

UNLOCKING NEW AI USE CASES WITH SHARING



Darshana Sivakumar, Databricks
David Talby, John Snow Labs
June 12th, 2024

Product safe harbor statement

This information is provided to outline Databricks' general product direction and is for **informational purposes only**. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

Agenda

- Importance of collaboration in the age of AI
- Databricks Data Intelligence Platform and collaboration features
- John Snow Labs' use of Databricks Marketplace and Delta Sharing
- Customer case study on AI collaboration
- Demo
- Key takeaways

Generative AI is taking the world by storm

91%



of organizations are experimenting with or investing in GenAI¹

75%



of CEOs say companies with advanced GenAI will have a competitive advantage²

40%



increase in performance of employees who used GenAI³

1. Laying the foundation for data and AI-led growth, [MIT Technology Review](#)

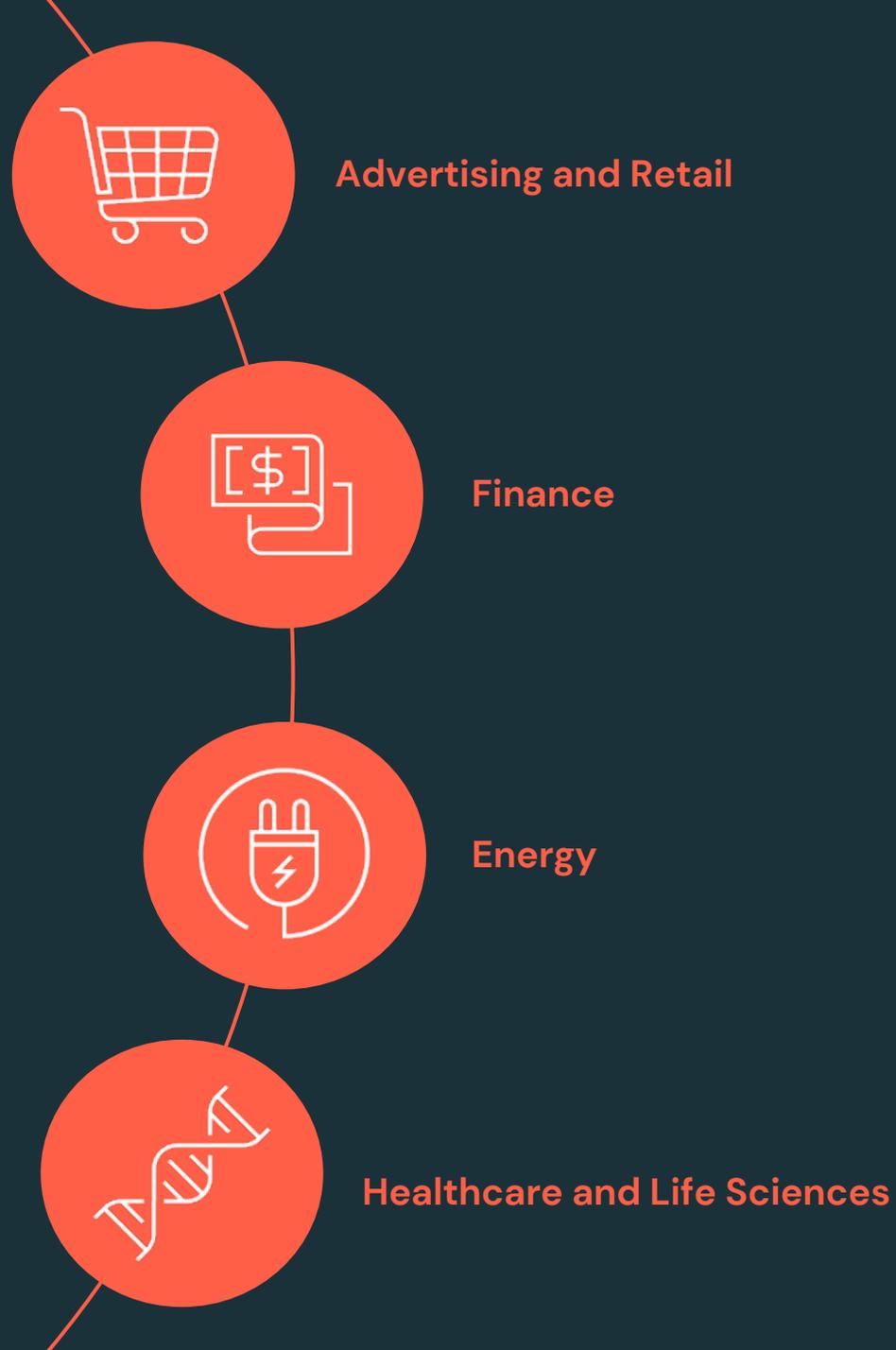
2. CEO decision-making in the age of AI, [IBM Institute for Business Value](#)

3. How generative AI can boost highly skilled workers' productivity, [MIT Management Sloan School](#)



AI is pervasive

...transcending industries
and user personas



Databricks Data Intelligence Platform

Unify your data and AI for better quality applications

Mosaic AI Gen AI

- Custom models
- RAG
- Vector Search

End-to-end AI

- MLOps/MLflow
- AutoML
- Feature engineering
- Model serving

Unity Catalog

Unified governance
End-to-end lineage
Monitoring

DBRX

2x smarter than Llama2
Extremely fast
Cost efficient
Open source



Databricks Data Intelligence Platform

Three main value-adds

Complete control

Decide how you want to build or use a model

Control model and data

Control who has access to what data and AI assets

Production quality

Use RAG to augment LLMs with enterprise-context

Features like lineage, monitoring and governance to enable building of higher quality models

Lower cost

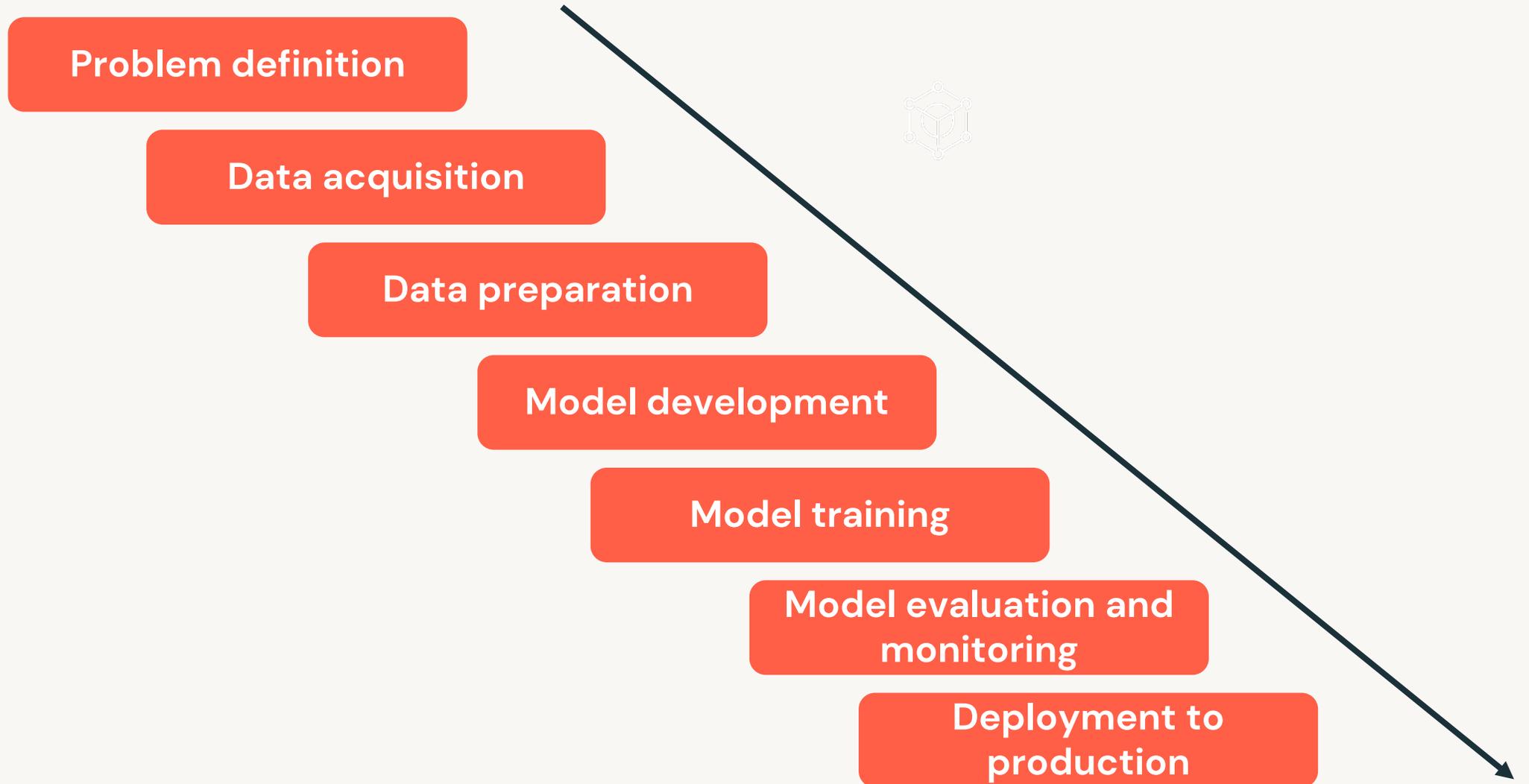
Optimized stack for building large models

