



SEMALYTIX

WHITE PAPER

Revolutionizing the development of patient journeys with large language models

February 2024

By Prof. Dr. Philipp Cimiano and Janik Jaskolski

Semalytix's mission is to improve patient-centricity, inform patient engagement, and provide insights for developing more patient-focused drugs and therapies.

Semalytix is a Germany-based AI startup that has developed PatientGPT™ and Pharos™.

Both are AI-powered patient-centricity solutions. They amplify the global patient voice, making it easier and more scalable to identify unmet patient needs and to gain a deeper understanding of how people genuinely experience diseases.

Semalytix's expertise in artificial intelligence, natural language processing, and Large-Language Models enables them to comprehensively anonymise and accurately analyze vast amounts of real-world patient experience data in over 25 languages. This unique approach positions Semalytix at the forefront of aiding pharmaceutical companies in designing patient-centric strategies, improving patient engagement and advancing patient-focused drug development.

The company is focused on establishing new, scalable, and reliable solutions that improve patient understanding for a simple reason, patient-centricity will only be achieved and maintained over time if complexity, cost, and the time needed to create insights, identify unmet needs, and even craft entire, patient experience data-based patient journeys are dramatically reduced.

Revolutionizing the development of patient journeys with large language models

Executive Summary

This white paper explores how large language models (LLMs) can revolutionize the capture of patients' journeys through disease, diagnosis and treatment. It explains how LLMs and other generative AI-based technologies are improving the speed and accuracy with which patient journeys are mapped and understood, which in turn supports the development of better care solutions.

The 'patient journey' refers to the path taken by an individual through their condition, from the first symptoms, through diagnosis, treatment(s), and eventually to optimal disease management or cure. It captures patients' experiences, interactions with healthcare stakeholders (doctors, surgeons, pharmacists) and key events along this path.

By offering a patient-centric view of disease and treatment, patient journeys are a powerful tool for healthcare stakeholders to better understand patients' needs, preferences and concerns, the drivers of their decision-making and the kind of information they seek. This ultimately helps identify more effective patient support mechanisms and better treatments, as well as supporting more targeted/effective marketing and commercialization efforts.

Yet patient journeys are challenging to capture in a timely, accurate and cost-effective manner. Creating a coherent patient journey requires careful screening and integration of multiple data types.

These include literature, claims and patient interview data, electronic health records and health care provider information. This data must also be constantly updated: patients' experiences evolve as their journey progresses, but also as new treatments, tools and procedures become available. There is also the challenge of ensuring that patient journeys go beyond objective milestones to truly capture and reflect patients' subjective experiences, their personal decision making, information needs, emotions and "moments-of-truth".

LLMs and other generative AI tools can help address these challenges. LLMs can i) facilitate and accelerate data capture and integration, ii) ensure patient journeys are dynamic, up-to-date, and accurately reflect patients' experience and voice at any given time, and iii) generate specific information on particular decisions taken along the patient journey.

Patient Journeys

Patient Journeys capture and visualize patients' path through their health condition, from early symptom onset, to seeking professional help, achieving a diagnosis and receiving treatment. Most patient journeys are complex. They comprise multiple decisions around seeking care and treatment choice, numerous touchpoints with the healthcare system (doctors, hospitals, pharmacists), and the physical and emotional impact of a given condition on patients' lives, family and social environment.

Patient journeys are typically displayed as a sequence of phases to help users identify and understand key steps in the patient journey, and to organize the main stages of a given condition.

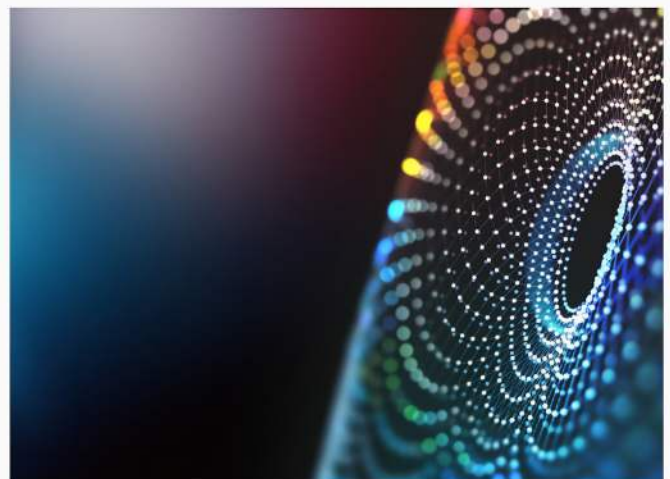
By highlighting patients' key experiences, decision points and encounters with healthcare professionals, patient journeys make explicit the main aspects of patient-doctor communication and information exchange. They also underscore different stakeholders' roles along patient journeys, and the kinds of information that patients consume during that journey.

