John Snow LABS

Oncology Use Cases

Muhammet ŞANTAŞ

Senior Data Scientist, John Snow Labs

Agenda

1. Introduction
2. Case Study Samples
3. Problem Description
4. Solution Steps
5. Coding

John Snow Labs

Provides State-of-the-Art Medical Language Models

100+ million	59% share	#1 Accuracy
Downloads on PyPI. “Most Widely Used NLP Library in the Enterprise.”	of Healthcare NLP teams use John Snow Labs	on 25 benchmarks in peer-reviewed papers
O’Reilly Media	**Gradient Flow**	**Papers with Code**

Peer-Reviewed, State-of-the-Art Accuracy

John Snow Labs Peer-Reviewed Papers

Deeper Clinical Document Understanding Using Relation Extraction

New state-of-the-art accuracy on:

2019 Phenotype-Gene Relations dataset
2018 n2c2 Posology Relations dataset
2012 Adverse Drug Events Drug-Reaction dataset
2012 i2b2 Clinical Temporal Relations challenge
2010 i2b2 Clinical Relations challenge

Mining Adverse Drug Reactions from Unstructured Mediums at Scale

New state-of-the-art accuracy on:

ADE benchmark
SMM4H benchmark
CADEC entity recognition dataset
CADEC relation extraction dataset

Biomedical Named Entity Recognition in Eight Languages with Zero Code Changes

New state-of-the-art accuracy on:

LivingNER dataset using a single model architecture in
English, French, Italian, Portuguese, Galatian, Catalan & Romanian

Accurate Clinical and Biomedical Named Entity Recognition at Scale

New state-of-the-art accuracy on:

2018 n2c2 medication extraction
2014 n2c2 de-identification
2010 i2b2/VA clinical concept extraction
8 different Biomedical NLP benchmarks

Oncology Case Studies from the NLP Summit

Mount Sinai

A Real-time NLP-Based Clinical Decision Support Platform for Psychiatry and Oncology

Roche

Accelerating Biomedical Innovation by Combining NLP and Knowledge Graphs

National Cancer Centre Singapore
SingHealth

Large Language Models to Facilitate Building of Cancer Data Registries

COTA

Leveraging Healthcare NLP Models in Regulatory Grade Oncology Data Curation

myNEO

Applying Natural Language Processing to Cancer Genomics

Fred Hutch Cancer Center

"Be Like Water" – NLP and GenAI Opportunities and Challenges in Healthcare

Real-World Oncology Use Case-I

John Snow LABS

ICD-10-CM Code Selection for Billing in an Oncology Hospital

PROBLEM: An oncology hospital needs to accurately select billable ICD-10-CM codes for insurance reimbursement. This process involves identifying the patient's current diagnoses, extracting the corresponding ICD-10-CM codes, and verifying that these codes are eligible for billing. Currently, this task is performed manually by medical coders, which is time-consuming, error-prone, and susceptible to human error.

Real-World Oncology Use Case-I

ICD-10-CM Code Selection for Billing in an Oncology Hospital

PROBLEM: An oncology hospital needs to accurately select billable ICD-10-CM codes for insurance reimbursement. This process involves identifying the patient's current diagnoses, extracting the corresponding ICD-10-CM codes, and verifying that these codes are eligible for billing. Currently, this task is performed manually by medical coders, which is time-consuming, error-prone, and susceptible to human error.

Real-World Oncology Use Case

John Snow LABS

Solution Steps

Step 1

Entity Extraction (NER)

Create a robust NER pipeline and extract oncological entities and body parts.

Step 2

Assertion Status Detection

Check the assertion status of the detected oncological entities.

Step 3

Relation Extraction

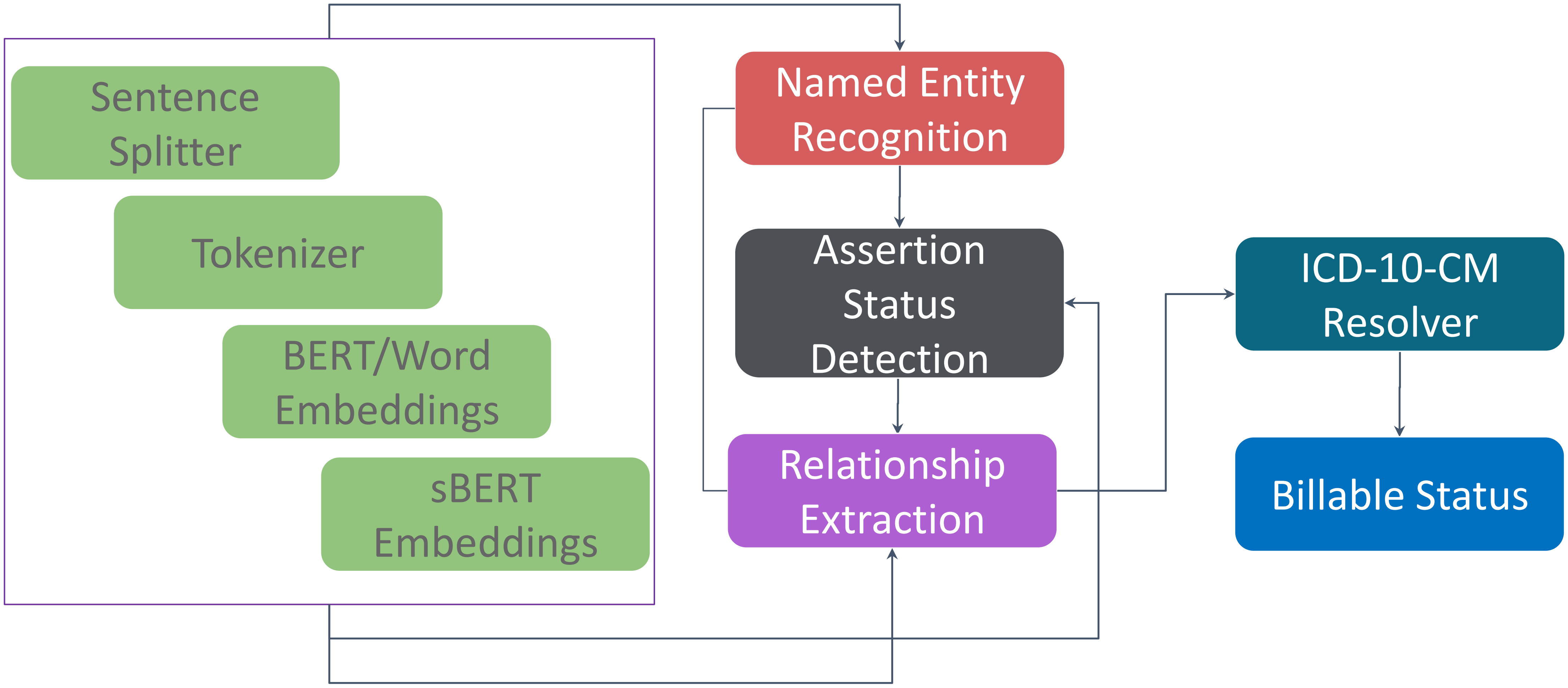
Extract relations between the oncological entities and body parts.

