



Applied Generative AI for Data Scientists

April 2025 Training & Certification

Break Until 1:05pm ET

Training Workshop Outline

Medical LLMs

Medical Chatbot

Human in the loop:
Generative AI Lab

LLM & RAG
with Spark NLP

Multi-Modal
Language Models

Information
Extraction

John Snow Labs

Provides State-of-the-Art Medical Language Models

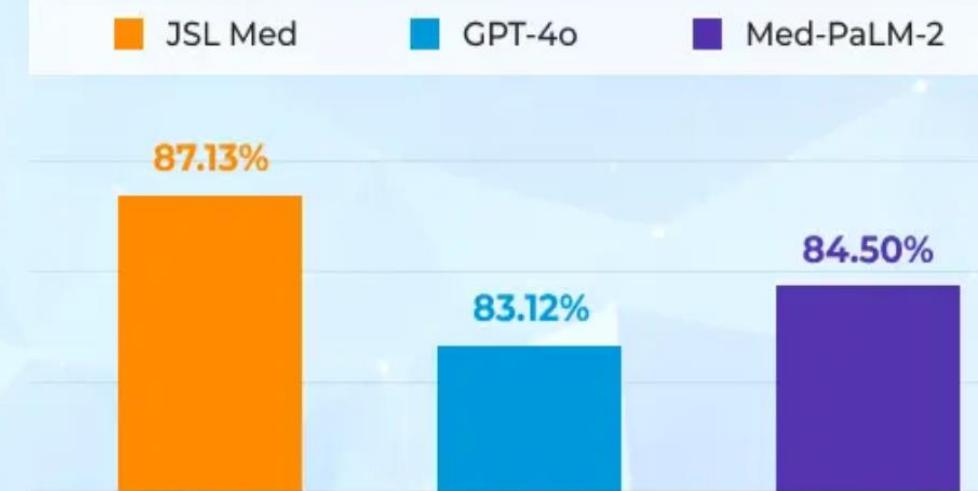


Use a Medical LLM that outperforms GPT-4o on tasks like clinical assessment (+6%), research comprehension (+4.2%), and clinical case analysis (+9%)

In a blind evaluation by practicing medical doctors, the Small Medical LLM was preferred over GPT-4o **45% to 92% more often** on clinical text summarization and information extraction

Designed for privacy and compliance, the Medical LLM runs inside your cloud

Average Performance on OpenMed Benchmarks



Designed for Enterprise Healthcare



Keep up with the state of the art

Bi-weekly releases for 8+ years



Run Privately in your infrastructure

Data never leaves your perimeter



Scale on a budget

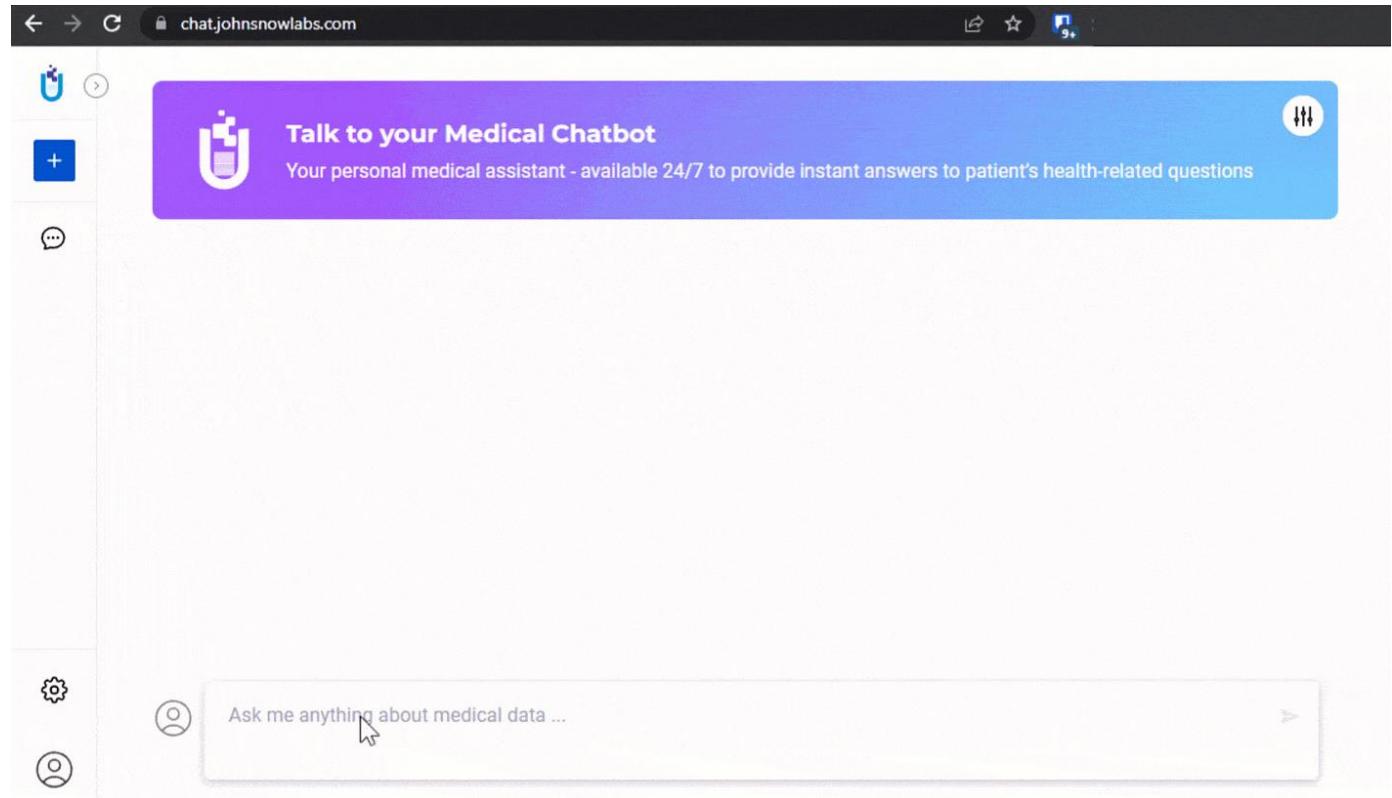
Priced per server, not per token



John Snow Labs' Medical Chatbot



- ✓ Delivers Superior Accuracy for Healthcare-Specific Tasks
- ✓ Runs privately behind your firewall:
No third-party APIs or data sharing
- ✓ Updates all medical knowledge daily
- ✓ Can consistently reproduce results
- ✓ Does not hallucinate, cites its sources
- ✓ Explains its answers
- ✓ Adds private knowledge bases
- ✓ Tunable guardrails and brand voice



John Snow Labs' Medical Chatbot



- ✓ Delivers Superior Accuracy for Healthcare-Specific Tasks → **Multi-LLM, Multi-Agent Architecture**
- ✓ Runs privately behind your firewall:
No third-party APIs or data sharing → **Comes with Medical Knowledge Bases**
- ✓ Updates all medical knowledge daily → **Built-in RAG Backend**
- ✓ Can consistently reproduce results → **Enterprise Grade Security & Scalability**
- ✓ Does not hallucinate, cites its sources → **Enterprise Grade Security & Scalability**
- ✓ Explains its answers → **Enterprise Grade Security & Scalability**
- ✓ Adds private knowledge bases → **Enterprise Grade Security & Scalability**
- ✓ Tunable guardrails and brand voice → **Enterprise Grade Security & Scalability**

Literature Review Agent



Literature Review

Literature review synthesizes the results of multiple scientific studies to help researchers draw conclusions, identify patterns or to derive a more precise estimate of the overall effect of a treatment or intervention.