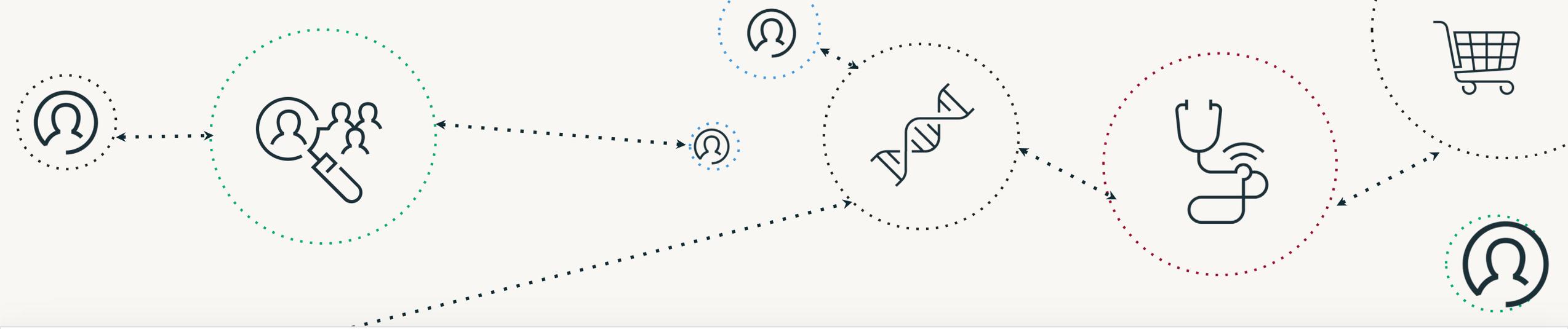
Combination of different techniques like tuned parallelism for increased compute utilization

Proven track record of lowering costs up to 10x

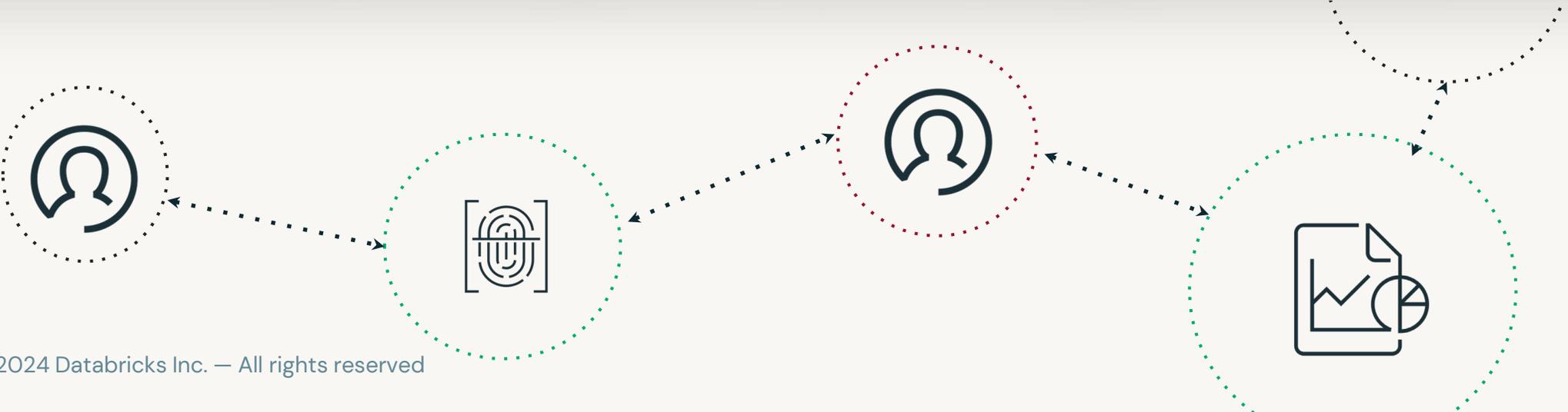


Collaboration throughout the AI Lifecycle





But data and AI asset sharing
is **complex and expensive**



Data and AI collaboration is complex

Difficult to access good quality data (training, evaluation, fine tuning)



Unable to train & evaluate high-quality models

Difficult to share non-structured data assets



Inability to share AI assets (notebooks, models, unstructured data)

Using external 3P model often requires exposing data to model provider



Trade-off between privacy and accessing 3P external models



Databricks platform makes Data and AI asset sharing easy

Discover high-quality data and AI assets through Marketplace

Share data sets, AI models, notebooks and solutions in a unified way

Take full control of data and AI assets



Discover high-quality data and AI assets through Marketplace

Share data sets, AI models, notebooks and solutions in a unified way

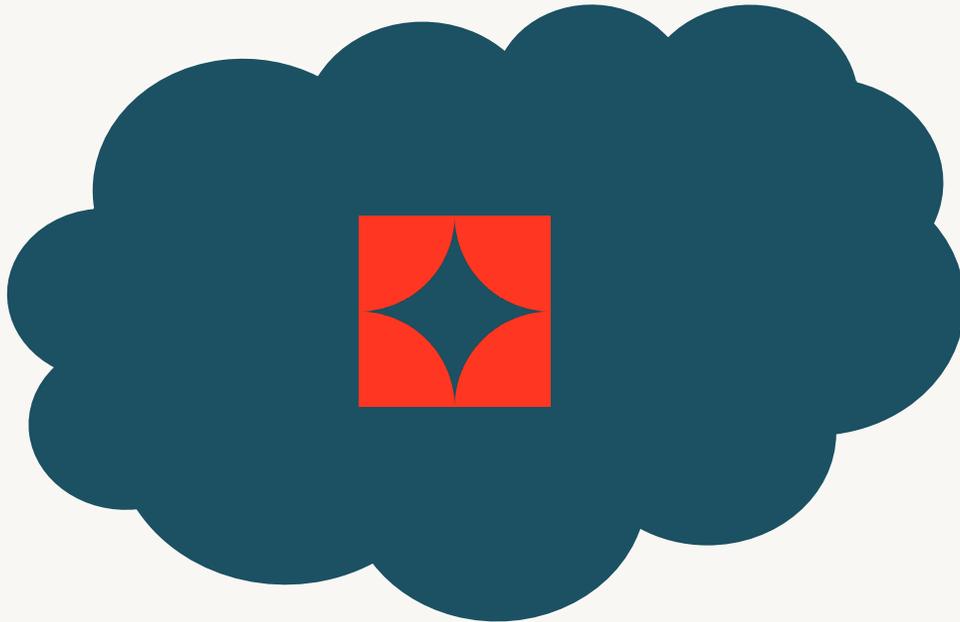
Secure, faster pace of innovation



Powered by Delta Sharing

How does AI model collaboration works?

An example

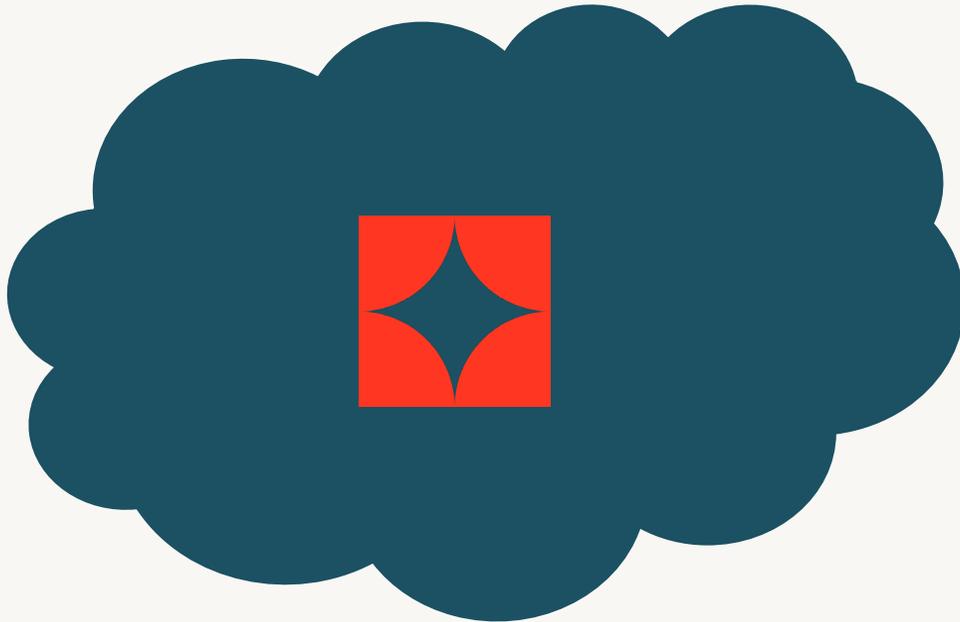


**FarmModel
(provider)**

FarmModel is a model provider the agriculture space

Model use cases: crop monitoring and soil analysis

FarmModel's needs



**FarmModel
(provider)**

Build and distribute both OSS and proprietary models

Distribute entire model packages and API endpoints

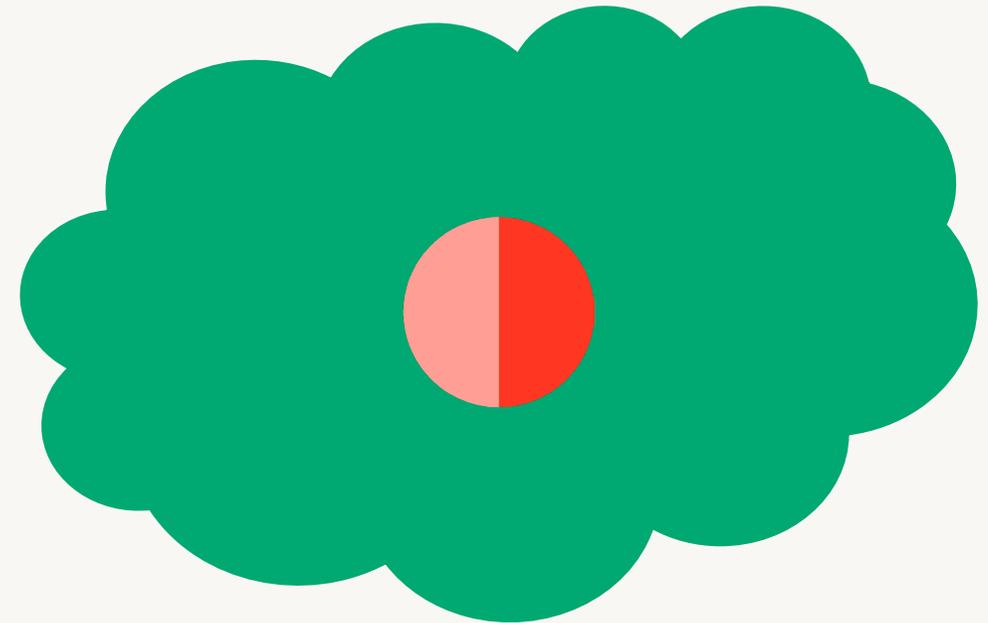
Want to meet customers where they are
- for privacy and easy of use

Want to acquire more customers

CropYieldAI – the consumer

Startup the agriculture space

Want to use satellite and drone images to monitor crop health, detect diseases, and assess yield potential.