Step 4

Code Mapping & Billing Status

Map the entities to their corresponging ICD-10-CM codes and check their billable status.

Real-World Oncology Use Case



(Resolvers can be mapped to each other using the pretrained mapper module or resolved from scratch using the same sBert Embeddings)

Out-of-the-Box Oncology NLP Models

Medical NER Models

ner_biomarker	ner_oncology_langtest
ner_biomarker_langtest	ner_oncology_limited_80p_for_benchmarks
ner_oncology	ner_oncology_posology
ner_oncology_anatomy_general	ner_oncology_posology_langtest
ner_oncology_anatomy_general_healthcare	ner_oncology_response_to_treatment
ner_oncology_anatomy_general_langtest	ner_oncology_response_to_treatment_langtest
ner_oncology_anatomy_granular	ner_oncology_test
ner_oncology_anatomy_granular_langtest	ner_oncology_test_langtest
ner_oncology_biomarker	ner_oncology_therapy
ner_oncology_biomarker_healthcare	ner_oncology_therapy_langtest
ner_oncology_biomarker_langtest	ner_oncology_tnm
ner_oncology_demographics	ner_oncology_tnm_langtest
ner_oncology_demographics_langtest	ner_oncology_unspecific_posology
ner_oncology_diagnosis	ner_oncology_unspecific_posology_healthcare
ner_oncology_diagnosis_langtest	ner_oncology_unspecific_posology_langtest
ner_oncology_emb_clinical_large	
ner_oncology_emb_clinical_medium	

Relation Extraction Models

re_oncology
re_oncology_biomarker_result
re_oncology_granular
re_oncology_location
re_oncology_size
re_oncology_temporal
re_oncology_test_result

Relation Extraction DL Models

redl_oncology_biobert
redl_oncology_biomarker_result_biobert
redl_oncology_granular_biobert
redl_oncology_location_biobert
redl_oncology_size_biobert
redl_oncology_temporal_biobert
redl_oncology_test_result_biobert

Assertion Status Detection Models

assertion_oncology
assertion_oncology_demographic_binary
assertion_oncology_family_history
assertion_oncology_problem
assertion_oncology_response_to_treatment
assertion_oncology_smoking_status
assertion_oncology_test_binary
assertion_oncology_treatment_binary

Sequence Classification Models

bert_sequence_classifier_biomarker
bert_sequence_classifier_response_to_treatment

32 NER Models

7 Relation Extraction Models

7 Relation Extraction DL Models

8 Assertion Status Detection Models

2 Sequence Classification Models

Out-of-the-Box Oncology NLP Models

Explore Oncology Notes with Spark NLP Models

This demo shows how oncological terms can be detected using Spark NLP Healthcare NER, Assertion (...)

Live Demo Colab

Identify Anatomical and Oncology entities related to different Treatments and Diagnosis from Clinical Texts

This demo shows how to extract more than 40 Oncology-related entities including those related (...)

Live Demo Colab

Identify Tests, Biomarkers, and their Results

This demo shows how to extract entities Pathology Tests, Imaging Tests, mentions of Biomarkers, (...)

Live Demo Colab

Identify Demographic Information from Oncology Texts

This demo shows how to extract Demographic information, Age, Gender, and Smoking status fro (...)

Live Demo Colab

Detect Assertion Status from Clinics Entities

This demo shows how to detect the assertion status of entities related to oncology (including diagnoses, (...)

Live Demo Colab

Detect Relation Extraction between different Oncological entity types

This demo shows how to identify relations between Clinical entities, Tumor mentions, Anatomical (...)

Live Demo Colab

Resolve Oncology terminology using the ICD-O taxonomy

This model maps oncology terminology to ICD-O codes using Entity Resolvers.