1 Search Criteria	2 Data Points	3 Inclusions & Exclusions	4 Results							
Literature Review - Results (98)										
Publication Date (Descending) CSV Print										
ARTICLE	MEDICATIONS	ADE	TESTS	PATIENT AGE	DEMOGRAPHICS	DIAGNOSES	OUTCOMES	INTERVENTION	SYMPTOMS	GENDER
Hypoparathyroidism Revealed by Unsuccessful Anti-epileptic Therapy. Raja Arrab, et al. Case Reports Cureus 2024 · N/A · PubMed · Full Article	oral valproate	N/A	serum phosphocalcic test	13	female	hypoparathyroidism, epilepsy	corrected diagnosis	oral valproate therapy	cerebral calcifications, metabolic disorders, hypocalcemia	female
Epileptic Seizures in a Pediatric Patient With Vein of Galen Aneurysmal Malformation and Obstructive Hydrocephalus: A Rare Case Report. Kiril Ivanov, et al. Case Reports Cureus 2024 · N/A · PubMed · Full Article	valproate	N/A	MRI, MR-angiography, ventriculostomy	6	male	Vein of Galen aneurysmal malformation, Obstructive hydrocephalus, Epilepsy	no further seizures	partial transarterial endovascular embolization, ventriculostomy	severe status epilepticus	male
Valproate-induced hyperammonemia, neuroleptic sensitivity, and cerebellar atrophy-A clinical conundrum in the management of bipolar disorder. Samarth S Shetty, et al. Case Reports Bipolar disorders 2024 · N/A · PubMed · Full Article	valproate, lithium, electroconvulsive therapy (ECT)	catatonia, seizures, hyperammonemia, essential tremors	N/A	N/A	male, male	bipolar disorder, intellectual development disorder	stability, symptom improvement	electroconvulsive therapy (ECT), tailored medication regimen	catatonia, seizures, hyperammonemia, essential tremors	male
Irregular Tremulous Movements and Infrequent Seizures: A Clinical-Electrophysiological Diagnosis of Benign Adult Familial Myoclonus Epilepsy. Kazuki Imon, et al. Case Reports Cureus	valproic acid, perampanel	N/A	video EEG, brain MRI, somatosensory evoked potentials (SEP)	31	male	benign adult familial myoclonus epilepsy	significant reduction of symptoms	valproic acid and perampanel	involuntary finger movements, rare seizures	male

Clinical Guidelines Agent



The screenshot shows the Clinical Guidelines Agent interface. On the left, a sidebar menu includes "John Snow LABS" logo, "+ New Chat" button, "Clinical Guidelines" (selected), "Literature Review", "Recent History", and a "TODAY" section with two collapsed items: "What are the first-line treatment rec..." and "What are the current recommended...". At the bottom are "Settings" and a search bar with placeholder "Ask me anything about medical data ...". The main content area has a purple header "Ask Clinical Guidelines" with a white "U" icon. It displays a patient history: "A 72-year-old male presents with gradually worsening hearing loss over the past 2 years. He reports difficulty understanding conversations, especially in noisy environments, and often asks people to repeat themselves. His family members have also noticed that he frequently turns up the volume on the television. He denies any pain, discharge, or ringing in his ears but mentions occasional dizziness and feeling unsteady when walking. He denies exposure to loud noises or ototoxic medications." Below this is a question: "He has no history of ear infections, ear surgeries, or head trauma. He does not have diabetes or hypertension but mentions occasional forgetfulness and difficulty concentrating. On physical examination, his ear canals are clear with no signs of cerumen impaction or infection. Tympanic membranes appear normal without perforation or fluid. He admits to feeling socially isolated as he finds it increasingly difficult to engage in group conversations." At the bottom of the main content area are two questions: "What is the most likely diagnosis for this patient presenting with gradually worsening hearing loss, difficulty understanding conversations, and no history of ear infections or noise exposure?" and "What diagnostic test should be conducted to confirm the presence and type of hearing loss in this patient?". A small "ANSWER" button is visible at the bottom right of the main content area.

A 72-year-old male presents with gradually worsening hearing loss over the past 2 years. He reports difficulty understanding conversations, especially in noisy environments, and often asks people to repeat themselves. His family members have also noticed that he frequently turns up the volume on the television. He denies any pain, discharge, or ringing in his ears but mentions occasional dizziness and feeling unsteady when walking. He denies exposure to loud noises or ototoxic medications.

He has no history of ear infections, ear surgeries, or head trauma. He does not have diabetes or hypertension but mentions occasional forgetfulness and difficulty concentrating. On physical examination, his ear canals are clear with no signs of cerumen impaction or infection. Tympanic membranes appear normal without perforation or fluid. He admits to feeling socially isolated as he finds it increasingly difficult to engage in group conversations.

What is the most likely diagnosis for this patient presenting with gradually worsening hearing loss, difficulty understanding conversations, and no history of ear infections or noise exposure?

What diagnostic test should be conducted to confirm the presence and type of hearing loss in this patient?

The Generative AI Lab

The No-Code NLP Platform:

- [Annotate Text & Images](#)
- [AI Assisted Annotation](#)
- [Train & Tune Models](#)
- [Models, Rules, and Prompts Hub](#)
- [Manage Projects & Teams](#)
- [Enterprise Security & Privacy](#)



The Generative AI Lab: Building Small, Task-Specific Language Models

ClinicalNER / Tasks / e5.txt

Switch Role
ANNOTATOR

Annotations Versions Progress

Completions +

Predictions

Model (SparkNLP Pre-annotation)
Created 1 minute ago

FILTER BY CONFIDENCE SCORE
only shows predictions above the selected confidence score
0.3 - 1

0 0.5 1

INDICATION:
The patient was a 85 year old Caucasian male, slipped and fell **PROBLEM**. He was diagnosed to have **inter-trochanteric left hip fracture PROBLEM**, severe osteoarthritis **PROBLEM** and **total hip arthroplasty TREATMENT** was performed on 2/2/2018. Sustained another fall post-op in the ward in the shower, complains of **left hip pain PROBLEM**.

TECHNIQUE: AP pelvis **TEST**

COMPARISON: Compared with **pre-op imaging TEST** taken on 1/28/2018, post-operative XR on 2/4/2018.

FINDINGS:
The initial radiographs **TEST** showed an inter-trochanteric left hip fracture **PROBLEM** with the lesser trochanter **PROBLEM** as a separate fragment **PROBLEM**. Bone density is low, soft tissue unremarkable.
Post-operative imaging **TEST** showed **prosthesis in situ TREATMENT**. No **loosening/fracture TREATMENT** seen.

Latest imaging **TEST** showed **traumatic periprosthetic fracture PROBLEM** at the level of the shaft of **the total hip prosthesis TREATMENT**. **Cerclage wire in situ TREATMENT**. No **loosening PROBLEM** or **luxation PROBLEM** seen.
Bone density is severely osteopenic **PROBLEM**.
Soft tissue swelling **PROBLEM** is unremarkable.

IMPRESSION:
Traumatic periprosthetic fracture **PROBLEM** at the level of the shaft of **the left total hip prosthesis TREATMENT**.