No two patient journeys are the same. They may also be quite specific to certain conditions: patients' experience of a chronic disease such as irritable bowel syndrome or diabetes will look very different from that of an acute infection, for instance. Nevertheless, all patient journeys typically include the following phases:

1. Awareness and recognition of symptoms
2. Seeking help
3. Presentation and diagnosis
4. Treatment selection
5. Brand selection and access
6. Switching or maintaining treatment
7. Lifestyle changes

Classical patient journeys focus on such pre-defined phases, including key diagnostic or treatment milestones and outcomes. However, there is additional value in capturing how patients feel, how they inform themselves, how they make decisions, and the factors that contribute to those decisions. For example, which symptoms lead patients to first consult a doctor? What are their biggest fears and concerns? What is most important to them?

Understanding patients' journeys in the fullest sense – as defined not simply by events and health system interactions, but also by patients' experiences, feelings and concerns – helps sponsors, health care providers and healthcare professionals more effectively address patients' needs. It also enables better communication, greater patient engagement, and more focused product development opportunities. Sponsors and healthcare providers with the best understanding of patients' needs will compete most successfully with fit-for-purpose products and support.



The Challenge

There is no one-size-fits-all approach to capturing patient journeys. It is an iterative process, ultimately leading to an increasingly detailed understanding of patients' experiences. The usefulness of any given patient journey is determined by the answers it provides to specific questions.

The most impactful patient journeys typically include multiple data sources and include patients' own voices, as captured for instance via social media or interview transcripts.

The most comprehensive patient journeys may include:

- Publicly available literature, including conceptual disease models and disease natural history data
- EHR data (patients' clinical history)
- Claims data (tracking treatments, diagnostic tests, hospital admissions)
- Patient registry data
- Patient interviews and case studies
- Health care provider interviews reflecting how healthcare professionals perceive patients' journeys

- Primary market research including survey data
- Social media, comprising patients' unfiltered voices

Integrating these multiple data sources and types into a coherent patient journey is challenging. It requires mastering data integration, and a deep understanding of data privacy laws and usage rights. Data collection and interpretation must be repeated regularly to capture evolving patient needs and standards-of-care. What's more, few of the listed data sources will directly reveal patients' motivations, emotions and decision-drivers.

Social media arguably offers the deepest insight into patients' experiences. Yet integrating and interpreting unstructured, large-scale data sources like social media requires natural language processing tools.

Enter Large Language Models. These new technologies are game-changing. They allow faster, more detailed capture of patient voices based on powerful machine-based training processes and predictive capabilities.



The Technological Opportunity: Large Language Models

Language models have long been used for machine-based translation, spell checks, text prediction and speech recognition. Their power has grown radically with the arrival of ChatGPT and other new models based on the so-called “Transformer Architecture”. These models have billions of parameters and are pre-trained on large amounts of text data (hundreds of terabytes of data corresponding to billions of text documents). GPT3, the training model underlying ChatGPT, features 175 billion parameters.

This pre-training makes them extremely strong text predictors, able to take into account contextual cues to generate highly plausible responses to queries. Notable large language models include those by OpenAI (ChatGPT, powered by GPT3.5 and its successor GPT4), META (LLaMa), Stanford University (Alpaca) or Anthropic (Claude).

Despite different data architectures, the basics of LLMs remain essentially the same across vendors /

providers. Pre-trained on large amounts of text, these models are typically fine-tuned on instructions or dialogues to ensure that their response fits what is expected in a given context. This model “alignment” is carried out using reinforcement learning.

The behaviour of LLMs is controlled by “prompting” - providing specific input instructions to get contextually appropriate answers that best reflect the underlying data. Using a range of different prompts helps achieve the most accurate answers.