Live Demo Colab

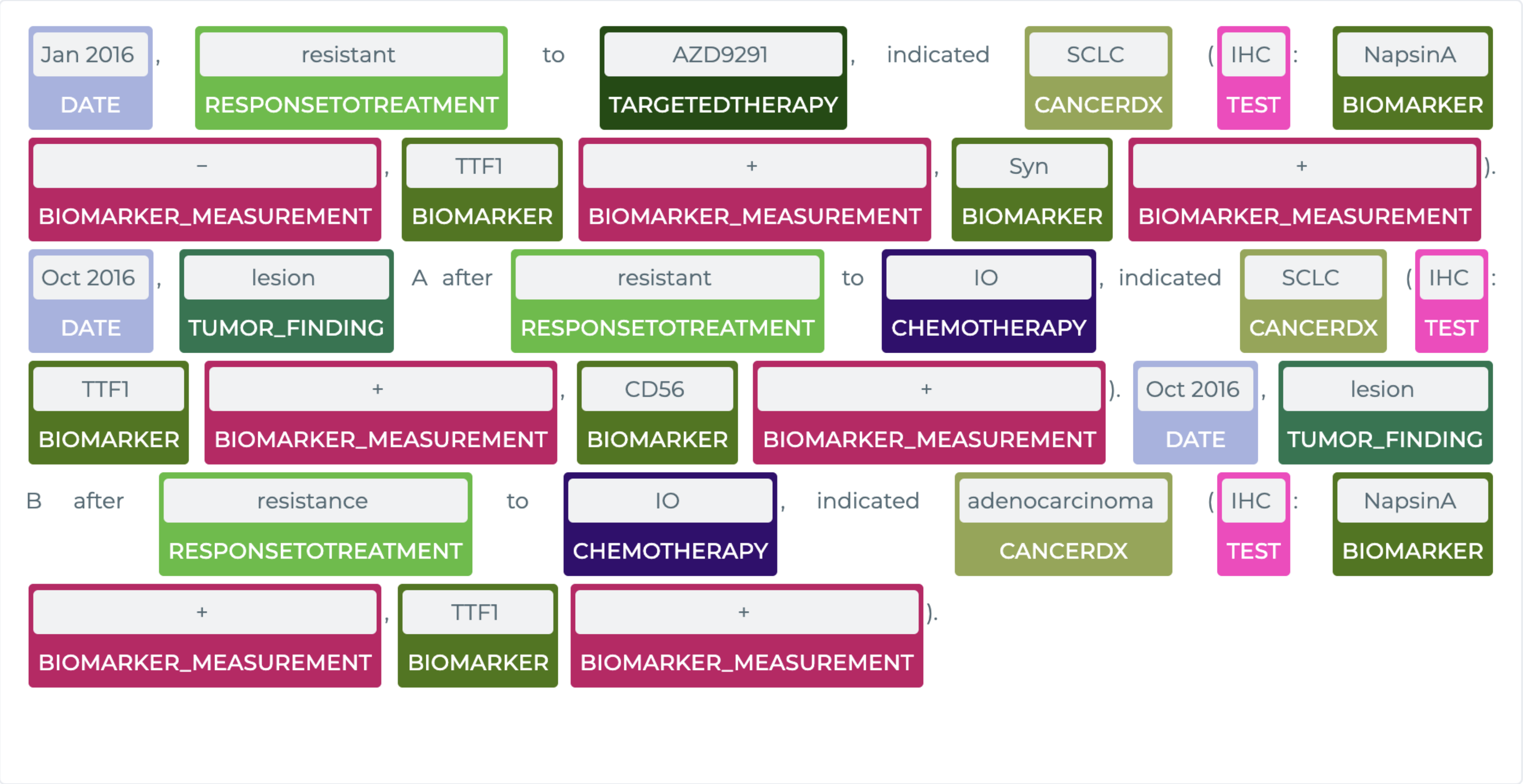
https://nlp.johnsnowlabs.com/oncology

Pre-trained Entity Recognition - Oncology

Text annotated with identified Named Entities

A 65-year-old Taiwanese woman had a history of debulking surgery **CANCERSURGERY**, bilateral **DIRECTION** oophorectomy **CANCERSURGERY** with omentectomy **CANCERSURGERY**, total anterior hysterectomy **CANCERSURGERY** with radical pelvic lymph nodes dissection **CANCERSURGERY** due to ovarian carcinoma **CANCERDX** (mucinous-type carcinoma **CANCERDX**, stage Ic **STAGING**) 1 year ago **RELATIVEDATE**. Patient's medical compliance was poor and failed to complete her chemotherapy **CHEMOTHERAPY** (cyclophosphamide **CHEMOTHERAPY** 750 mg/m2 **DOSAGE**, carboplatin **CHEMOTHERAPY** 300 mg/m2 **DOSAGE**). Recently, she noted a palpable right **DIRECTION** breast **SITEBREAST** mass **TUMOR_FINDING**, which enlarged rapidly to about 15 cm **TUMORSIZE** in size and nearly occupied the whole right **DIRECTION** breast **SITEBREAST** in 2 months **DURATION**. Core needle biopsy **PATHOLOGYTEST** revealed metaplastic **HISTOLOGICALTYPE** carcinoma **CANCERDX**. Neoadjuvant chemotherapy **CHEMOTHERAPY** with the regimens of Taxotere **CHEMOTHERAPY** (75 mg/m2 **DOSAGE**), Epirubicin **CHEMOTHERAPY** (75 mg/m2 **DOSAGE**), and