View 1800 Characters Per Page

The “Old School” Use Case:

Define the task for a new model

Import documents to Annotate

AI-Assisted Pre-Annotation

Manual work by domain experts

Train or tune a language model

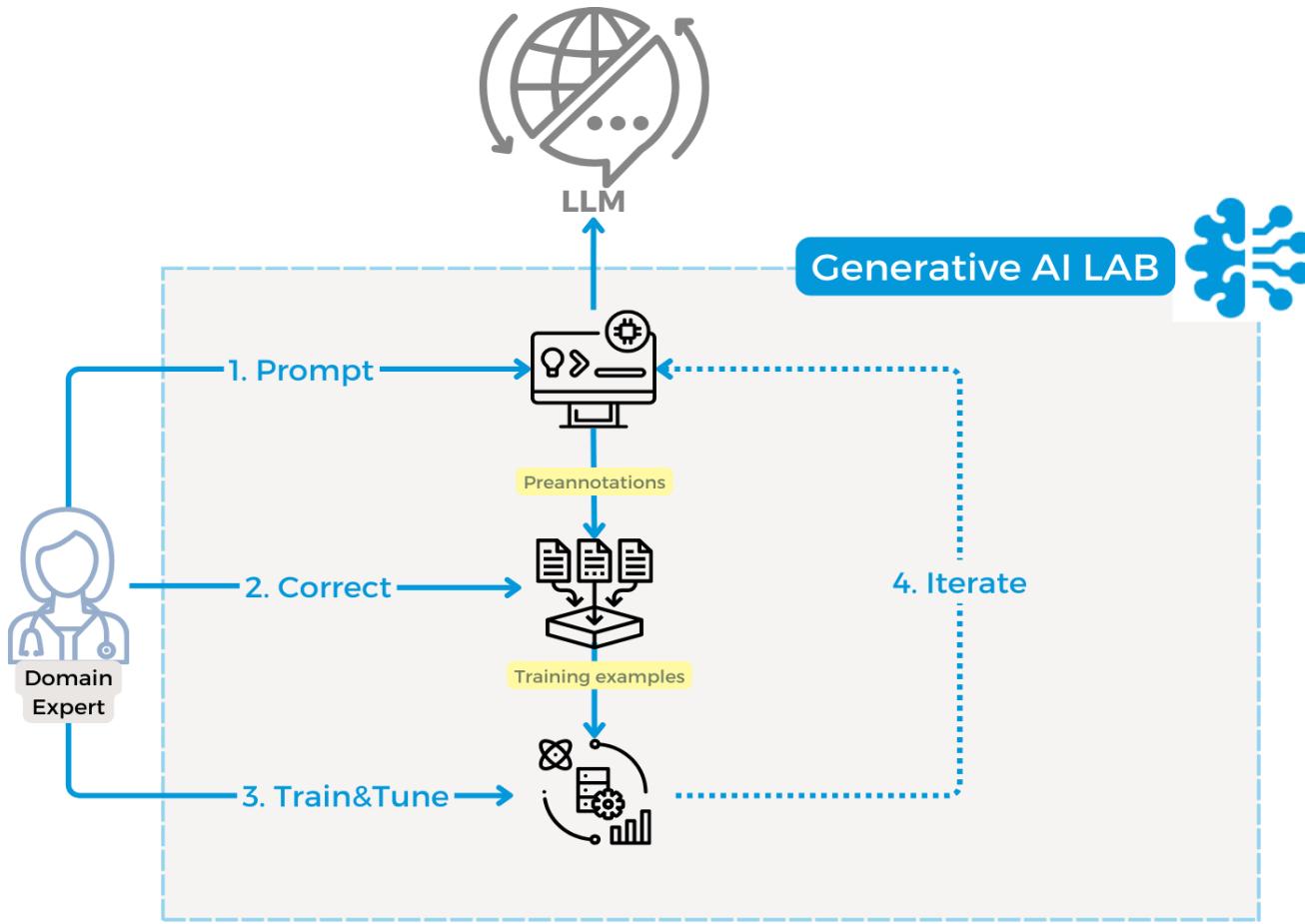
Export the model or the data



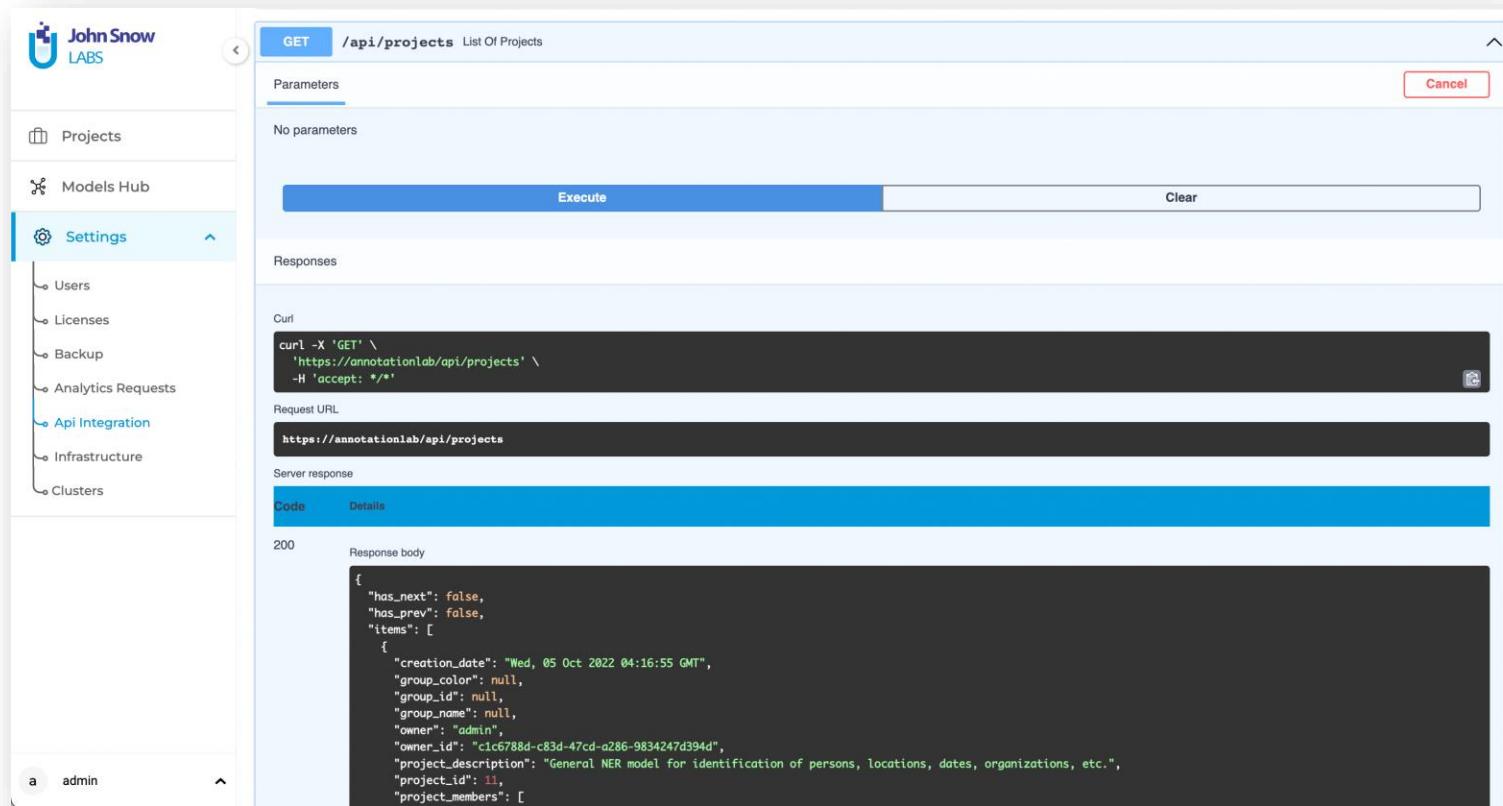
Generative AI Lab Use Cases

- 1. Use LLMs to bootstrap task-specific models**
- 2. Human-in-the-loop workflows to validate LLMs when regulatory-grade accuracy is required**
- 3. Organize and share models, prompts, and rules within one private enterprise hub**

