

“Some AI companies stand out via outstanding academic validation; some via successful customers and deployments; and yet others by using AI for good. John Snow Labs is utterly unique in doing all three.”

### Making Healthcare AI Perform in the Real World

Healthcare is a hard industry to innovate in. Conflicting interests, siloed systems, messy data, heightened compliance, security concerns, skeptic customers, and long validation cycles all contribute. Yet, nothing is more important than improving our collective health and well-being. It therefore gives us great pleasure to name John Snow Labs as the fastest growing AI company of 2022, as living proof that combining business success and better healthcare can go hand in hand.

John Snow Labs offers state-of-the-art clinical and biomedical Natural Language Processing (NLP) models are used to predict and optimize medical outcomes. Some of these use cases include accelerating clinical trials, predicting disease progression, providing clinical decision support, analyzing real-world evidence, detecting and preventing adverse drug events, and creating a more comprehensive view of each patient’s journey.

The data science teams of several large healthcare systems have put the software to action – and then publicly described their success in industry conferences. Kaiser Permanente applied it to optimize hospital patient flow models: the solution enabled real-time decision-making and strategic planning, by predicting bed demand, safe staffing levels, and hospital gridlock. Mount Sinai uses it to predict the aggression level of psychiatric patients based on a combination of free-text clinical notes and structured data. Providence Health has applied the software to de-automatically identify – for the first time ever – more than 700 million patient notes.

In the pharmaceutical space, Roche is a longtime customer and uses the software within its Navify platform. John Snow Labs’ software automatically reads and understands pathology reports, radiology reports, and next generation sequencing reports to extract tumor characteristics and other information used to help a cancer patient’s medical team. Novartis has used the software to extract information from hundreds of thousands of clinical trial documents – replacing a process done manually by over 200 people and thus enabling a new drug to get to market several months sooner. Merck has also publicly described several use cases for optimizing the pharma supply chain that are newly enabled by being able to accurately read and understand free-text documents.



John Snow Labs also provides its software to dozens of Health IT companies: Per the latest NLP Industry Survey, 59% of AI teams in the healthcare and life science industries rely on it. This ranges from goliaths like McKesson, Optum, and GE Healthcare to startups like Diameter Health, Omny Health, and Cityblock Health.

### A Peer-Reviewed Technical Foundation

It is rare for any solution in healthcare to apply equally across providers, payers. Pharma, and health IT. The bedrock of John Snow Labs’ technology and community is Spark NLP – the open-source software library for natural language processing. Used by 33% of all enterprise AI teams and 59% of healthcare AI teams, Spark NLP has secured the spot as the most widely used NLP library in the enterprise for every one of the past four years.

Spark NLP has delivered a furious pace of innovation, releasing new software every two weeks for over four years, and growing to provide over 5,000+ pre-trained NLP models available in 375 languages. These enables data scientists who use the library to automate just about any NLP task, from spelling checking to question answering, with a few lines of code.

John Snow Labs’ commercial products - Spark NLP for Healthcare, Spark OCR, and the Annotation Lab – all build on Spark NLP and hence enjoy the speed and scalability benefits it provides. Most importantly, they deliver state-of-the-art accuracy: Peer-reviewed, reproducible, public papers and code that show the library providing the best accuracy ever shown on 20 academic benchmarks in the clinical and biomedical space.

This new level of accuracy, recently enabled by research advances in deep learning and transfer learning, enable Spark NLP for Healthcare to serve as the connective tissue for accurately understanding unstructured medical information. This in turn enables data scientists or domain experts – as clinicians – to make better decisions, whether it be a form of the recommended treatment or how to appropriately staff an emergency room.