Pre-trained Entity Recognition - Biomarkers



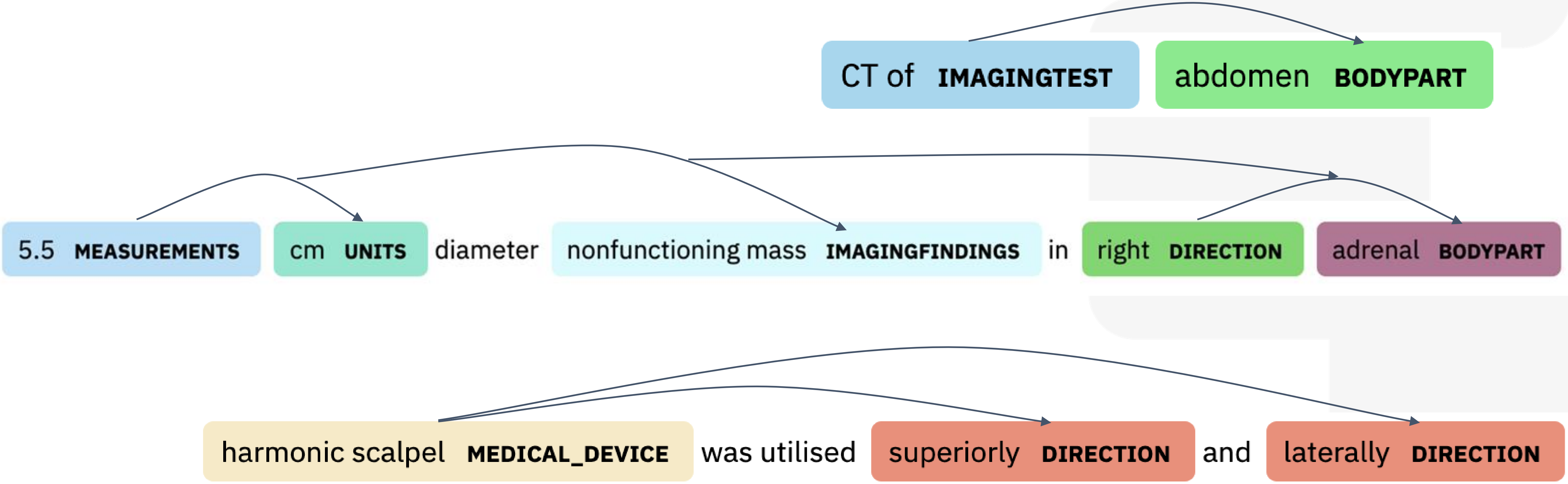
Assertion Status Detection

"Mother with a lung cancer, a patient is diagnosed as breast cancer in 1991 and then admitted to Mayo Clinic in Oct 2000, went under chemo for 6 months, discharged in April 2001 with a prescription of 2 mg metformin 3x per day. No sign of gynecological disorder but she suffers from acute cramps if she doesn't take her drug."

Chunk	Entity	Assertion
lung cancer	Oncological	Family
breast cancer	Oncological	Past
chemo	Treatment	Past
gynecological disorder	Disorder	Absent
acute cramps	Disorder	Conditional

Relation Extraction

"This is a 52-year-old inmate with a 5.5 cm diameter nonfunctioning mass in his right adrenal shown by CT of abdomen. During the umbilical hernia repair, the harmonic scalpel was utilised superiorly and laterally."



Entity Resolution to Standard Terminologies

This is a 52-year-old AGE inmate with a 5.5 MEASUREMENTS cm UNITS diameter nonfunctioning mass SYMPTOM in his GENDER right DIRECTION adrenal BODYPART shown by CT of IMAGINGTEST abdomen BODYPART . During the umbilical hernia repair PROCEDURE , the harmonic scalpel MEDICAL_DEVICE was utilised superiorly DIRECTION and laterally DIRECTION .