From LLMs to task specific AI Models



Building Enterprise-Grade Human-in-the-Loop Pipelines



Versioning Human & AI Work

Full Audit Trails

Custom Approval Workflows

API Integration

Authentication & Authorization

Private On-Premise Deployment

Scale to Large Teams & Projects

Spark NLP

Open-Source Language Understanding at Scale

140+ million

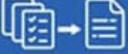
Downloads on PyPI.
“Most Widely Used NLP
Library in the Enterprise.”

60% growth

In Spark NLP downloads
since the 5.0 release
for RAG & LLM pipelines

8 years

Straight with frequent
releases & upgrades

Entity Recognition I love Lucy PERSON	Text Classification 	Spelling & Grammar abc She become the first... ✓ She became the first	Information Extraction They met Last week DATE -> 29-04-2020
Question Answering 	Speech to Text 	Image Classification 	Reading Comprehension 
Translation  [je t'aime -> i love you]	Summarization 	Paraphrasing You bet! > For sure.	Emotion Detection 

Split Text <ul style="list-style-type: none"> Sentence Detector Tokenizer Normalizer nGram Generator Word Segmentation 	Clean Text <ul style="list-style-type: none"> Spell Checker Grammar Checker Writing Style Checker Stopword Cleaner Summarization 	<p>100,000+ Pre-trained Pipelines, Models & Transformers</p> <table border="1"> <tbody> <tr><td>BERT</td><td>ELMO</td><td>TAPAS</td></tr> <tr><td>ALBERT</td><td>DeBERTa</td><td>USE</td></tr> <tr><td>Longformer</td><td>ELECTRA</td><td></td></tr> <tr><td>T5</td><td>NMT</td><td>VIT</td></tr> <tr><td>DistilBERT</td><td>RoBERTa</td><td></td></tr> <tr><td></td><td>XLM-RoBERTa</td><td></td></tr> <tr><td>Wav2Vec2</td><td>XLNet</td><td></td></tr> </tbody> </table>	BERT	ELMO	TAPAS	ALBERT	DeBERTa	USE	Longformer	ELECTRA		T5	NMT	VIT	DistilBERT	RoBERTa			XLM-RoBERTa		Wav2Vec2	XLNet		<p>250+ Languages</p> 
BERT	ELMO	TAPAS																						
ALBERT	DeBERTa	USE																						
Longformer	ELECTRA																							
T5	NMT	VIT																						
DistilBERT	RoBERTa																							
	XLM-RoBERTa																							
Wav2Vec2	XLNet																							
Understand Grammar <ul style="list-style-type: none"> Stemmer Lemmatizer Part of Speech Tagger Dependency Parser Translation 	Find in Text <ul style="list-style-type: none"> Text Matcher Regex Matcher Date Matcher Chunker Question Answering 																							
Trainable & Tunable 	Scalable 	Fast Inference 	Hardware Optimized   																					
			Community 																					

Spark NLP

Apache 2.0 license

Documentation:
www.sparknlp.org

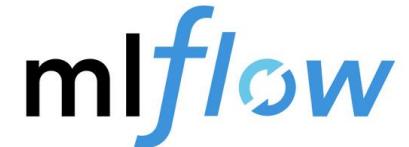
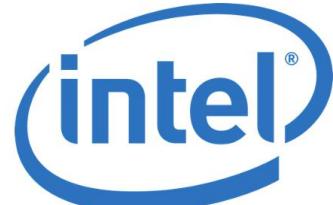
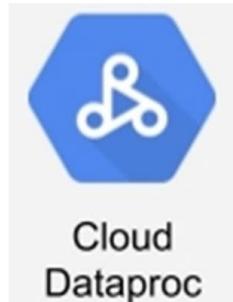
Community:
spark-nlp.slack.com

Own Curated Model Hub:
www.sparknlp.org/models

Optimized, Tested, Supported Integrations



CLOUDERA



kaggle



Spark NLP: Built for Scale



IT & Software > IT Certifications > Natural Language Processing (NLP)

Spark NLP for Data Scientists

Unlock your NLP power with Spark NLP, the most popular NLP library in enterprises

Bestseller 4.3 ★★★★☆ (57 ratings)

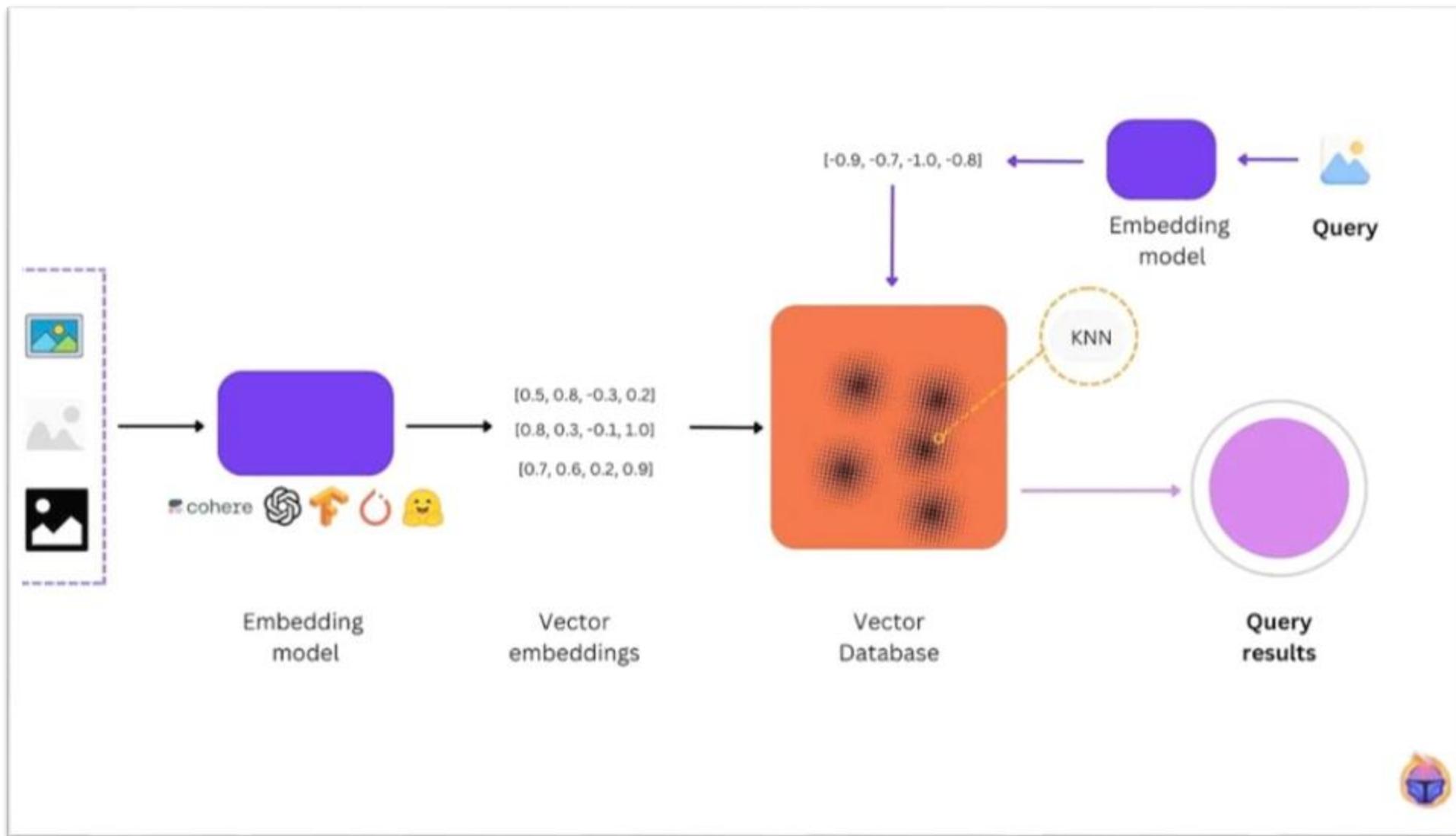
Compose multiple language models into custom pipelines

