the internal neural networks of an LLM are continuously refined by human trainers through a process known as Reinforcement Learning from Human Feedback (RLHF), the model gradually acquires basic grammar, facts about the world, and reasoning abilities. It also becomes ever more cognizant of biases inherent in the data.

What makes *LLMs* so impressive

Versatility

LLMs excel in a wide range of tasks, including copywriting, translation, summarization, brainstorming, code generation, email drafting, and more.

Rapid Results

LLMs generate deliverables that, up until now, could have taken weeks or months to produce in mere seconds.

Accessibility

LLMs have placed never-before-imagined capabilities in the hands of the general public, demanding no prior experience, formal education, or training.

While LLMs exhibit remarkable abilities and have already profoundly transformed how knowledge workers do their jobs globally, there's much to consider regarding their use in healthcare. Successful utilization of LLMs in this sector demands a commitment to reliability, relevance, and pertinence, strict compliance with regulations, and the minimization of latency in the conversational interface to ensure swift responses to queries.

To win over the trust and dollars of health systems, a patient-facing generative AI conversational interface must be:

- Adaptive by nature
- Flexible and contextual
- Fortified with stringent safeguards

2.3

What an Effective Solution Looks Like

To be clear, generative AI is a tool to be used with great care and caution in addressing pain points such as staffing shortages, evolving patient demands, and lengthy wait times. When deployed correctly within defined boundaries, it offers a significant opportunity for hospitals and healthcare systems. Its capacity to facilitate human-like interactions enables the automation of repetitive tasks—freeing up skilled human resources that can make more valuable contributions elsewhere.

A viable solution that makes the most of this opportunity would have to be: ***hybrid, responsible, easy to implement, and transformative.***

HYBRID

A hybrid solution securely combines the conversational user interface and experience we've come to expect from products such as ChatGPT with internal, vetted health system data that can be controlled and closely monitored.

To strike the perfect balance for a cutting-edge digital patient experience while safeguarding personally identifiable information (PII), this solution must be clear of:

Hallucinations
Instances in which a conversational interface incorrectly (but convincingly) answers a question with a fabricated response.

Toxicity
The generation of disrespectful, offensive, or harmful content.

Fragility
The likelihood that a seemingly insignificant change to a prompt will completely change an output.

Irrelevance
Responses unrelated to the task or question at hand.

Privacy violations
PII breaches, under HIPAA law, can levy fines of up to $50,000 per violation.

RESPONSIBLE

Responsibility in the context of generative AI means safeguarding your AI solution by implementing the following measures:

Control

Control entails restricting data inputs to trusted sources such as an organization's FAQ pages, CSV files, medical databases, and other trustworthy internal sources.

Compliance

Compliance with applicable laws and regulations, including HIPAA, to protect confidential and sensitive patient information.