CropYieldAI
(consumer)

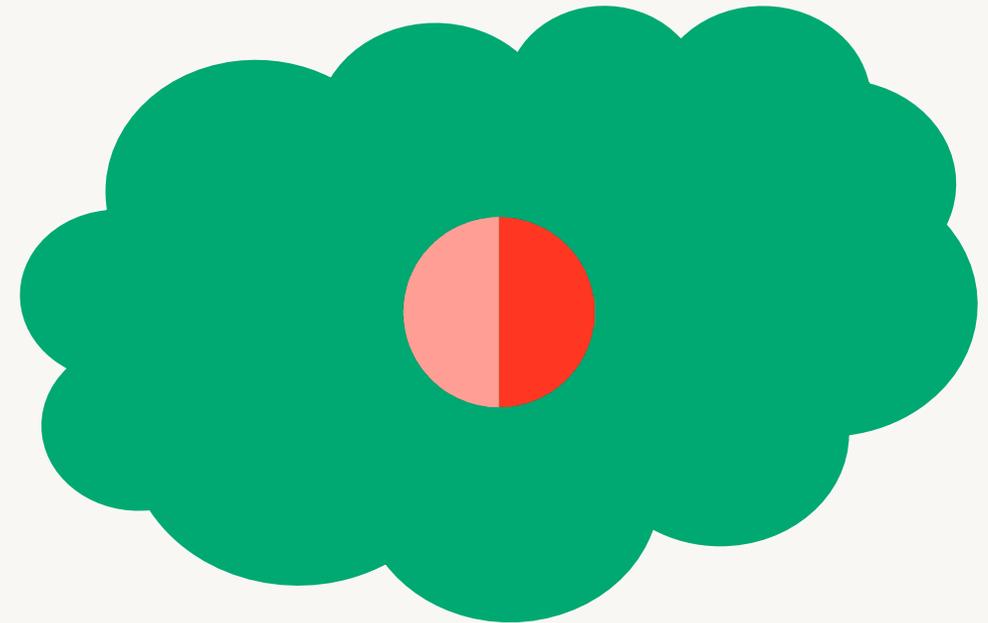
CropYieldAI's needs

Fine tune an existing OSS model for their use case – lower the cost of creating the model

Acquire fine tuning data from 3rd party provider

Cautious about internal data privacy

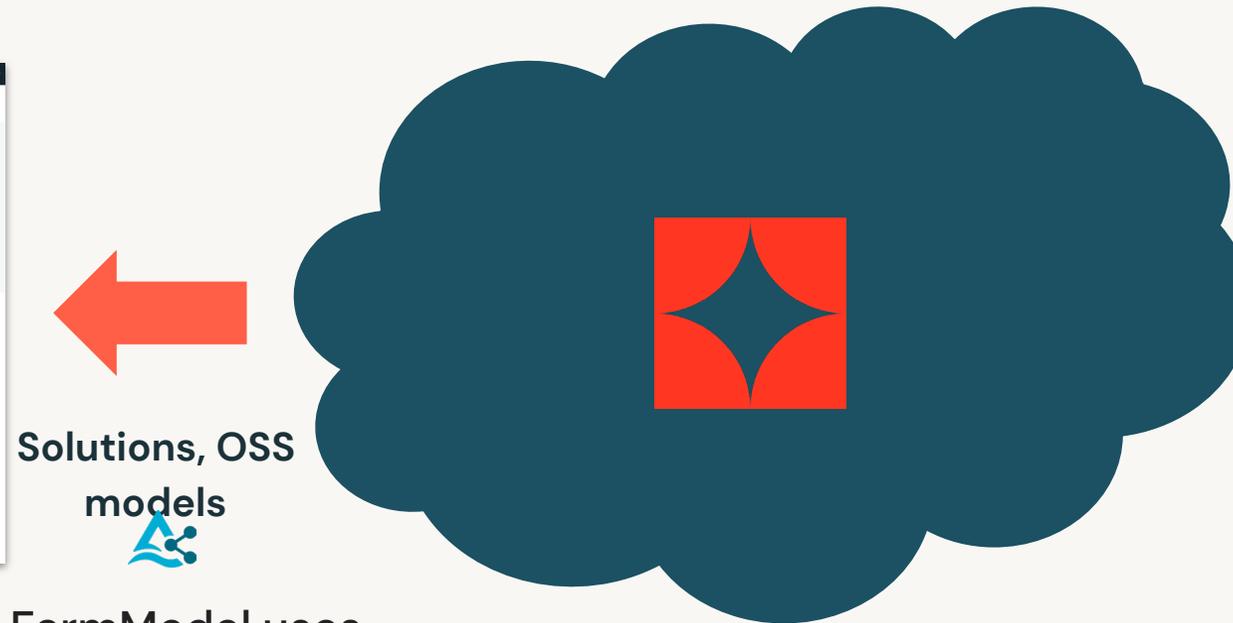
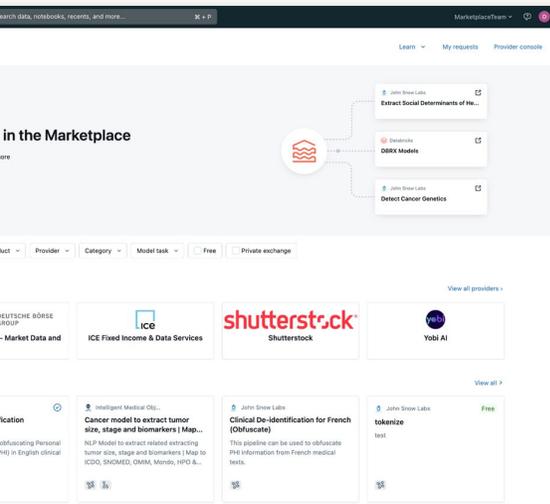
Use Databricks as the end-to-end, performant AI platform



CropYieldAI
(consumer)

FarmModel publishes content on Marketplace

CropYieldAI is a consumer of FarmModel's OSS model



FarmModel uses
Databricks
Marketplace to
distribute both open
models and solutions

Farm Model
(provider)

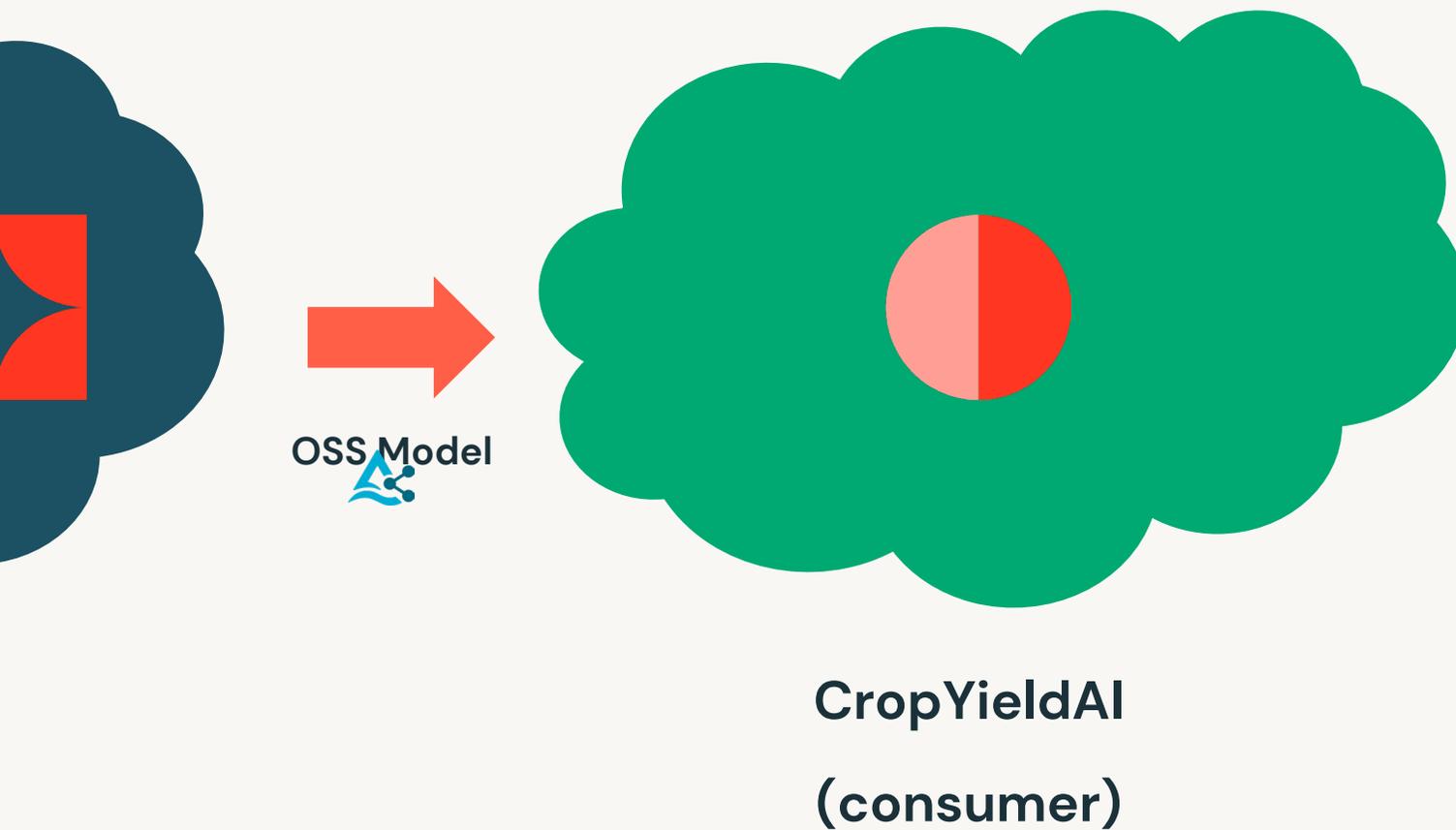
OSS Model


CropYieldAI
(consumer)