Natively scalable:
Extends the Spark ML Pipeline

Speed Optimized:
Builds for Intel, Nvidia, Apple

Own optimized codebase

RAG LLM Use Case: Populate a Vector Database

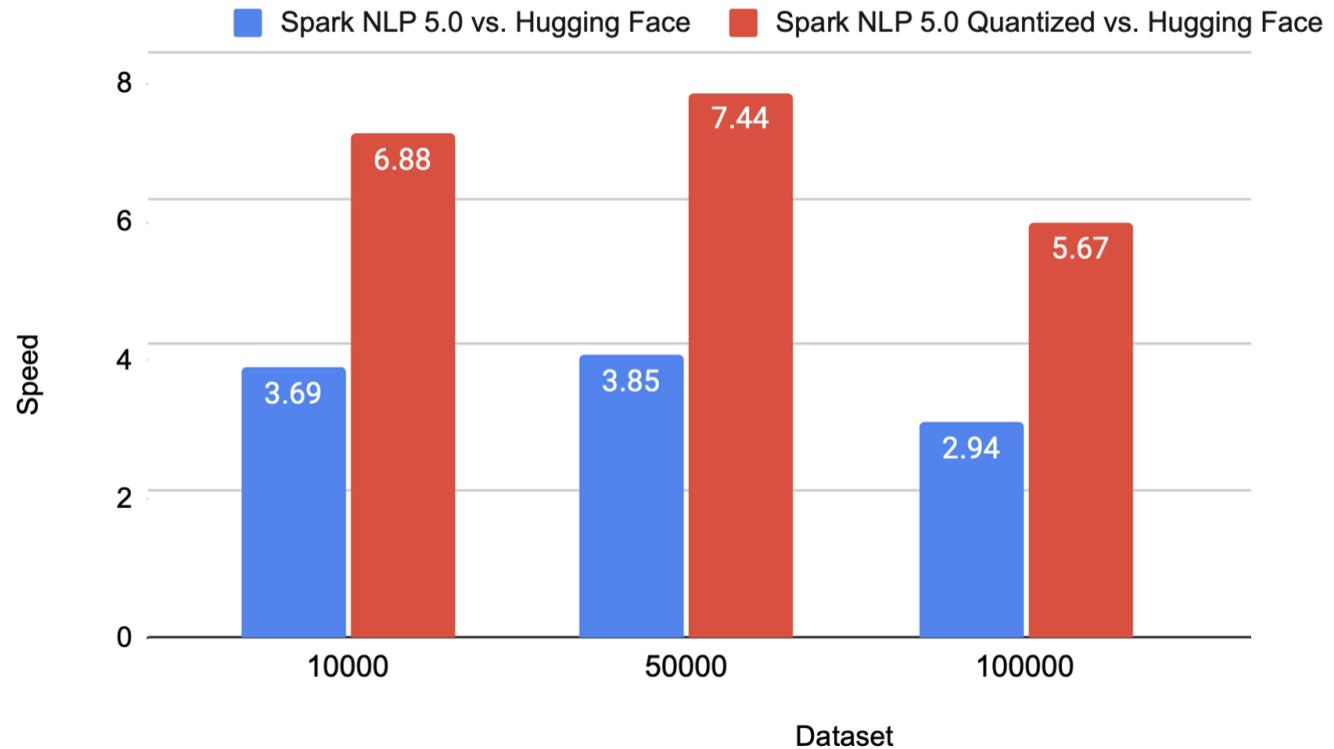


Calculate Embeddings 3x-7x Faster Than Hugging Face on a Single Server

- HPE Server:
 - AMD EPYC 7542 32-Core Processor
 - 80G memory
- Spark NLP based on Onyx Runtime vs. Hugging Face based on PyTorch

Comparison of Speed: Spark NLP vs Hugging Face in HPE Server

Spark NLP has demonstrated a performance improvement of 2.11 to 7.44 times over Hugging Face.