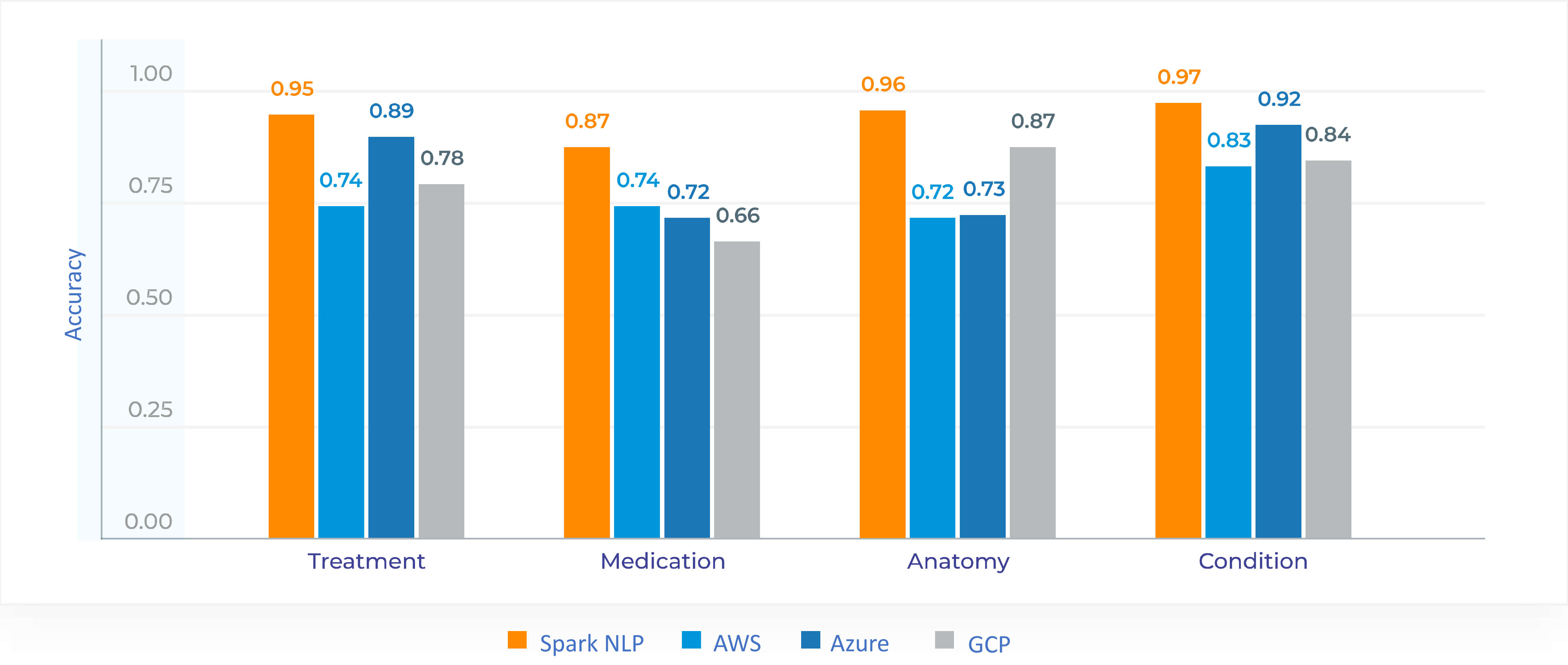
Entity Resolution

ICD10CM, Snomed, RxNorm, CPT-4, ICD10CPS, RxCUI, ICDO, UMLS, ATC, HPO,

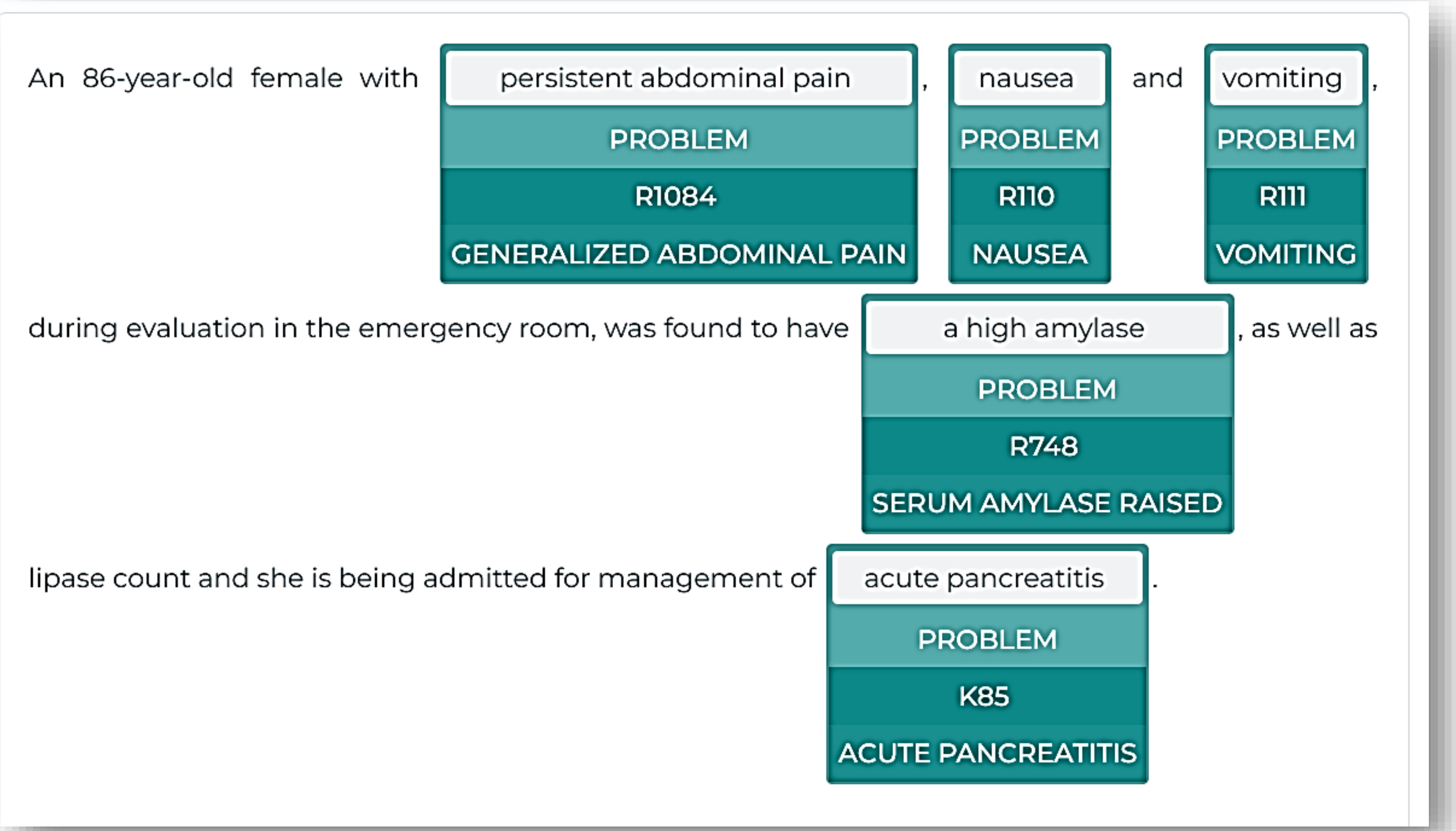
Term	Vocab	Code	Explanation (ground truth)
CT	CPT-4	76497	Unlisted computed tomography procedure
CT of abdomen	CPT-4	74150	Computed tomography, abdomen; without contrast material
umbilical hernia repair	CPT-4	49587	Repair umbilical hernia, age 5 years or older; incarcerated or strangulated
nonfunctioning mass, right adrenal	ICD10CM	D35.01	Benign neoplasm of right adrenal gland