Databricks
Marketplace



CropYieldAI uses Databricks Mosaic AI Platform



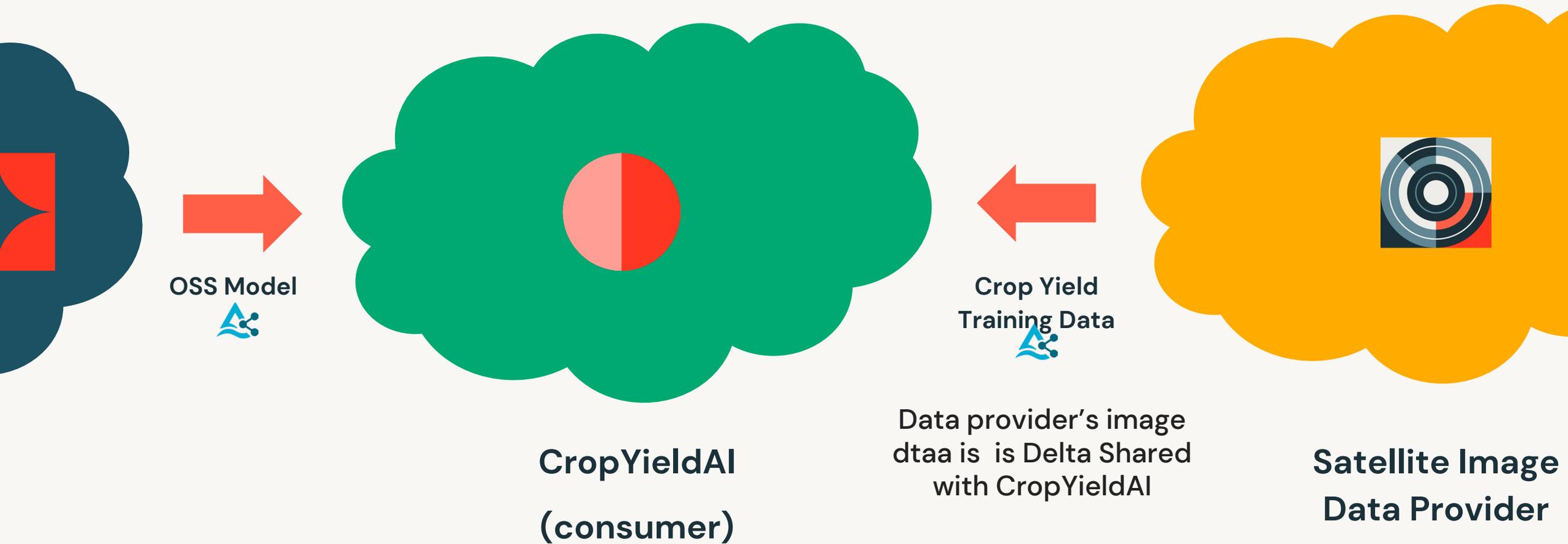
AI model works out of box with Databricks' suite of AI capabilities – batch or real time inference, use as-is or fine-tune

UC for governance to ensure high quality and control

Sensitive customer data stays with CropYieldAI

Move model to cloud/region with best GPU resources and pricing

CropYieldAI uses 3rd party image data



Benefits of AI model + asset collaboration

Lower cost



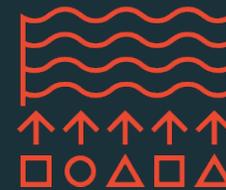
Reduce total cost of ownership of AI models: acquisition, development, infrastructure

Production quality



Acquire the model that fits your use case and augment with single platform for entire AI lifecycle

Complete control



Have complete control over the model and associated data



Share all assets: tables, volumes, AI models, notebooks



Share AI Models

OSS + proprietary models
Real-time and batch inference
Use as-is or fine-tune

Share Code

Share notebooks
Access Solutions Accelerators

/LiveRamp



Assets on Databricks Marketplace

Share All Data Types

Structured data
File-based data (e.g. images, videos, PDFs)

shutterstock



Databricks Marketplace powered by Delta Sharing

Open Marketplace for your data and AI needs



Databricks Marketplace

2000+ listings

320% YoY Growth in marketplace listings

300% YoY Growth in marketplace providers

John Snow Labs

Provides State-of-the-Art Medical Language Models

100+ million

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

O'Reilly Media

59% share

of Healthcare NLP teams
use John Snow Labs

Gradient Flow

#1 Accuracy

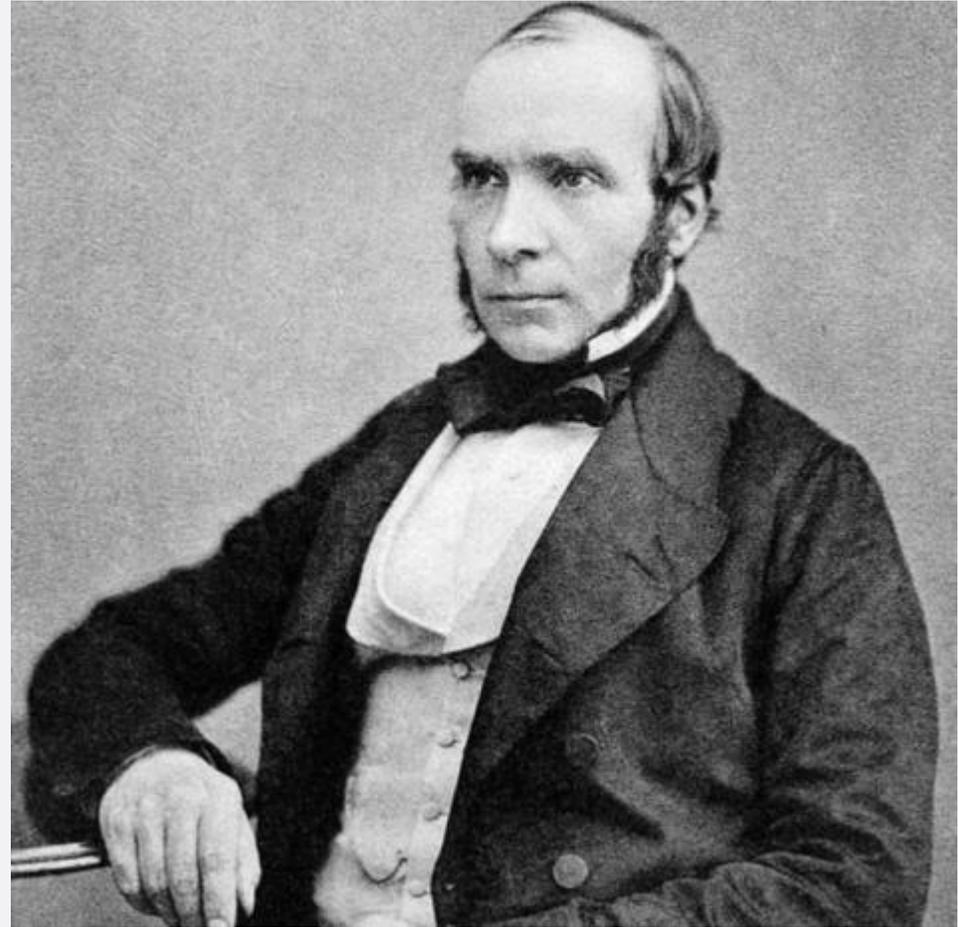
on 25 benchmarks in
peer-reviewed papers

Papers with Code

The Story of John Snow

Dr. John Snow was a physician in London during the reign of Queen Victoria.

In 1854, he played a key role in stopping a cholera outbreak which had killed hundreds of people within a few weeks.