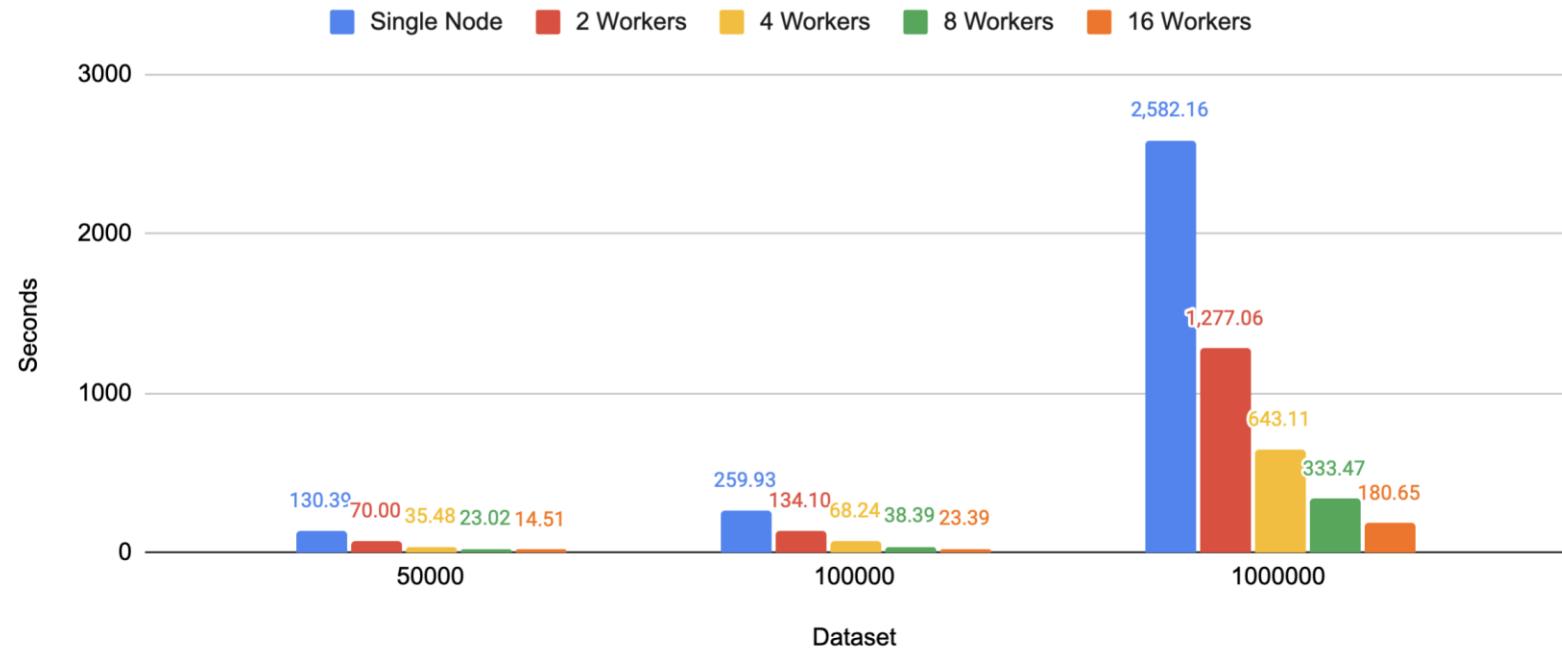


Scale Up With Zero Code Changes

- Databricks Single Node Cluster
 - 13.0 ML (includes Apache Spark 3.4.0, Scala 2.12)
 - c6i.8xlarge
 - 32 Cores
 - 64 GB Memory
- By natively scaling on the Databricks cluster and adding more executors, Spark NLP 5 achieves near-linear speedup.

Comparison of Speed: Spark NLP vs Hugging Face in Databricks multi-node Cluster

By natively scaling on the Databricks cluster and adding more executors, Spark NLP achieves nearly linear speed enhancements.



Processing of 1,000,000 records was reduced
from 43 hours to 3 minutes with zero code changes

LLM Inference Use Case: What if my text processing pipelines needs an LLM?

```
data = spark.createDataFrame([
    ("Is coffee consumption linked to improved cognitive function?",),
    ("What are the implications of quantum computing on data security?",),
    ("Describe the process of photosynthesis in plants.",),
    ("Can renewable energy effectively replace fossil fuels globally by 2040?",),
    ("What is the role of artificial intelligence in healthcare?",),
    ("How does the stock market influence the global economy?",),
    ("Explain the theory behind black holes and event horizons.",),
    ("What are the long-term effects of climate change on coastal ecosystems?",),
    ("Discuss the impact of virtual reality technology on education.",),
    ("What measures can be taken to reduce urban air pollution effectively?",)
], ["text"])
```

Question Answering, Text Generation, Summarization, Translation, Extraction

In Spark NLP, an LLM is just another pipeline step

- Distribute LLM inference on your hardware
- Tens of thousands of curated GGUF models
- Native integration with Llama.cpp
- Optimized builds for:
 - Intel processors
 - Nvidia CUDA GPUs
 - Apple Silicon chips

```
from sparknlp.annotator import *
from sparknlp.base import *

data = (spark.createDataFrame([["<|start_header_id|>user<|end_header_id|>
Make a coherent and concise summary:\n\nThere are only two movies I would give a 1/10 to, this stinker and "The Man who Fell to Earth." I remember seeing Protocol at a theater in the early 80s when I was in high school. ...<|eot_id|>
<|start_header_id|>assistant<|end_header_id|>
"""]]).toDF("text"))

document_assembler = (
    DocumentAssembler()
        .setInputCol("text")
        .setOutputCol("document")
)

gguf_model = (
    AutoGGUFModel.pretrained("Meta-Llama-3.1-8B-Instruct.Q4_K_M")
        .setInputCols("document")
        .setOutputCol("completions")
        .setBatchSize(2)
        .setNPredict(400)
        .setNGpuLayers(99)
        .setTemperature(0.4)
        .setTopK(40)
        .setTopP(0.9)
        .setPenalizeNL(True)
)

pipeline = Pipeline().setStages([
    document_assembler,
    gguf_model
])

results = pipeline.fit(data).transform(data)
results.select("completions").show(truncate=False)
```

Unified Pipelines: Combining Medical LLM & NLP Models



Previous attachment

Example of True Negative relation:

- She continues to tolerate treatment with **crizotinib** well. She does have continued **diarrhea** with 4 to 5/day and some **shortness of breath**.



Example of True Positive relation:

- Pt was started on **Alectinib** in October; developed worsening **bradycardia** with HR decreasing to the mid 30's.

AI-Enhanced Oncology Data: Unlocking Insights from EHRs with NLP and LLMs

Scott Newman – Senior VP, Life Sciences at MiBA

Zach Liu – Senior Research Scientist at MiBA

Based on annotation on 140 sampled sentence groups,
the evaluation resulted a F1 score of 0.93
(recall: 0.95, precision 0.91)

Visual NLP

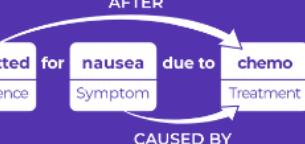
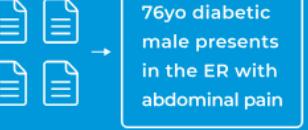
	Form understanding		Object Character Recognition		Table detection & extraction		Noisy image enhancement				
	Visual document classification		Visual entity recognition		Signature detection		Image de-identification				
Read	Enhance		Recognize	Extract		Generate					
	Text or PDF		Binarizer		Adaptive scaler		Regions		Text from Images (OCR)		Extracted text & layout (HOCR)
	Scanned PDF		Adaptive Tresholding		Noise Scorer		Tables		Data from Tables (CV)		Structured data tables
	DOCX		Erosion		Remove objects		Signatures		Named Entities from Forms		Highlighted named entities
	Image		Skew corrections		Morphology opening		Dates		De-identification		De-identified text, PDF, or DICOM
	DICOM		Scaler		Cropper		Brands		Visual Document Classification		Bounding boxes & coordinates

300+ live demos &
Python notebooks:

[nlp.johnsnowlabs.com
/demos](https://nlp.johnsnowlabs.com/demos)

Peer-reviewed
papers:

[johnsnowlabs.com/
peer-reviewed-papers/](https://johnsnowlabs.com/peer-reviewed-papers/)