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/



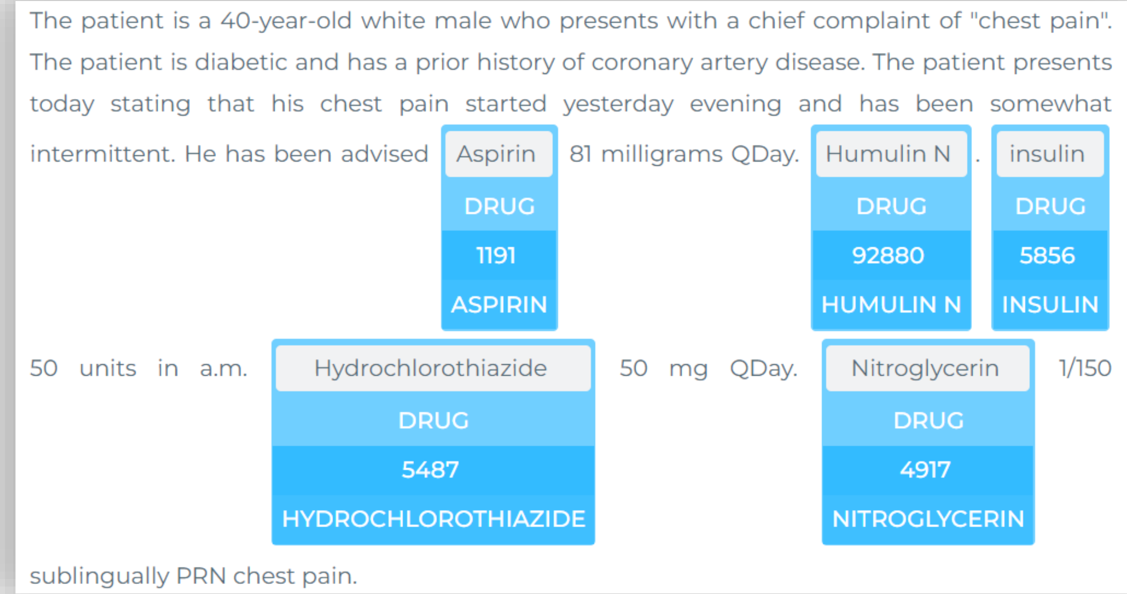
Benchmark: Extracting ICD-10-CM Codes

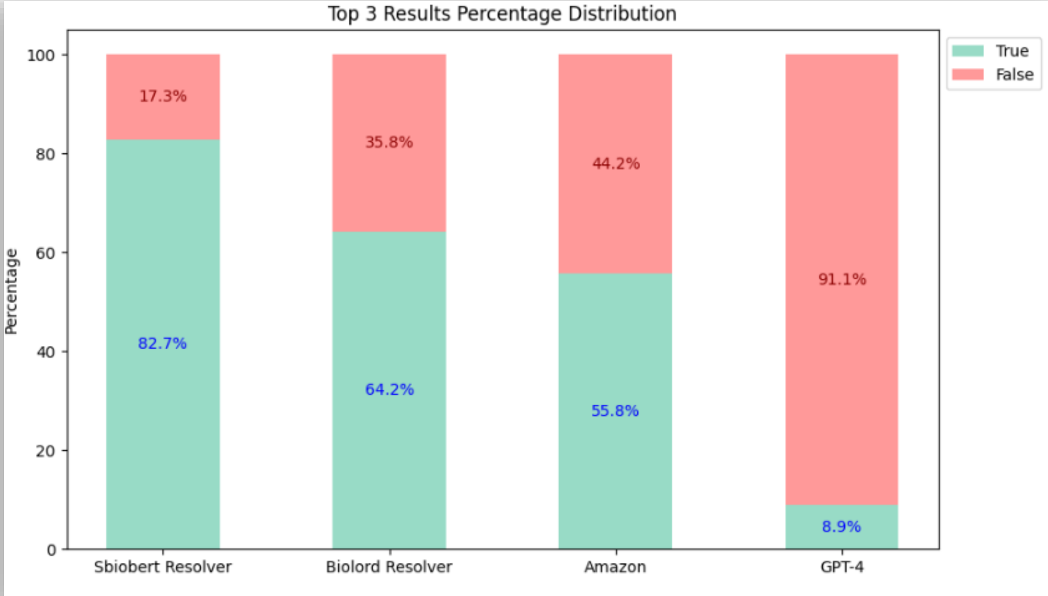


Extracting ICD-10-CM codes is done with a **76% success rate vs. 26% for GPT-3.5 and 36% for GPT-4.**

https://www.johnsnowlabs.com/comparing-spark-nlp-for-healthcare-and-chatgpt-in-extracting-icd10-cm-codes-from-clinical-notes/