LLMs excel at summarizing large amounts of data, answering open questions, engaging in dialogue (chatbots), and performing sentiment analysis. They can be applied to open-ended tasks and questions without having to be explicitly programmed for answering these specific questions. They can generate content based on relatively vague and natural human instructions.

Deriving patient journeys from self-reported patient experience data using LLMs

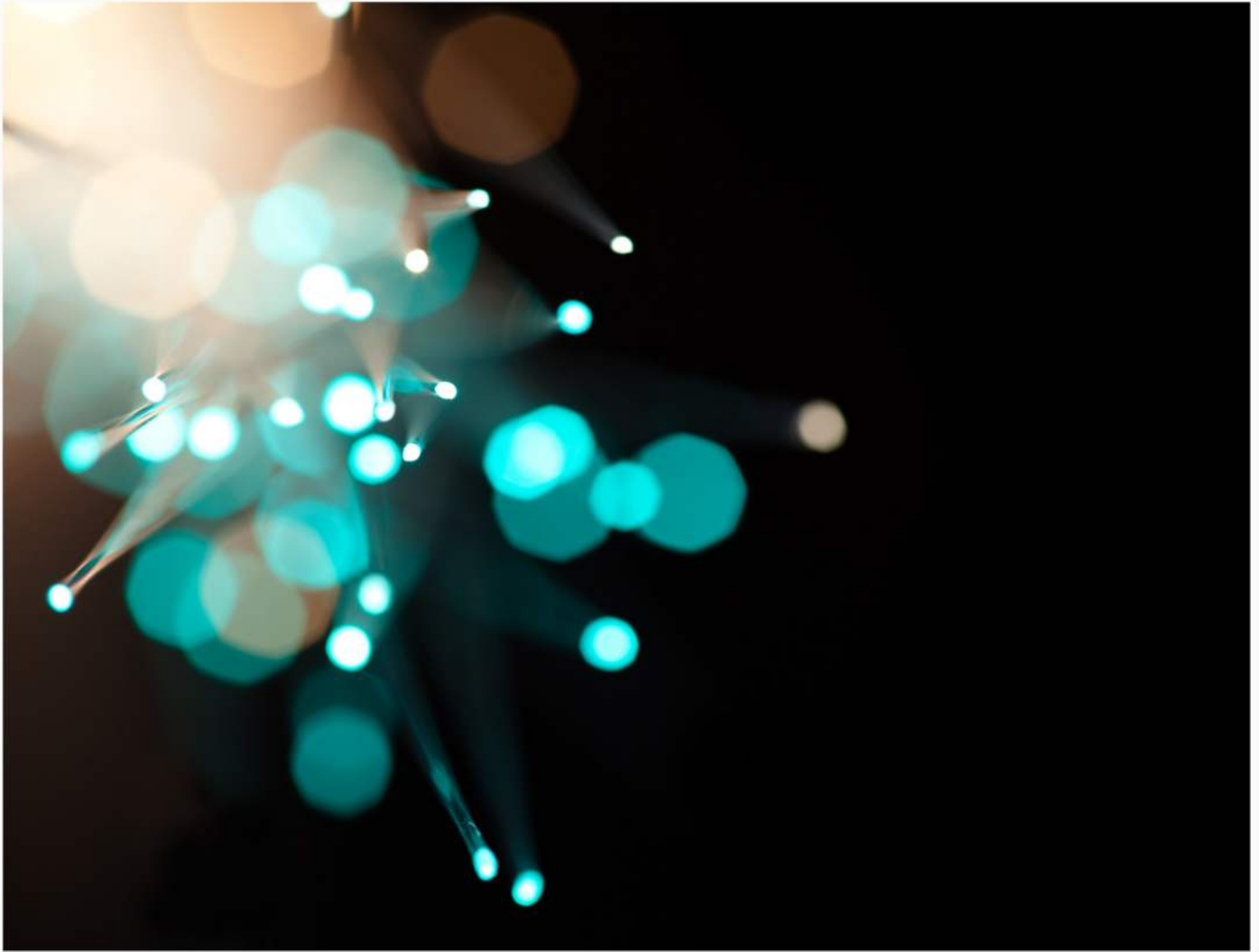
LLMs therefore offer a huge opportunity to transform how patient insights are captured. Fueled by huge datasets of patients’ self-reported experiences of their condition and treatments, LLMs can reproduce patient voices at unprecedented scale and speed. This in turn can radically improve the quality and usefulness of patient journeys.