Data Sharing... circa 1854

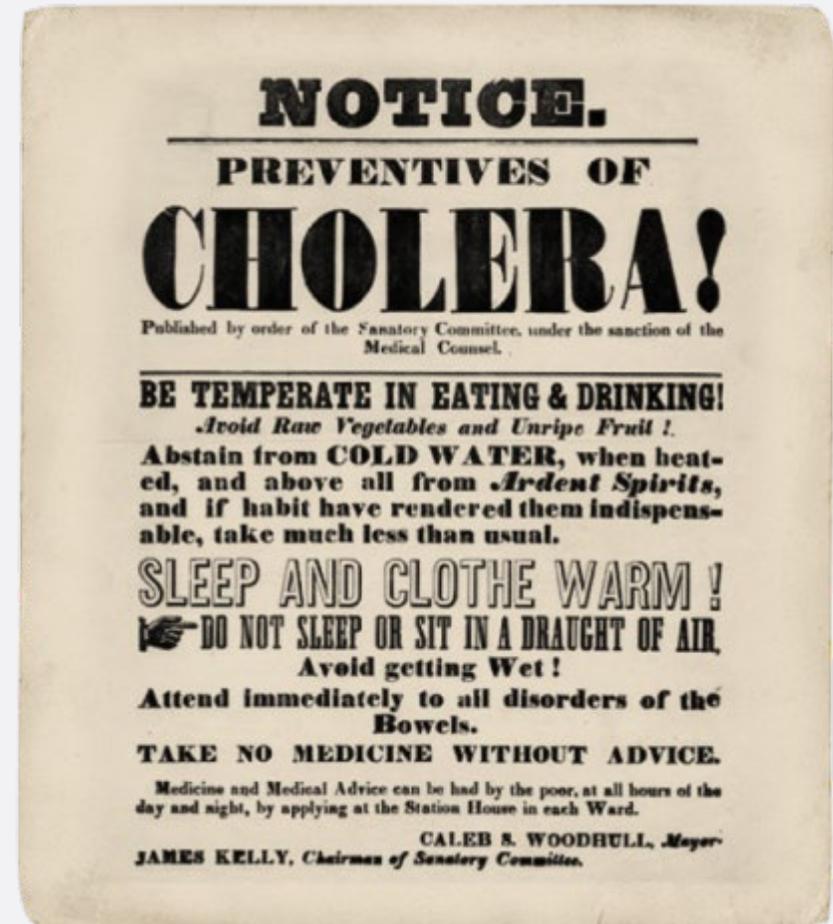


John Snow built a map of all the households where people died.

He concluded that the fault was one public water pump that all the victims had used.

The pump was taken out of service and the cholera outbreak subsided.

However, the big insight wasn't the pump but the fact that cholera was a waterborne disease.



John Snow Labs Case Studies



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



Accelerating Biomedical Innovation by Combining NLP and Knowledge Graphs



Extracting what, when, why, and how from Radiology Reports in Real World Data Projects



Automated Classification and Entity Extraction from Essential Clinical Trial Documents



Question Answering on Clinical Guidelines



Identifying opioid-related adverse events from unstructured text



Adverse Drug Event Detection using Spark NLP



Lessons Learned De-Identifying 700 Million Patients Notes with Spark NLP



Understand Patient Experience Journey to Improve Pharma Value Chain

2,400+ Expert Curated Datasets



Each dataset goes through 3 levels of quality review

- 2 Manual reviews are done by domain experts
- Then, an automated set of 60+ validations enforces every datum matches metadata & defined constraints



Data is normalized into one unified type system

- All dates, units, codes, currencies look the same
- All null values are normalized to the same value
- All dataset and field names are SQL and Hive compliant



Data and Metadata

- Data is available in both CSV and Apache Parquet format, optimized for high read performance on distributed Hadoop, Spark & MPP clusters
- Metadata is provided in the open Frictionless Data standard, and every field is normalized & validated



Data Updates

- Data updates support replace-on-update: outdated foreign keys are deprecated, not deleted

AI Models on the Databricks Marketplace

Databricks Marketplace
powered by  DELTA SHARING

Learn  My requests  Provider console 

Product: 1 

Provider: 1 

Category: 1 

Model task 

Free

Private exchange

[Reset filters](#)

60 filtered results

Sort: Name (A-Z) 

| | | | | |
|--|--|---|--|---|
| <p> John Snow Labs</p> <p>ICD-10-CM to UMLS Code Mapper</p> <p>This model expertly translates ICD-10-CM codes to UMLS (Unified Medical Language System) concepts.</p> <p></p> | <p> John Snow Labs </p> <p>Clinical De-identification (Mask)</p> <p>This model specializes in masking Personal Health Information (PHI) in English clinical notes.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification (Obfuscate)</p> <p>This model excels in obfuscating Personal Health Information (PHI) in English clinical notes.</p> <p></p> | <p> John Snow Labs Free</p> <p>Clinical De-identification for Arabic (Mask)</p> <p>This pipeline can be used to deidentify PHI information from Arabic medical texts. The PHI information will be masked i</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Arabic (Obfuscate)</p> <p>This pipeline can be used to obfuscate PHI information from Arabic medical texts.</p> <p></p> |
| <p> John Snow Labs </p> <p>Clinical De-identification for French (Mask)</p> <p>This pipeline can be used to de-identify PHI information from French medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for French (Obfuscate)</p> <p>This pipeline can be used to obfuscate PHI information from French medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for German (Mask)</p> <p>This pipeline can be used to deidentify PHI information from German medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for German (Obfuscate)</p> <p>This pipeline can be used to deidentify PHI information from German medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Italian (Mask)</p> <p>This pipeline can be used to de-identify PHI information from Italian medical texts.</p> <p></p> |
| <p> John Snow Labs</p> <p>Clinical De-identification for Italian (Obfuscate)</p> <p>This pipeline can be used to obfuscate PHI information from Italian medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Portuguese (Mask)</p> <p>This pipeline can be used to mask PHI information from Portuguese medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Portuguese (Obfuscate)</p> <p>This pipeline can be used to obfuscate PHI information from Portuguese medical texts.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Romanian (Mask)</p> <p>This pipeline can be used to de-identify PHI information from Romanian medical texts. The PHI information is masked.</p> <p></p> | <p> John Snow Labs</p> <p>Clinical De-identification for Romanian (Obfuscate)</p> <p>This pipeline can be used to obfuscate PHI information from Romanian medical texts.</p> <p></p> |
| <p> John Snow Labs</p> <p>Clinical De-identification for Spanish (Mask)</p> | <p> John Snow Labs</p> <p>Clinical De-identification for Spanish (Obfuscate)</p> | <p> John Snow Labs</p> <p>Clinical Text Summarization</p> <p>This LLM can be used to summarize clinical</p> | <p> John Snow Labs </p> <p>Detect Cancer Genetics</p> <p>Extracts biological and genetics entities</p> | <p> John Snow Labs</p> <p>Detect Drug Side Effect Narratives</p> <p>Classify health-related text in colloquial</p> |

Why Databricks Marketplace?



Fully Managed API



Click to Deploy

Pre-built NLP & LLM pipelines, optimized speed & latency, without a DevOps effort.



Managed Operations

Infrastructure, uptime, failover, security, scaling, monitoring, and updates are taken care of.



Private & Secure



Runs in your infrastructure

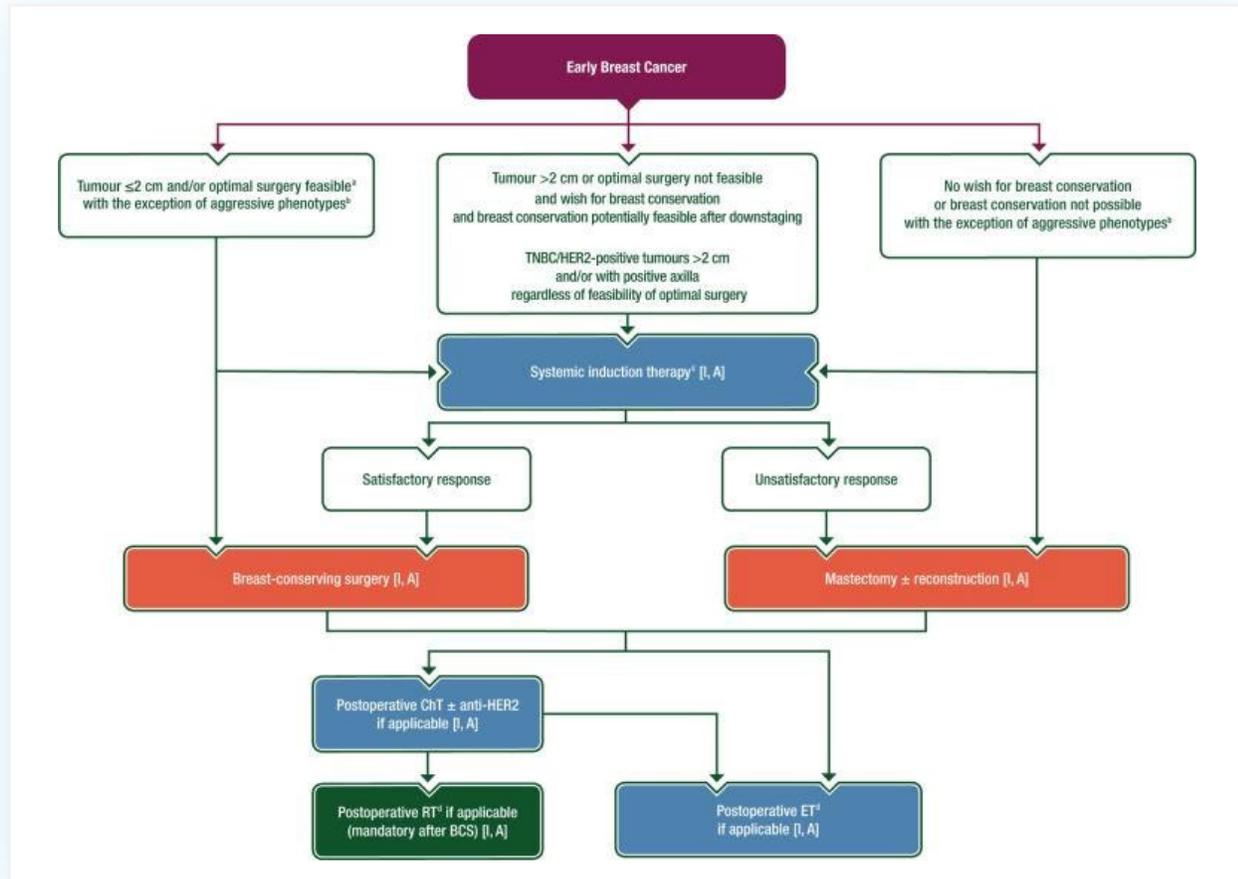
No data is sent outside your network, to John Snow Labs or anyone else.