Entity Recognition	Entity Linking	Assertion Status	Relation Extraction												
<p>EGFR BIOMARKER positive RESULT</p> <p>invasive HISTOLOGICAL_TYPE</p> <p>adenocarcinoma DIAGNOSIS</p> <p>classified as T1bN1M0 STAGING</p>	<p>Abdominal pain ICD10CM: R10.84</p> <p>Dermatitis MedDRA: 10012431</p> <p>Hernia repair SNOMED: 50465008</p>	<p>Fever and sore throat → PRESENT</p> <p>No stomach pain → ABSENT</p> <p>Father with Alzheimer → FAMILY</p>													
De-Identification	Question Answering	Summarization	Data Enrichment												
<p>Katia was born on April 29th PATIENT was born on DATE Olga was born on March 28th</p>	<p>Do preoperative stains reduce arterial fibrillation after CABG?</p> <p>YES</p>	 <p>76yo diabetic male presents in the ER with abdominal pain</p>	<p>Amoxicillin → RxNorm: 722 → drug class: antibiotic → brand: Amoxil, Larotid</p>												
Algorithms		Content													
Information Extraction <ul style="list-style-type: none"> Document Classification Entity Disambiguation Contextual Parsing Patient Risk Scoring 	Data Obfuscation <ul style="list-style-type: none"> Name Consistency Gender Consistency Age Group Consistency Format Consistency 	Medical Language Models Medical LLMs for: <p>Q&A RAG Extract Summarize</p> <p>Sizes: S M L Quantizations: q4 q8 q16</p>	Medical Terminologies <table border="1"> <tr><td>SNOMED-CT</td><td>CPT</td><td>UMLS</td></tr> <tr><td>ICD-10-CM</td><td>RxNorm</td><td>HPO</td></tr> <tr><td>ICD-10-PCS</td><td>ICD-O</td><td>LOINC</td></tr> <tr><td>MedDRA</td><td>NDC</td><td>MeSH</td></tr> </table>	SNOMED-CT	CPT	UMLS	ICD-10-CM	RxNorm	HPO	ICD-10-PCS	ICD-O	LOINC	MedDRA	NDC	MeSH
SNOMED-CT	CPT	UMLS													
ICD-10-CM	RxNorm	HPO													
ICD-10-PCS	ICD-O	LOINC													
MedDRA	NDC	MeSH													
Clinical Grammar <ul style="list-style-type: none"> Deep Sentence Detector Medical Spell Checking Medical Part of Speech Terminology Mapping 	Zero-Shot Learning <ul style="list-style-type: none"> Entities by Prompt Relations by Prompt Classification by Prompt Relative Data Extraction 	<p>2,500+ Pretrained Models</p> <table border="1"> <tr><td>Clinical Text Signs, Symptoms, Treatments, Findings, Procedures, Drugs, Tests, Labs, Vitals, Sections, Adverse Effects, Risk Factors, Anatomy, Social Determinants, Vaccines, Demographics, Sensitive Data</td><td>Biomedical Text Clinical Trial Design, Protocols, Objectives, Results; Research Summary & Outcomes; Organs, Cell Lines, Organisms, Tissues, Genes, Variants, Expressions, Chemicals, Phenotypes, Proteins, Pathogens</td></tr> </table>		Clinical Text Signs, Symptoms, Treatments, Findings, Procedures, Drugs, Tests, Labs, Vitals, Sections, Adverse Effects, Risk Factors, Anatomy, Social Determinants, Vaccines, Demographics, Sensitive Data	Biomedical Text Clinical Trial Design, Protocols, Objectives, Results; Research Summary & Outcomes; Organs, Cell Lines, Organisms, Tissues, Genes, Variants, Expressions, Chemicals, Phenotypes, Proteins, Pathogens										
Clinical Text Signs, Symptoms, Treatments, Findings, Procedures, Drugs, Tests, Labs, Vitals, Sections, Adverse Effects, Risk Factors, Anatomy, Social Determinants, Vaccines, Demographics, Sensitive Data	Biomedical Text Clinical Trial Design, Protocols, Objectives, Results; Research Summary & Outcomes; Organs, Cell Lines, Organisms, Tissues, Genes, Variants, Expressions, Chemicals, Phenotypes, Proteins, Pathogens														

Healthcare NLP

3 tasks where LLMs underperform:

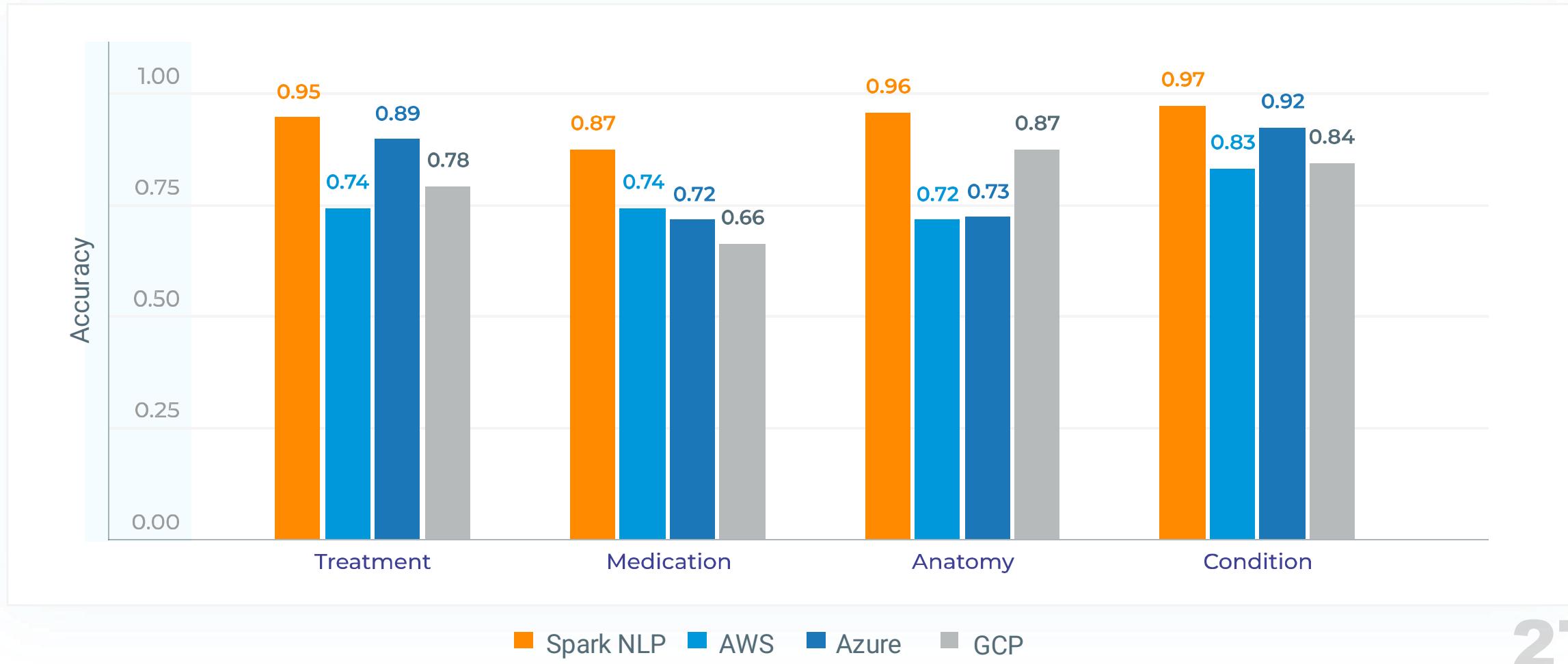
Information Extraction

De-Identification

Mapping Terms to Codes

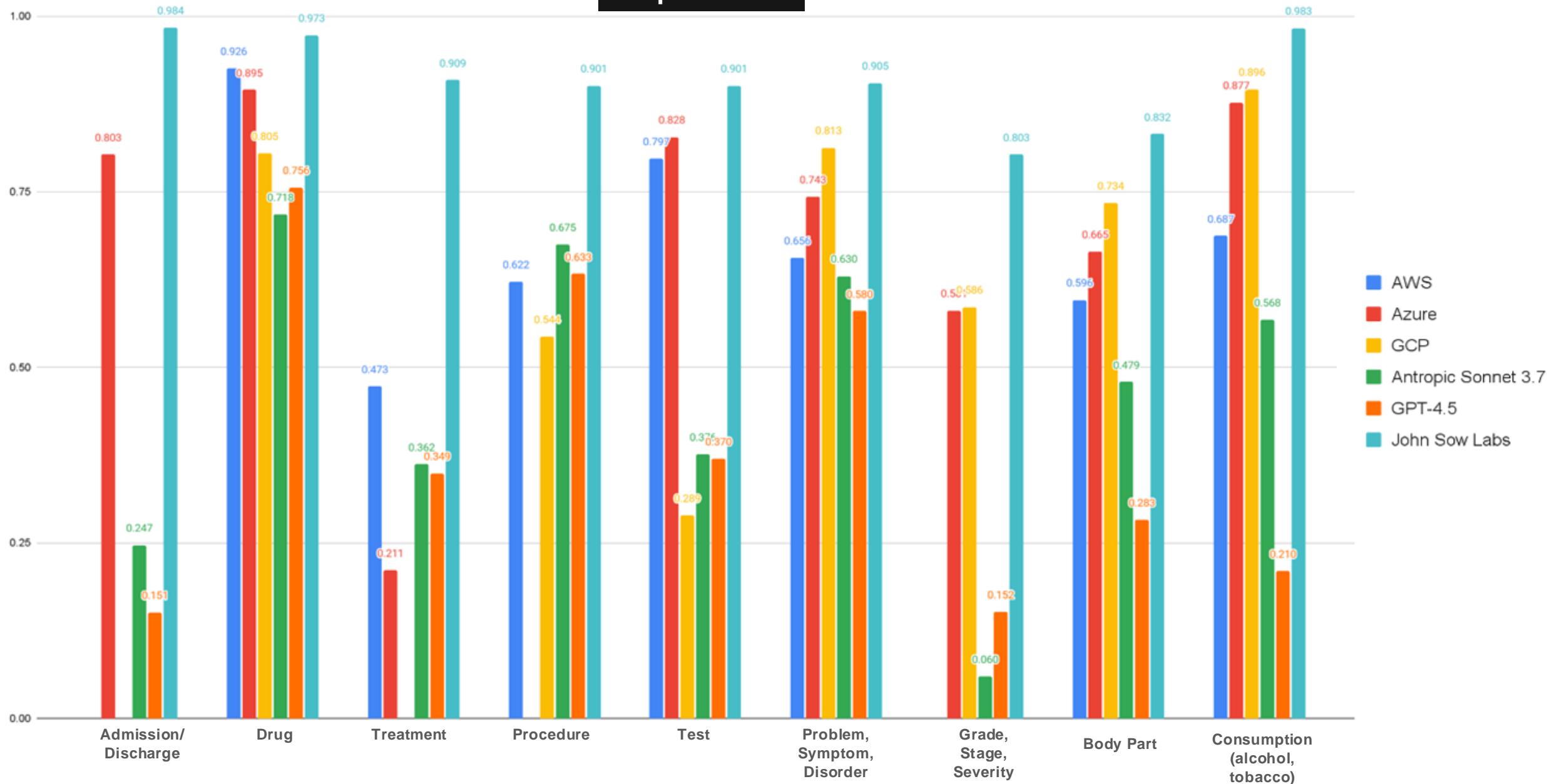
4-6x Fewer Errors than AWS, Azure, & GCP

www.johnsnowlabs.com/comparison-of-key-medical-nlp-benchmarks-spark-nlp-vs-aws-google-cloud-and-azure/



John Snow Labs vs Cloud & Frontier Models on Entity Recognition

April 2025



Clinical Text De-Identification of PHI Data

Healthcare NLP, Azure, Amazon, GPT-4o, and Claude 3.7 Sonnet PHI Recognition and Benchmark Comparison (Sample size: 45172 PHI entities).

Metric / Entity	Healthcare NLP			Azure			Amazon			GPT-4o			Claude 3.7 Sonnet		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
AGE	0.96	1.00	0.98	0.94	0.45	0.61	1.00	0.41	0.58	0.87	0.50	0.64	0.73	0.55	0.63
CONTACT	0.96	0.97	0.97	0.73	0.88	0.80	0.78	0.72	0.75	0.67	0.53	0.59	0.74	0.43	0.54
DATE	0.97	0.99	0.98	0.91	0.99	0.95	0.90	0.97	0.93	0.79	0.72	0.75	0.84	0.83	0.83
IDNUM	0.98	0.94	0.96	0.78	0.93	0.85	0.95	0.86	0.91	0.70	0.92	0.80	0.70	0.95	0.80
LOCATION	0.93	0.92	0.93	0.89	0.87	0.88	0.52	0.74	0.61	0.82	0.72	0.76	0.79	0.87	0.83
NAME	0.92	0.94	0.93	0.92	0.89	0.90	0.85	0.76	0.80	0.79	0.82	0.80	0.82	0.86	0.84
O	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Macro Avg	0.96	0.97	0.96	0.88	0.86	0.85	0.86	0.78	0.80	0.80	0.74	0.76	0.80	0.78	0.78
Non-PHI	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
PHI	0.96	0.97	0.96	0.91	0.92	0.91	0.81	0.85	0.83	0.81	0.77	0.79	0.81	0.84	0.83
Macro Avg	0.98	0.98	0.98	0.95	0.96	0.95	0.90	0.92	0.91	0.90	0.88	0.89	0.90	0.92	0.91
cost per 1M doc	\$2,418			\$13,125			\$14,525			\$21,400			\$23,330		

[Submitted on 21 Mar 2025 (v1), last revised 31 Mar 2025 (this version, v2)]

Can Zero-Shot Commercial APIs Deliver Regulatory-Grade Clinical Text Deldentification?

Veysel Kocaman, Muhammed Santas, Yigit Gul, Mehmet Butgul, David Talby

RxNorm Entity Resolution Benchmarks

PAST MEDICAL HISTORY: Diabetes.

PAST SURGICAL HISTORY: Hernia repair.

ALLERGIES: He has no allergies.

MEDICATIONS: Listed in the chart and include

diltiazem	B12	Prevacid
DRUG	DRUG	DRUG
3443	11248	83156
DILTIAZEM	VITAMIN B12	PREVACID

Coumadin	Lasix	metformin	folic acid
DRUG	DRUG	DRUG	DRUG
202421	202991	6809	4511
COUMADIN	LASIX	METFORMIN	FOLIC ACID

<https://medium.com/john-snow-labs/state-of-the-art-rxnorm-code-mapping-with-nlp-comparative-analysis-between-the-tools-by-john-snow-794b353dc38e>

	Top-3 Accuracy	Top-5 Accuracy	Cost
Healthcare NLP	82.7%	84.6%	\$4,500
Amazon Comprehend Medical	55.8%	56.2%	\$24,250
GPT-4 (Turbo)	8.9%	8.9%	\$44,000
GPT-4o	8.9%	8.9%	\$22,000

Medical Language Models Workshop Outline

Clinical Data
Abstraction

Patient Risk
Adjustment

Summarization
and Q&A

Text
De-Identification

DICOM
De-Identification

Terminology
Server

How to Get The Most from This Training

- 1. Try the libraries and tools hands-on**
- 2. Ask questions**
- 3. Learn how the tools apply to use cases:
www.johnsnowlabs.com/customers**

Thank you!