Benchmark: Extracting RxNorm Codes





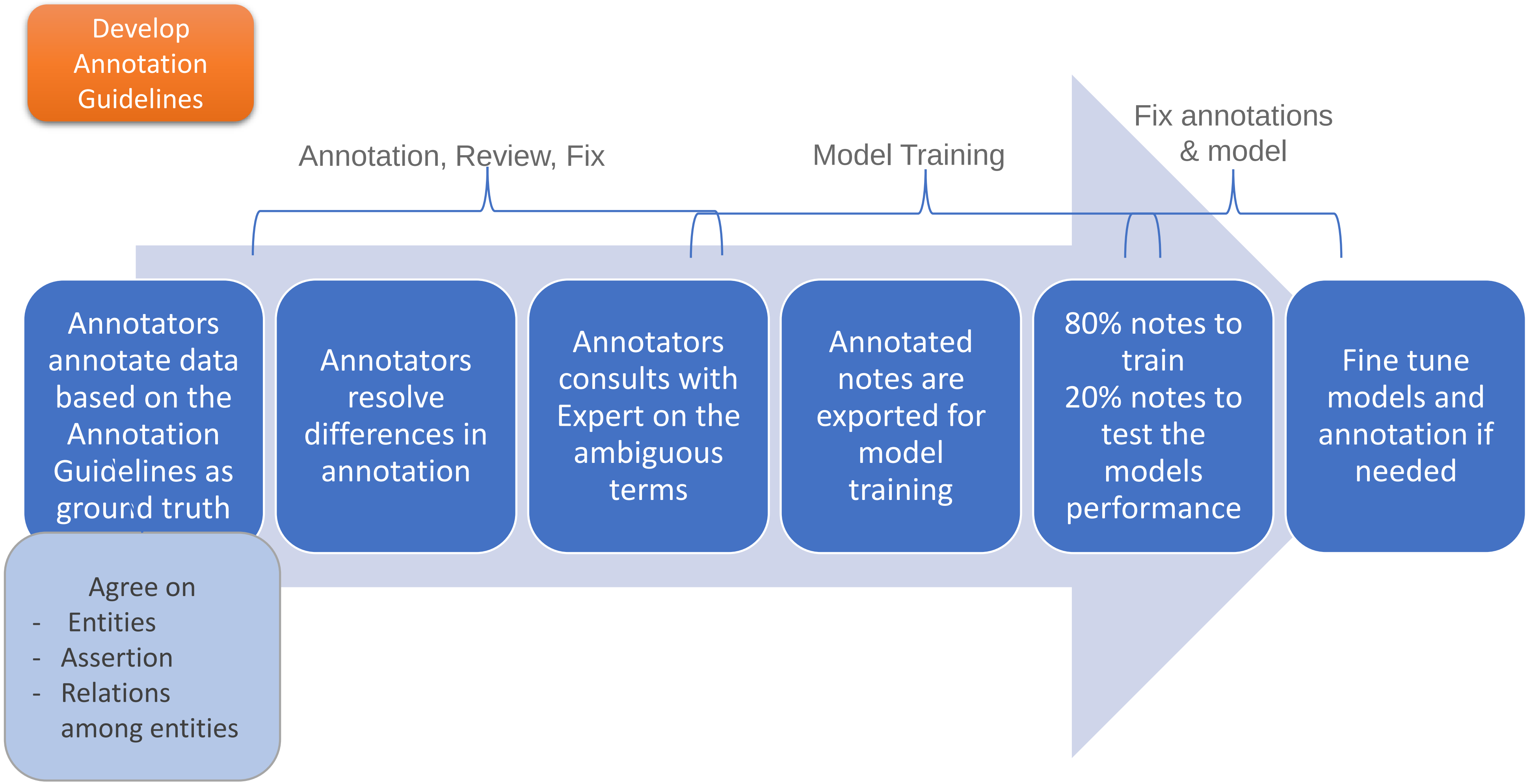
	Top-3 Accuracy	Top-5 Accuracy	Cost
Healthcare NLP	82.7%	84.6%	$4,500
Amazon Comprehend Medical	55.8%	56.2%	$24,250
GPT-4 (Turbo)	8.9%	8.9%	$44,000
GPT-4o	8.9%	8.9%	$22,000

Extracting RxNorm codes is done with a **82.7% success rate vs. 55.8% for Amazon and 8.9% for GPT-4.**

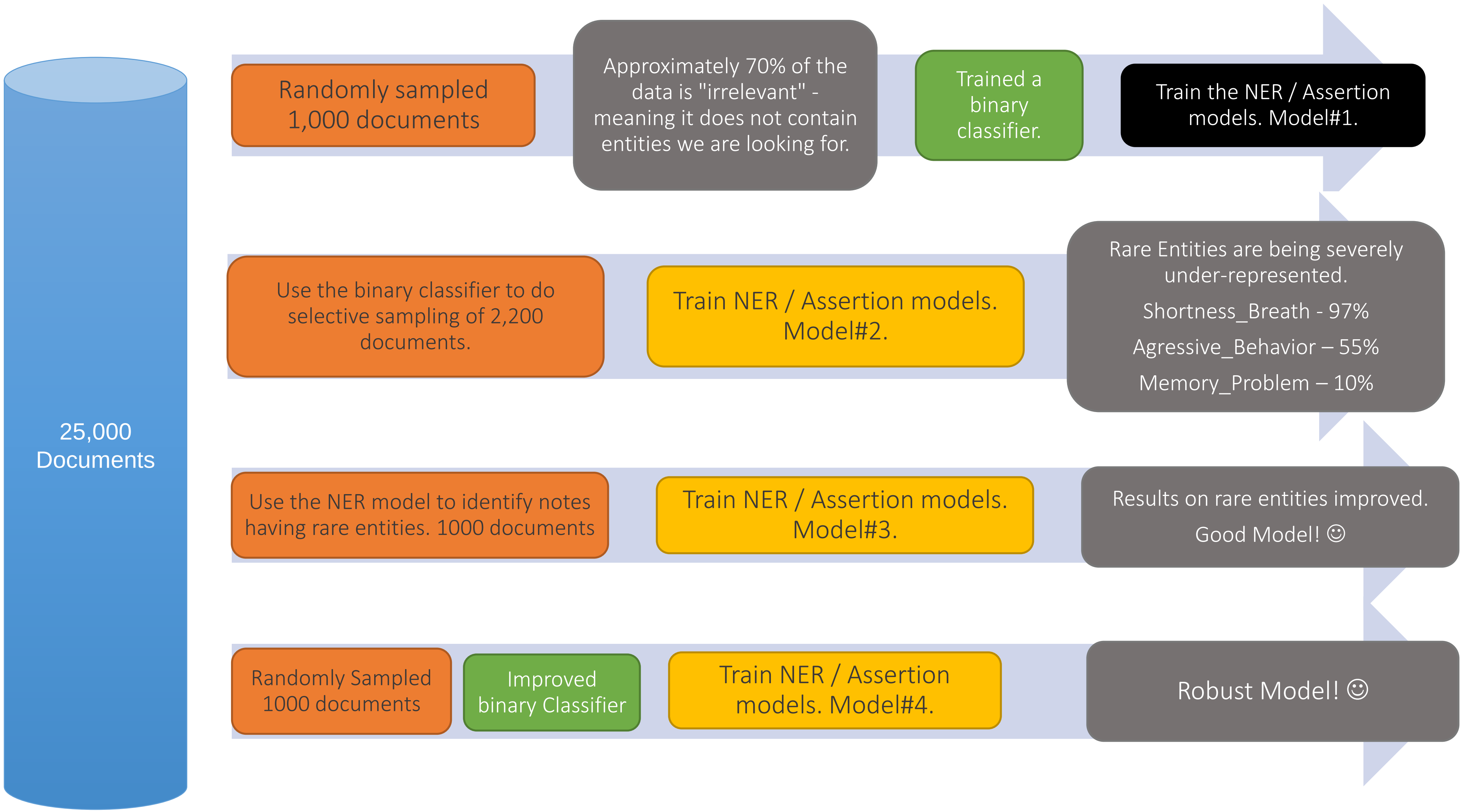
Also **5x times cheaper!**

RxNorm Code Mapping with NLP: Comparative Analysis between the tools by John Snow Labs, Amazon, and GPT-4