Build and Keep your IP

Code, models, and prompts trained on your data is your IP. No one else will ever see it.

Case Study: Clinical Guidelines for Cancer Patients



- A clinical guideline is a document that guides decisions and criteria regarding diagnosis, management, and treatment.
- modern medical guidelines are based on an examination of current evidence within the paradigm of evidence-based medicine.
- They usually summarize consensus statements on best practice.
- A doctor is obliged to know the medical guidelines of their profession and must decide whether to follow them for an individual patient.

Where Is The Clinically Relevant Oncology Data?

Lab Reports

| TESTS | RESULTS | FLAG | UNITS | REFERENCE INTERVAL | LAB |
|--|---------|------|-------------|--------------------|-----|
| CBC With Differential/Platelet | | | | | |
| WBC | 5.1 | | x10E3/uL | 4.0 - 10.5 | 01 |
| RBC | 4.94 | | x10E6/uL | 4.10 - 5.60 | 01 |
| Hemoglobin | 15.1 | | g/dL | 12.5 - 17.0 | 01 |
| Hematocrit | 46.2 | | % | 36.0 - 50.0 | 01 |
| MCV | 94 | | fL | 80 - 98 | 01 |
| MCH | 35.6 | | pg | 27.0 - 34.0 | 01 |
| MCHC | 32.7 | | g/dL | 32.0 - 36.0 | 01 |
| RDW | 13.2 | | % | 11.7 - 15.0 | 01 |
| Platelets | 201 | | x10E3/uL | 140 - 415 | 01 |
| Neutrophils | 44 | | % | 40 - 74 | 01 |
| Lymphs | 44 | | % | 14 - 46 | 01 |
| Monocytes | 9 | | % | 4 - 13 | 01 |
| Eos | 3 | | % | 0 - 7 | 01 |
| Basos | 0 | | % | 0 - 3 | 01 |
| Neutrophils (Absolute) | 2.2 | | x10E3/uL | 1.8 - 7.8 | 01 |
| Lymphs (Absolute) | 2.3 | | x10E3/uL | 0.7 - 4.5 | 01 |
| Monocytes (Absolute) | 0.5 | | x10E3/uL | 0.1 - 1.0 | 01 |
| Eos (Absolute) | 0.1 | | x10E3/uL | 0.0 - 0.4 | 01 |
| Baso (Absolute) | 0.0 | | x10E3/uL | 0.0 - 0.2 | 01 |
| Immature Granulocytes | 0 | | % | 0 - 1 | 01 |
| Immature Grans (Abs) | 0.0 | | x10E3/uL | 0.0 - 0.1 | 01 |
| Comp. Metabolic Panel (14) | | | | | |
| Glucose, Serum | 95 | | mg/dL | 65 - 99 | 01 |
| BUN | 12 | | mg/dL | 5 - 26 | 01 |
| Creatinine, Serum | 1.02 | | mg/dL | 0.76 - 1.27 | 01 |
| eGFR | >59 | | mL/min/1.73 | >59 | 01 |
| eGFR AfricanAmerican | >59 | | mL/min/1.73 | >59 | 01 |
| Note: Persistent reduction for 3 months or more in an eGFR <60 mL/min/1.73 m2 defines CKD. Patients with eGFR values >=60 mL/min/1.73 m2 may also have CKD if evidence of persistent proteinuria is present. Additional information may be found at www.kdoqi.org . | | | | | |

Procedure Notes

General Surgery: Procedure Note Sample

Patient Name: Mary Jones **MRN:** 63342714808
Date of Birth: 2/14/1981 **Admit Type:** Inpatient
Age: 32 **Room:** OR 1
Gender: Female **Procedure Date:** 6/25/2013

Surgical Staff: Matt T. Johnson, MD
Referring MD: Alex Smith, MD
Procedure: Laparoscopic Cholecystectomy

Pre-OP Diagnosis: Chronic right upper quadrant and epigastric abdominal pain
 Cholelithiasis with chronic cholecystitis
Post-OP Diagnosis: Chronic right upper quadrant and epigastric abdominal pain
 Cholelithiasis with chronic cholecystitis

Patient Profile: The patient is a 32 year old female. The patient has symptoms of abdominal pain patient chart for documentation of history and physical. Previously obtained CT scan found in the gallbladder. The patient has failed previous conservative treatment. I recommended due to the patient's progressive symptoms. The alternatives, risks surgery were discussed with the patient. The patient verbalized understanding of the alternatives to surgery. The patient wished to proceed with operative intervention witnessed informed consent was placed on the chart. Prior to initiation of the procedure was performed: patient identification and proposed procedure were verified by the and the anesthesiologist, and the operative site was verified by the patient and the verification was performed in the pre-op area.

Anesthesia: General - Endotracheal
Findings: Gallbladder:
 - Thickened gallbladder wall.
 - Acute and chronic inflammation.
 - Multiple multilobulated green gallstones were seen.

Description of Procedure:
 Preoperative Medications / Therapy:
 - Ampicillin Sodium (Unasyn) 3 gm IV given prior to incision.
 - Knee high pneumatic compression stockings.
 Abdominal Prep and Drape:
 - The patient was placed on the standard operating table in the supine surgical position and compression were well padded. An OG tube was placed orally. The patient was sterilely prep povidone iodine solution (Betadine) and draped in the usual fashion.
 Laparoscope Insertion and Accessory Port Placement:
 - A 10 mL solution of 0.5% bupivacaine with epinephrine was infiltrated into the proposed incision site. A small puncture incision was made in the skin infraumbilical area and was carried down through subcutaneous tissue to the fascia. Bleeding was controlled with electrocautery. An incision was made in the peritoneum and 0 Vkril stay sutures were placed on both sides of the fascial incision. A 10 mm Hasson cannula was inserted through the opening into the peritoneal cavity and was fixed with stay sutures. The peritoneal cavity was then insufflated with CO2 to a pressure of 14 mmHg.
 - A 0 degree, 10 mm laparoscope was inserted through the port into the peritoneal cavity. If the peritoneal cavity revealed no evidence of bowel injury or bleeding.
 - Local anesthetic was infiltrated into the tissues at the proposed accessory port sites. Using incisions, one 10 mm port was placed subxiphoid area and two 5 mm ports were placed right midclavicular line and right subcostal anterior axillary line under laparoscopic vision.

SOAP Notes

SUBJECTIVE:
 The patient explained to me that he has been feeling much better since his last visit. When I asked the patient to rate his intensity on a scale of 0 to 10 with nothing and 10 being his most intense pain, Arthur gave his Right Sacroiliac Region as his last office visit. The activities that aggravate the patient's condition changed and include when he bends, sits, stands up, works out and runs. The patient also said his symptoms are improved when he gets adjusted and receives the combination of therapies at the office. The patient also informed me that he is doing all of the exercises he is supposed to do on a regular basis.

OBJECTIVE:
 Observation of the patient's active range of motion revealed decreased lumbar rotation with pain and right lateral flexion with pain. I observed spasms in Arthur's lower lumbar. I noted that the patient had moderate trigger points on his right piriformis. While palpating the patient, I noted severe tender taut fibers in the Right Sacroiliac Region. Orthopedic testing revealed Nachlas was positive, Right Test was positive, Laguerre's Test was positive and Hibb's Test was positive. Examination of the patient in the prone position, revealed a half of an inch functional short right leg length.

ASSESSMENT:
 The prognosis for the patient at this time is good because the patient is responsive to treatment.

PROCEDURE/PLAN:
 Eight minutes of continuous ultrasound at 1 Mhz was used to increase the blood flow to the muscle tonicity, and to decrease the discomfort over Arthur's Right Sacroiliac Region (97035). The Graston technique was applied for four minutes to the patient's Right Sacroiliac Region to improve muscle and ligament performance and possibly reduce any peripheral nerve impingement caused by adhesions (97110).

Pathology

Clinical History: Large Gastric Mass
Specimen: Gastric Mucosa
Diagnosis:
 Stomach, Partial Gastrectomy:
 - Malignant Epithelioid Gastrointestinal Stromal Tumor
 - Tumor Size 10 x 9 x 8 cm
 - Cell Type: Epithelioid and Spindled
 - High cellularity; present
 - Mucosal Invasion: Focally present adjacent to ulceration
 - Mucosal ulceration present
 - Mitotic Count: 10/50 HPF
 - Myxoid background: Focally present
 - Foci of necrosis present
 - CD117, vimentin, and CD34: uniformly positive

Gross Description
 The specimen consists of an approximately 5 x 7 cm portion of gastric mucosa that is lobulated mass which is 10 x 9 x 8 cm. The central portion of the mass appears to be the mucosa away from the area of ulceration is partially removed from the underlying appears encapsulated and lobular. Gross sections show the lesion to consist of several areas of gray to gray-tan pattern with an area of central necrosis showing a fairly uniform appearance with areas of gray white and somewhat lobular in appearance. Areas of yellow necrosis are scattered throughout the tumor. Representative portions submitted.