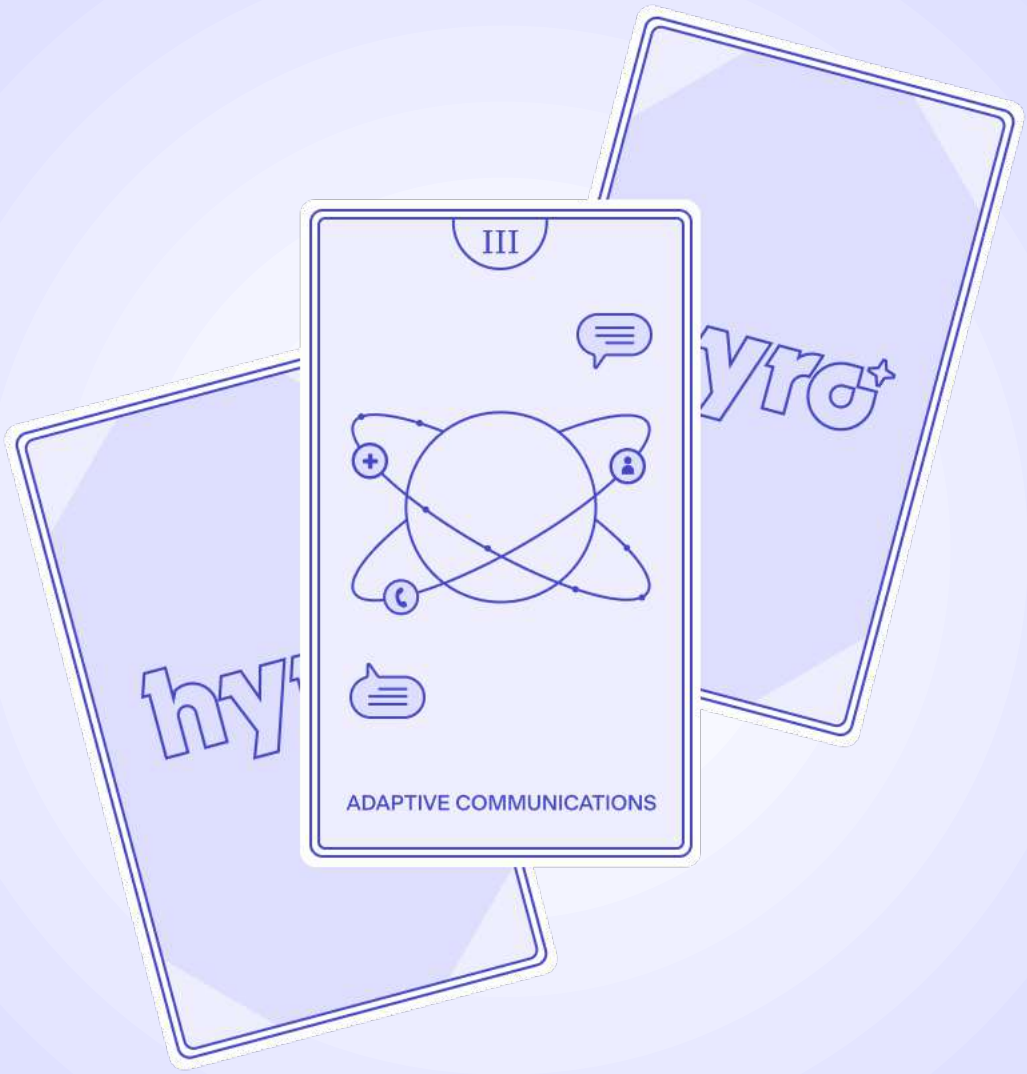
Explainability

Explainability provides system decision-makers with complete visibility into the solution's logic pathways from input to output and allows them to investigate the root cause of any inaccurate or misleading utterance.

This solution should **easily integrate** into your workflows, as well as your contact center, telephony systems, mobile app, and website, ensuring seamless adoption. It should be **transformative** in that it should immediately and significantly reduce the load on your staff, freeing them to focus on higher-value interactions and functions and demonstrate tangible ROI.

03

How Hyro Integrates LLMs Into Its Adaptive Communications Platform



3.1

Hyro's Adaptive Communications Platform

Traditional chatbots and voice solutions (e.g., IVR) are intent-based. Intent, within the conversational AI space, means rigid strings of "if X then Y." For example, if a user says "green," then the chatbot responds "go."

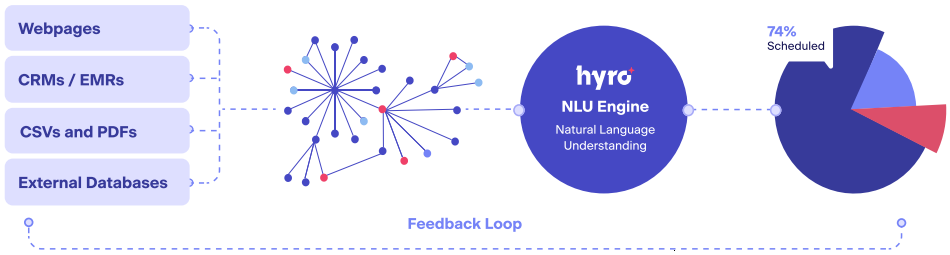
Firmly believing that human-computer interaction can and should be more flexible and less limited than these bounded charts of intent, Hyro developed a new language-based approach to conversational AI called Adaptive Communications.

Adaptive Communications in a Nutshell

Hyro incorporates knowledge graphs, computational linguistics, and natural language understanding (NLU) to create plug-and-play voice and chat AI assistants that easily scale to meet the demands of health systems and their patients.