Model Training Process overview



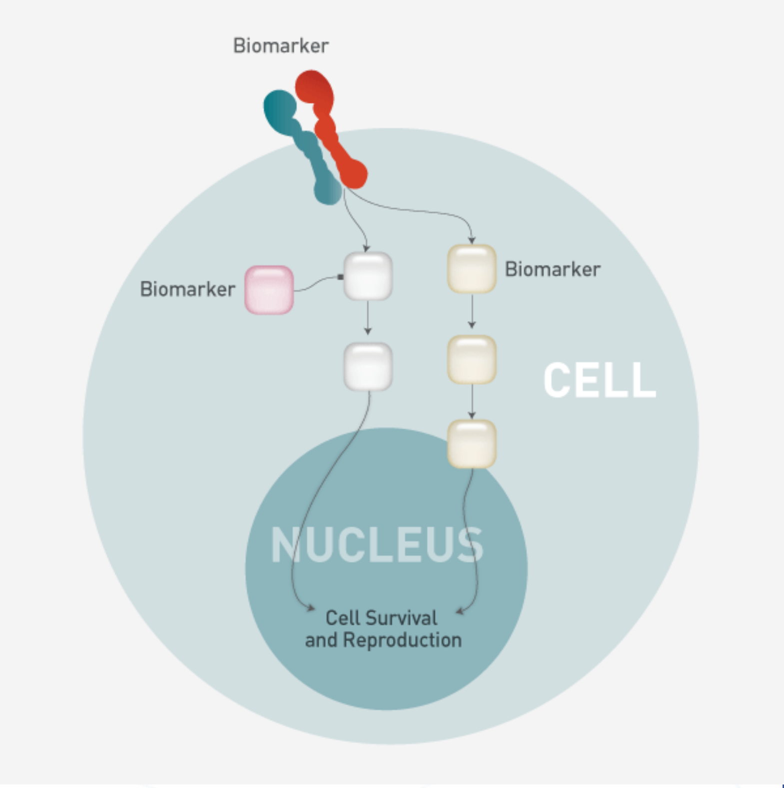
Model Training and Evaluation Process



Real-World Oncology Use Case-II

Biomarker and Biomarker Result Table Generation from Oncology Notes

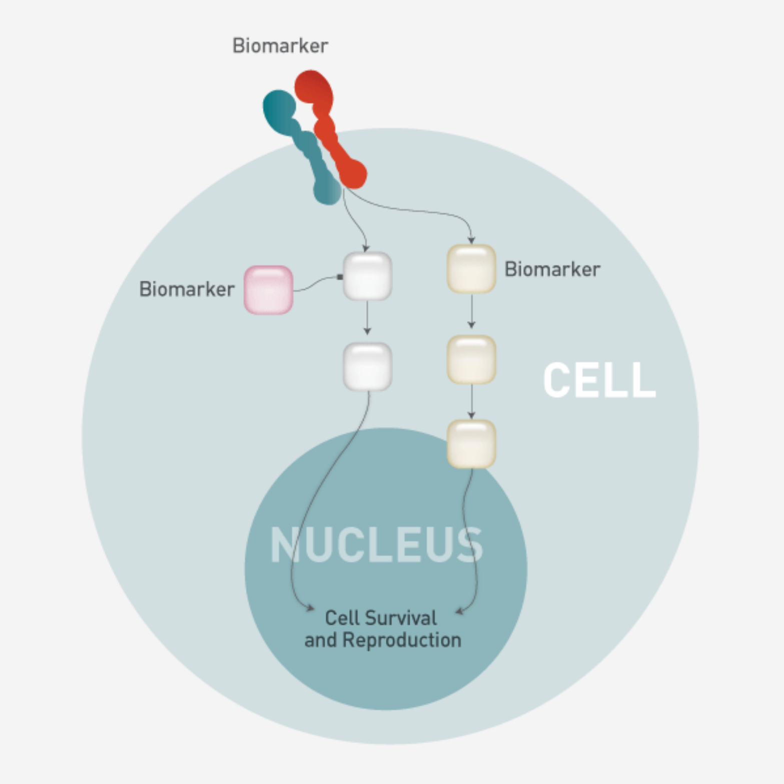
PROBLEM: Oncology researchers require a method to efficiently extract and organize biomarker and biomarker result information specifically from sections within oncology notes dedicated to biomarker analysis. This information is crucial for data analysis and biomarker research. Currently, manually extracting this data from extensive oncology notes is time-consuming, labor-intensive, and susceptible to human error.



Real-World Oncology Use Case-II

Biomarker and Biomarker Result Table Generation from Oncology Notes

PROBLEM: Oncology researchers require a method to efficiently extract and organize biomarker and biomarker result information specifically from sections within oncology notes dedicated to biomarker analysis. This information is crucial for data analysis and biomarker research. Currently, manually extracting this data from extensive oncology notes is time-consuming, labor-intensive, and susceptible to human error.



Real-World Oncology Use Case-II

John Snow LABS

Solution Steps

Step 1	Step 2	Step 3
Sequence Classification	**Entity Extraction (NER)**	**Relation Extraction**
Classify the sequences to determine whether they are related to the biomarker or not.	Create a robust NER pipeline and extract biomarker and biomarker result entities from the sequences related to biomarker.	Extract relations between the biomarker and biomarker results.

Let’s code!