Microscopic Description
 Sections through the neoplasm show it to be primarily a high cellular neoplasm. The cells are in part arranged in fascicles and clusters with enlarged elongate nuclei having relatively fine nucleoli. In some areas, the fascicles have an interwoven appearance. Mitotic figure up to 10/50 HPF. A few areas show foci of necrosis with the cells appearing to be surrounded by somewhat myxoid stroma. Foci of displayed necrosis are present. The lesions appear circumscribed, although not specifically encapsulated. It focally involved the mucosa and shows full thickness ulceration. The tumor immediately beneath the mucosal area of ulceration has a nearly lobular somewhat spindled growth pattern. Some areas of the tumor have a slightly more rounded nuclei and somewhat epithelioid appearance. The cells appear to be arranged in groups and clusters. Some of the cells have cytoplasmic vacuoles. These areas also show a prominent mitotic activity. Some mitotic figures are abnormal and atypical. The tumor contains numerous relatively open vascular channels which appear to be part of the neoplasm. The tumor has a pseudo capsule and in some areas appear to be nearly covered. Immunostains are strongly positive for CD117 (C-kit), CD34, and Vimentin, Smooth muscle actin, Desmin, Synaptophysin, S-100, and Ck8/18 are negative.

Comment
 Immunostains were performed on the core biopsy and demonstrate that the tumor cells are positive for CD117. The findings are consistent with the above diagnosis.

Discharge Summaries

PHYSICIAN HOSPITAL DISCHARGE SUMMARY

Provider: Ken Cure, MD
Patient: Patient H Sample **Provider's Pt ID:** 6910828 **Sex:** Female
Attachment Control Number: XA728302

HOSPITAL DISCHARGE DX

- 174.8 Malignant neoplasm of female breast: Other specified sites of female breast
- 163.8 Other specified sites of pleura.

HOSPITAL DISCHARGE PROCEDURES

- 32650 Thoracoscopy with chest tube placement and pleurodesis.

HISTORY OF PRESENT ILLNESS

The patient is a very pleasant, 70-year-old female with a history of breast cancer that originated in the early 70's. At that time she had a radical mastectomy with postoperative radiotherapy. In the early 70's, she had a chest wall recurrence and was treated with further radiation therapy. She then went through many years until the late 80's when she developed bone metastases with involvement of the right humerus, and left sacral area. She was started on Tamoxifen at that point in time and has when she developed shortness of breath and was found to have a larger pleural effusion. Two occasions and has rapidly reaccumulated so she was admitted at this time for thoracoscopy note, her CA15-3 was 44 in the mid 90's and recently was found to be 600.

HOSPITAL DISCHARGE PHYSICAL FINDINGS

Physical examination at the time of admission revealed a thin, pleasant female in mild respiratory distress. She had decreased breath sounds three fourths of the way up on the right side, mostly clear although there were a few scattered rales. Cardiac examination revealed a regular rhythm without murmurs. She had no hepatosplenomegaly and no peripheral clubbing, cyanosis, edema.

HOSPITAL DISCHARGE STUDIES SUMMARY

A chest x-ray showed a large pleural effusion on the right.

HOSPITAL COURSE

The patient was admitted. A CT scan was performed which showed a possibility that the lung was involved and that there were some adhesions. The patient then underwent thoracoscopy which confirmed pleural peel of tumor and multiple adhesions which were taken down. Two chest tubes were placed. These were left in place for approximately four days after which a TALC slurry was infused and removed the following day. Because of the significant pleural peel and the trapped lungs, it is felt that pleurodesis will not be successful and this was explained to the patient and the family prior to the procedure.

Discharge Summaries

Where Is The Clinically Relevant Oncology Data?



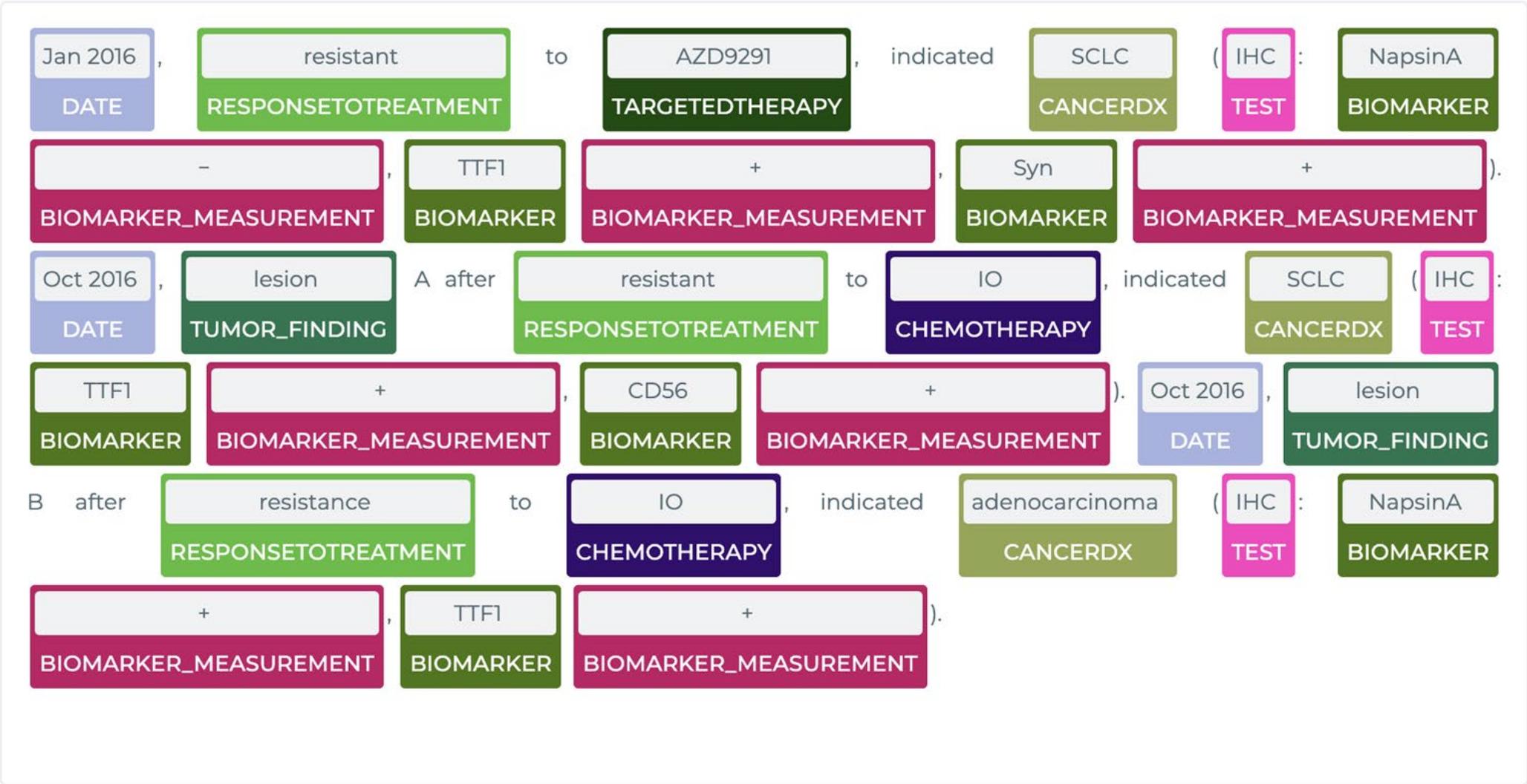
A detailed study found that **only 40%** of about 300 data points required for clinical decision support were available in structured data

Recognize 400+ Medical Entities

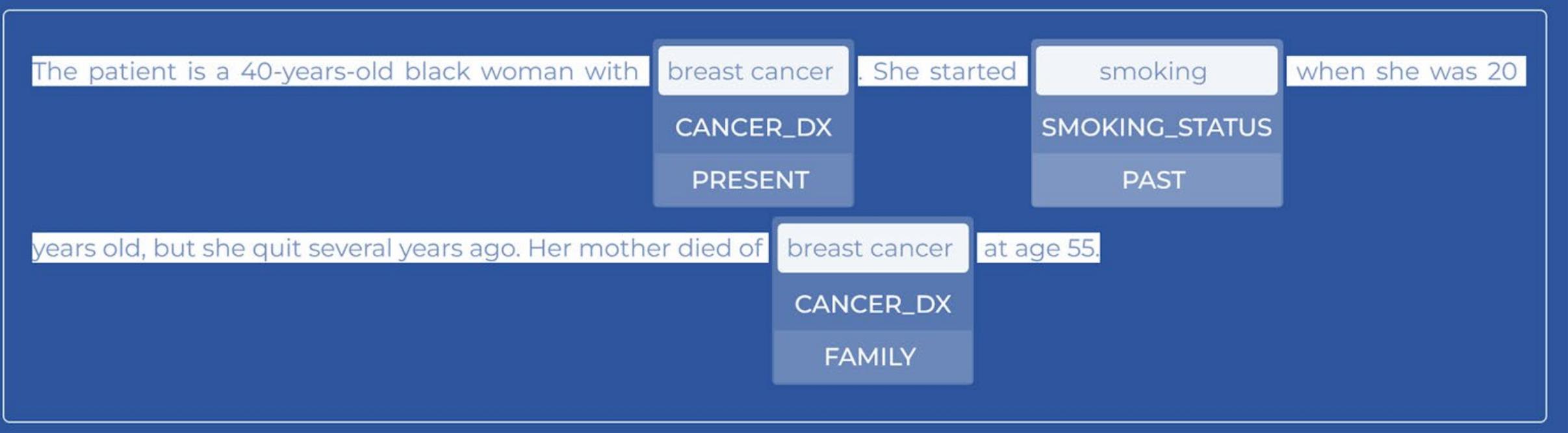
Text annotated with identified Named Entities