### Riding the Trends of AI for Healthcare

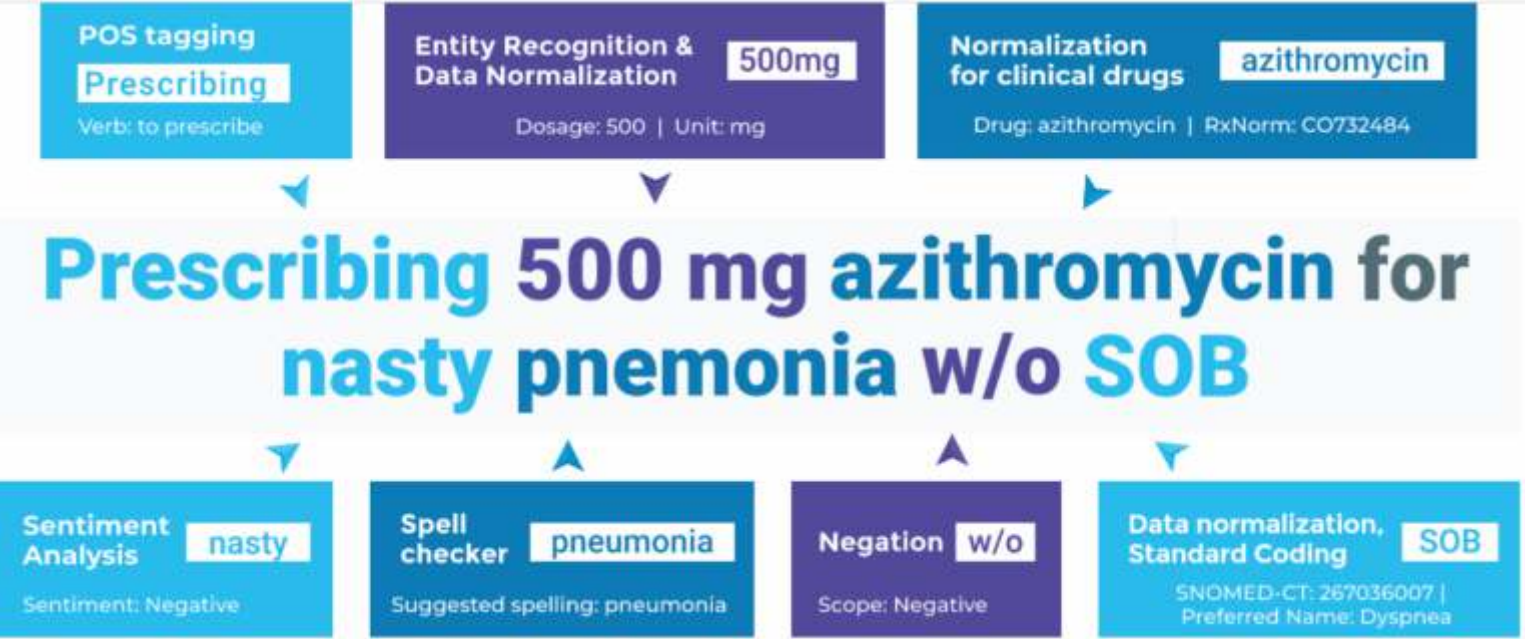
According to the latest industry survey from Gradient Flow, there are several trends shaping AI in healthcare. First, practitioners are shifting from exclusively data scientists to domain experts. In fact, more than half (61%) of respondents identified clinicians as their target users, followed by healthcare providers (45%) and health IT companies (38%).

This is good news: AI is becoming more widely used, and the barriers to entry are getting lower for those outsiders of technical roles. It also shows that practitioners understand the unique challenges of the healthcare industry, such as clinical language and jargon, stricter protocols around information sharing, and the critical need for accuracy. John Snow Labs’ Annotation Lab is a no-code tool used today by medical doctors and bioinformaticists to train and tune AI models without getting data scientists involved. Now in its third year and its 33rd release, the tool offers a robust toolset and support for large teams, right when the industry is ready for it.