Knowledge graphs serve as maps built upon an organization's pre-existing data sources, ranging from APIs and PDFs to CSVs. Hyro enhances these maps by incorporating layers of natural language understanding, enabling AI assistants to comprehend and analyze various words within a sentence based on their composition and grammar. As an enterprise modifies information across its internal data sources, Hyro's platform automatically updates its knowledge graph to ensure the information delivered to end users stays precise and current.

An inherent issue with intent-based tools is their reliance on machine learning, which necessitates significant resources and time for training and fine-tuning to achieve accuracy. In contrast, Hyro's Adaptive Communications Platform provides a more cost-effective alternative with faster deployment, no maintenance requirements, and ease of scalability.



01
Scrapes Data
Hyro scrapes information automatically from pre-approved internal sources such as webpages, databases, EMRs, APIs, and PDFs.

02
Maps Data to Knowledge Graph
Hyro translates the scraped information into a knowledge graph, which self-updates as the data changes.

03
Makes it Conversational
Data is made queryable by natural language on all channels.

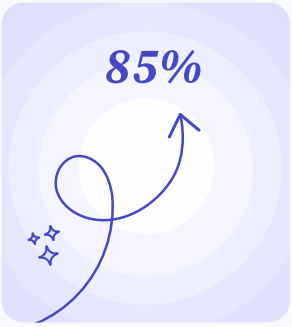
04
Extracts Key Insights
Decision-makers can access real-time patient journey analytics.

Hyro's platform has gained widespread adoption among leading healthcare enterprises in the US, including Intermountain Health, Novant Health, Weill Cornell Medicine, and Hackensack Meridian Health. Hyro is a Gartner Cool Vendor in Conversational AI and Natural Language Technology and is recommended by the veteran consulting firm as a go-to solution for organizations aiming to meet patient digital demands without substantial business resource allocation.

Bottom Line

Hyro's Adaptive Communications Platform can:

Launch in days, not months



Automate over 85% of repetitive tasks

Extract rich conversational insights

Hyro's Platform Results

-99%
In call center average hold time

+47%
In appointments scheduled online

5.5x
Average of ROI within six months

+35%
Average savings in annual operational costs

3.2

Hyro's GPT-Powered Assistant, Spot

As per Hyro's State of Healthcare Call Centers Report, health system executives' top concerns about the use of LLM-powered solutions are HIPAA compliance (64%), accuracy and reliability (38%), and lack of data control (35%). Spot, Hyro's GPT-powered assistant, differs from currently available generative AI solutions through a unique composition that renders it 100% HIPAA-compliant, accurate, and controllable. Let's unpack what that means:

01

Complete Data Control

- Spot pulls its answers directly from HIPAA-compliant internal organizational data. No external or unauthorized sources are ever used to form its outputs.
- It comes with embedded proprietary mechanisms to prevent 'hallucinations' & off-target responses.
- End-to-end closed API guarantees 100% data protection, concealing patient and health system information from unwanted parties.

02

Visibility and Explainability

- Users can monitor, track, and review the exact information sources Spot pulls from to compose its responses.
- Gain visibility into patient requests and top requested services.
- Identify system knowledge gaps and sanitize existing data.
- Monitor engagement and patient goal completion rates.

03

Enhanced Patient Experience

- Spot automatically channels the most relevant content on your website through GPT to generate instant, human-centric answers.
- Steers patients to the services and information they need by navigating them to relevant pages for deeper exploration.
- Holds intuitive multi-turn conversations, retaining context throughout the interaction and able to handle complex queries with diverse dialects, syntax, synonyms, and slang.

How Spot works is intuitive and simple.