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

Tuned for Clinical, Biomedical, and Patient Voice Text

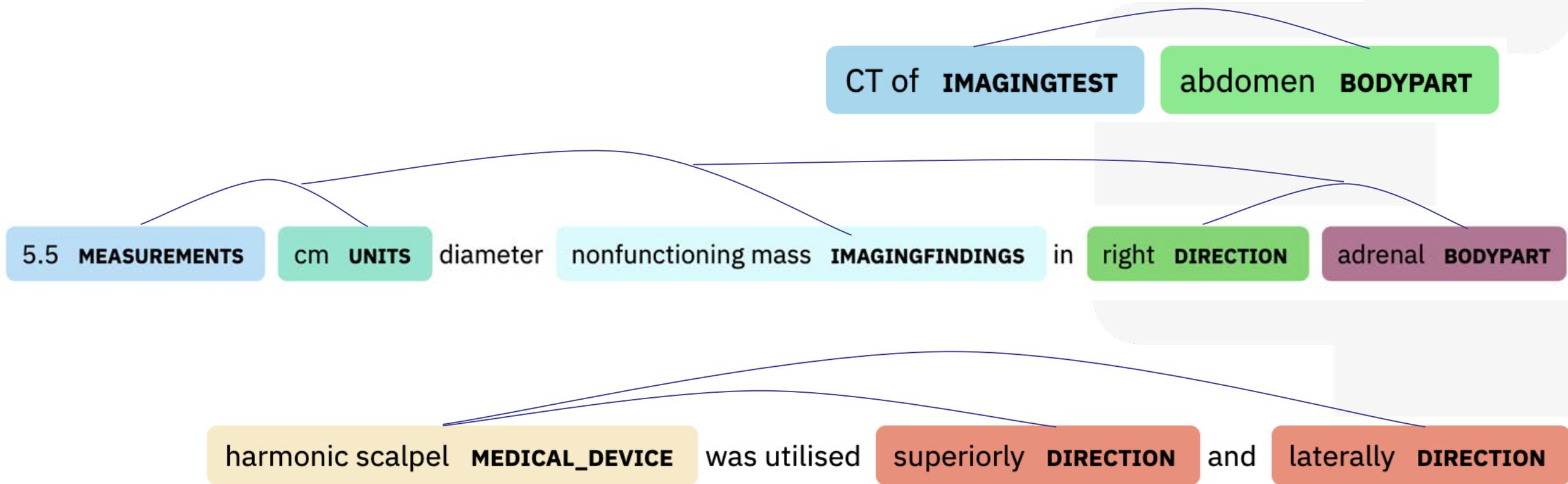


Understand Present vs. Absent vs. Possible, Past vs. Present, Patient vs. Someone Else

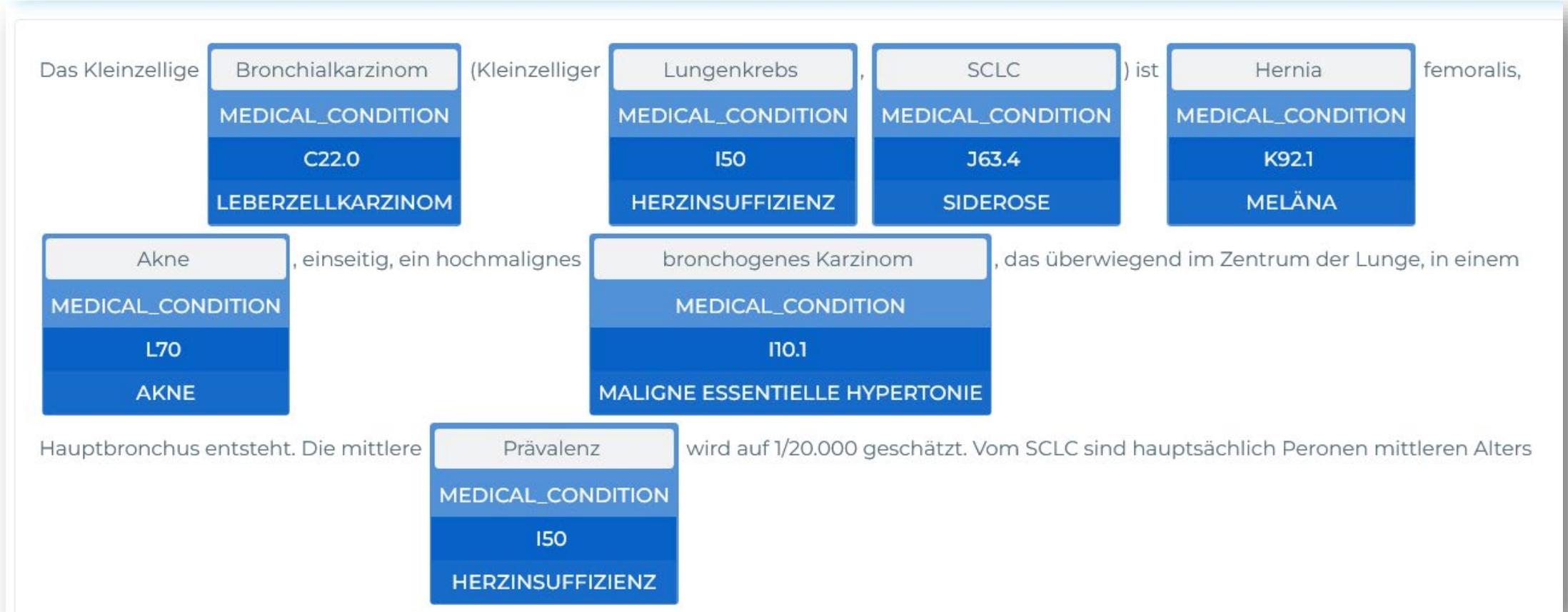


100+ Pre-Trained Medical Relation Types

"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."

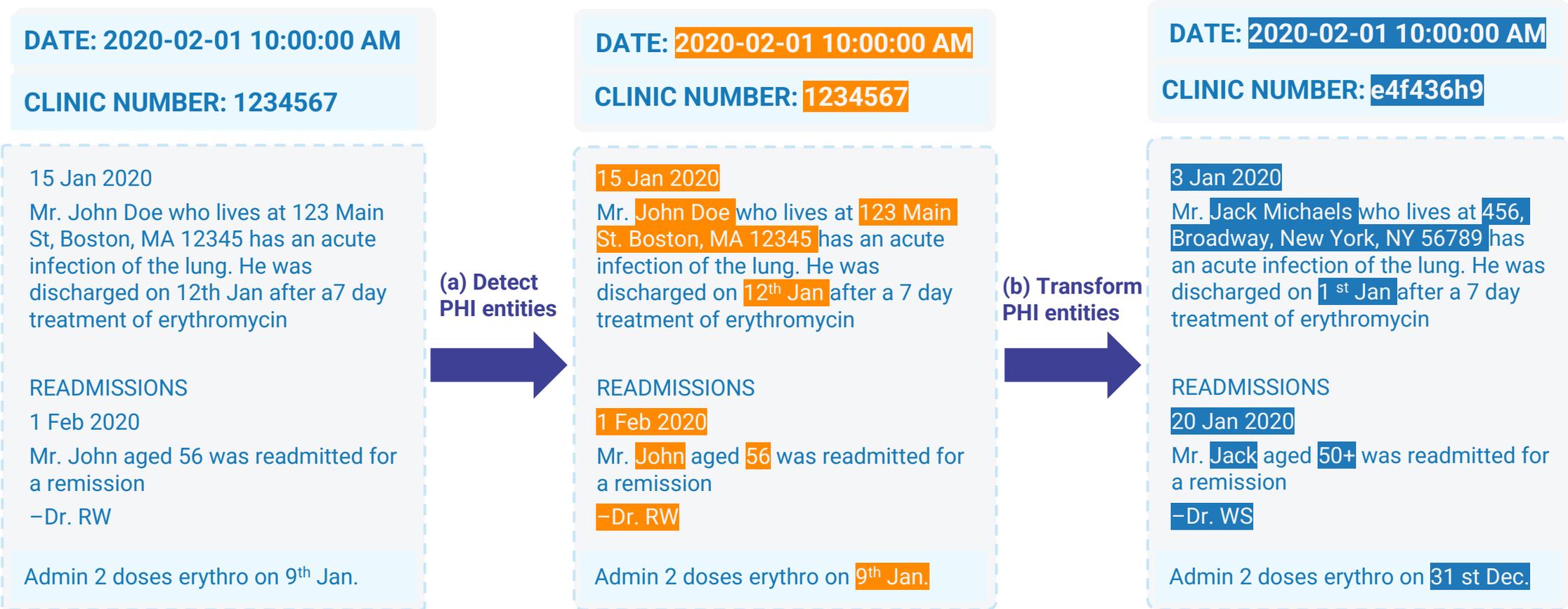


Multi-Lingual Support, 10+ Terminology Code Sets



Named entity recognition & ICD-10-GM code normalization in German

De-Identification of Medical Data



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

State-of-the-Art Medical LLM

| | |
|---------------|-------|
| JSL-MedLX-70B | 91.82 |
| Med-PaLM2 | 84.09 |
| GPT-4 | 82.97 |
| Llama3-FT-Med | 77.71 |

* on the Open Medical LLM Leaderboard Benchmark

MedQA (USMLE)

- 1,273 real-world questions from the US Medical License Exams (USMLE) to test general medical knowledge

PubMedQA

- 500 questions constructed to test reasoning over biomedical research texts, especially their quantitative contents

MedMCQA

- 4,183 questions from Indian medical entrance exams (AIIMS & NEET PG) spanning 2.4k healthcare topics, designed to address real-world medical entrance exam questions

MMLU

- College-level questions on Clinical knowledge (265), Medical genetics (100), Anatomy (135), Professional medicine (272), College biology (144), and College medicine (173)

Key Takeaways

- Databricks Marketplace lets use discover, share, and run fully managed AI models.
- Models are deployed and run within your infrastructure, critical for high-compliance industries like healthcare and finance.
- 60+ Medical Language Models are available to you today – and they get ongoing updates, accuracy improvements, and support.