Second, the survey also shows a growing interest in open-source solutions and healthcare-specific models, respectively. Customers are rightfully weary of buzzwords and want solutions they can test themselves, which have been tailored to the unique challenges of healthcare data. As such, it’s not surprising that the survey found that a majority of respondents (53%) choose to rely on their own data to validate models, rather than on third-party or software vendor metrics. Respondents from mature organizations (68%) had a clear preference for using in-house evaluation and for tuning their models themselves.

As the developer of a major open-source library, an extensive healthcare-specific NLP library, and a large collection of pretrained healthcare-specific models, John Snow Labs meets all the demands of today’s AI users. The team building the software includes not only data scientists and engineers but also medical doctors and biomedical domain experts, ensuring





that models are clinically valid and work on real medical cases and documents.

Third, concerns around data privacy and cybersecurity are top of mind for all healthcare software buyers. The survey shows a decrease in the use of cloud-based services compared to the previous year, and confirmed the absolute need for compliance. Designed for this industry from the ground up, Spark NLP is designed to run in air-gap environments, requiring no data sharing with John Snow Labs or any third party. It also runs on the customer's infrastructure, making it easy to comply with data residency regulations and mesh with each organization's security architecture and privacy controls. This is important, as threats escalate and personally identifiable information (PII) is on the line.

Growing Faster by Doing Good

John Snow Labs also impressed on living up to its missing of using AI for good. This starts with its focus on healthcare, its massive open-source contributions, and its continued investment in providing the gloal AI community and free and open-source software and models. The company also provides its commercial software under a free license to academic researchers and teachers, an offer that's already been taken by more than 150 universities worldwide.

But it does not end there. John Snow Labs is also contributing to the education of the AI community, as the creator and host of the NLP Summit and Healthcare NLP Summit. These free conferences, the last of which attracted more than 10,000 online attendees, have become the largest event for the applied NLP community, gathering speakers and best practices from across all major open-source projects, industries, and use cases.

John Snow Labs has also become a carbon neutral company – minimizing and then fully offsetting its entire carbon footprint for the entire operating history of the company. This was done with the express goal of showing that there is no inherent tradeoff between training state-of-the-art NLP models and wasting substantial amount of electricity or compute power.

Most impressively, John Snow Labs has achieved its growth and community contributions without ever having raised external investment or debt. This was a strategic early decision – enabling complete focus on customer needs, in contrast to investor needs. The way the team tells it, fast growth came not despite the lack of investment – but because of it.

# Spark NLP for Healthcare

Peer-reviewed state-of-the-art accuracy

### 4-6X Fewer Errors

Than AWS, Azure, or GCP

### #1 Accuracy

on 20 benchmarks in peer-reviewed papers

### 5 of 8

of the world's largest pharma are customers

Entity Recognition & Data Normalization  
**500 mg**  
Dosage: 500 | Unit: mg

Sentiment Analysis  
**nasty**  
Sentiment: Negative

Data normalization, Standard Coding  
**SOB**  
SNOMED-CT: 267036007 | Preferred Name: Dyspnea

## Prescribing 500 mg azithromycin for nasty pneumonia w/o SOB.

POS tagging  
**Prescribing**  
Verb: to prescribe

Normalization for clinical drugs  
**azithromycin**  
Drug: azithromycin | RxNorm: CO732484

Spell checker  
**pneumonia**  
Suggested spelling: pneumonia

Negation  
**w/o**  
Scope: Negative

### Which NLP Libraries does your organization use? (check all that apply)

(share of respondents in the Healthcare sector)

Spark NLP	54%
spaCy	25%
Hugging Face	24%
nlk	22%
Stanford CoreNLP	16%
Other	16%
Gensim	11%
Allen NLP	8%
Rasa NLU	6%

## The most widely used NLP library far in Healthcare, by far

Source: gradientflow.com

## 600+

pretrained clinical pipelines & models out of the box

## 100+

clinical named entity recognition(NER) models

## 400+

different entities from various taxonomies

johnsnowlabs.com

info@johnsnowlabs.com

+1 (302) 768-5227