Accurate responses generated in milliseconds

The patient asks for information

↓

Spot searches only relevant web pages

↓

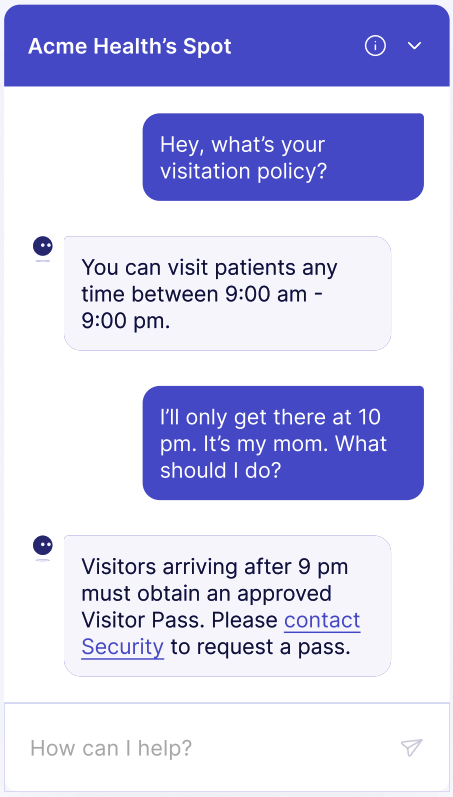
It weighs the accuracy of the answer against its Knowledge Graph

↓

This mapped knowledge base constrains the algorithm's response, keeping it relevant and valid

↓

The model generates a single cohesive response that provides the patient with a clear answer



Does Spot deliver?

Based on a sample of 26,000 patient conversations, as analyzed by Hyro customer Summa Health, Spot offers:

98%

Accuracy rate

Questions that were answered with the correct response or service

73%

Resolution rate

Questions and requests that were resolved end-to-end by Spot

In short, Spot combines the conversational prowess of large language models (LLMs) with HIPAA-compliant and PII-secure data scraping capabilities. It enables healthcare systems to answer large volumes of open patient inquiries—as opposed to the traditional and outdated search box—without having to train the model. This platform represents a new frontier in digital patient-provider interaction.

The Three Pillars of *Responsible Generative AI* in Healthcare

Hyro has established these three core pillars for **responsible** generative AI communications in healthcare

Control

All data used to form Hyro's generative AI outputs is comprised exclusively of internal and approved health system sources.

Compliance

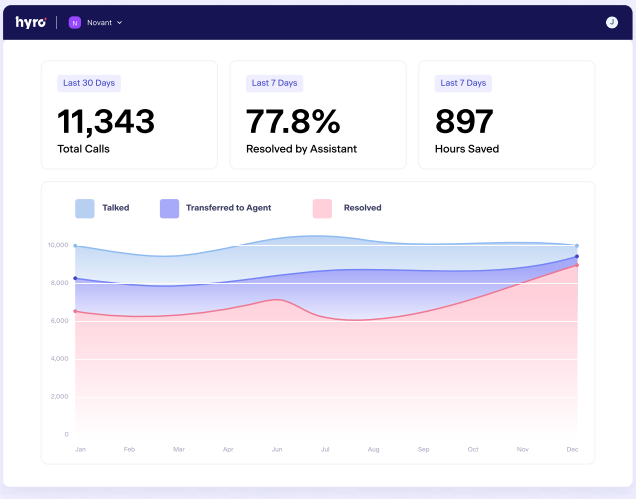
Knowledge graphs act as safeguards to redact patients' personally identifiable information (PII) and meet HIPAA security requirements, preventing data transfer and breaches.

Explainability

By integrating GPT and cross-referencing responses with knowledge graphs, Hyro can present the precise reasoning behind all outputs as well as the data sources used to generate them.

ANALYTICS

Hyro enables healthcare systems to extract actionable insights and analytics that the "black box" of machine learning can't access. This valuable information is automatically collected in real-time, easily shareable across the organization, and aids in enhancing operational efficiencies, data-driven decision-making, and addressing customer pain points.



Hyro, the leading Adaptive Communications Platform for healthcare, enables health systems to automate workflows and conversations across their most valuable platforms, services and channels—including call centers, websites, SMS, mobile apps and more. Hyro's plug-and-play approach helps organizations recapture time and investment lost to building and maintaining chat and voice solutions.

Hyro's clients, which include Intermountain Health, Baptist Health, and Novant Health, benefit from AI assistants that are 60x faster to deploy, easy to maintain, and simple to scale—generating better conversations, more patient goals completed, and revenue-driving insights.

Learn More →

Visit us at: www.hyro.ai