Semalytix has developed a new methodology to automate patient journey mapping. This mapping reflects the patient experience as reported in social media and other patient-reported data sources. The approach relies on a huge data lake of patient experiences across dozens of indications, including millions of social media posts from patients with clear, self-disclosed diagnoses of a given disease. This data lake, combined with LLMs’ generative capacity, can be used to answer almost any question about the patient journey.

Using this methodology, we can generate a default

patient journey by prompting a LLM to answer standard questions relevant to any condition:

- What are early symptoms that lead a patient to seek a health care professional?
- What tests or procedures are applied to arrive at a diagnosis?
- What is the emotional impact of the diagnosis on the patient?
- Which sources do patients consult for more information about the disease or possible treatments?
- How does the diagnosis affect patients’ social and family environment?
- What are the factors that patients consider when opting for a treatment?
- What positive and negative experiences are associated with given treatments?
- What side effects do patients have to cope with? How do they do that? Which side effects are intolerable?



This “first draft” patient journey can be further refined using more specific questions. For example, users can ask questions based on their own experience, or to help fill knowledge gaps around specific treatments or side effects.

The methodology has multiple benefits:

- Rapidly generated patient-journeys with no/low need for costly data integration
- Focused on subjective experience and what patients disclose on social media
- Up-to-date, as answers to key questions can be generated at regular intervals
- Can be used to analyze trends over time
- Provides answers to specific questions around specific decisions
- No data science expertise is needed

Notwithstanding the many strengths of this

methodology, it cannot on its own capture the full truth: LLM-generated journeys should be complemented with data from other sources such as those mentioned above to obtain a complete picture.

LLMs’ strengths in generating text that goes beyond their training data can also lead them to „hallucinate“ – generate content that is not backed by the input data. (The models „compress“ their training data, but do not remember the text explicitly.)

To address the problem of hallucinations, Semalytix has developed a method to provide relevant content from the patient data lake as part of the prompt. This ensures that the LLM relies only on relevant data when generating summaries, rather than „inferring“ information from the compressed data representation.

Use case: Deriving a Patient Journey for Breast Cancer Patients

The methodology described above is part of PatientGPT, an LLM-powered product by Semalytix that supports the fast, agile and comprehensive generation of patient insights from social media. Patient journey generation is one module of PatientGPT.

To prove PatientGPT's ability to generate patient journeys that truly reflect the patient experience, we compared its performance with that of an in-person study of breast cancer patients in Spain, carried out by Ciria-Suarez et al..

Those researchers used interviews to map the journeys of 21 Spanish breast cancer patients , yielding insights for seven key phases: Diagnosis, Surgery, Chemotherapy, Radiotherapy, Follow-up Care, Relapse, Metastatic Breast Cancer.

Many of the subjective experiences described by PatientGPT corresponded to the experiences summarized in the published journeys, demonstrating that the LLM-powered methodology is able to accurately summarize patients' experiences.

More information on this and other case studies is available on request.

¹Ciria-Suarez L, Jiménez-Fonseca P, Palacín-Lois M, Antoñanzas-Basa M, Fernández-Montes A, Manzano-Fernández A, Castelo B, Asensio-Martínez E, Hernando-Polo S, Calderon C. Breast cancer patient experiences through a journey map: A qualitative study. PLoS One. 2021 Sep 22;16(9):e0257680. doi: 10.1371/journal.pone.0257680. PMID: 34550996; PMCID: PMC8457460.



About the Authors

Prof. Dr. Philipp Cimiano is the Chief Technology Officer and Co-Founder of Semalytix. He is an internationally leading academic in the fields of semantic technologies, knowledge representation and natural language processing.

Janik Jaskolski is the Chief Product Officer and Co-Founder of Semalytix. He is an innovative leader in AI life-science technology and has been instrumental in advancing patient-focused drug development through the development of AI/NLP technology.

For Further Information

If you would like to discuss this report, please contact the authors.

