

Leveraging Healthcare NLP Models in Regulatory Grade Oncology Data Curation



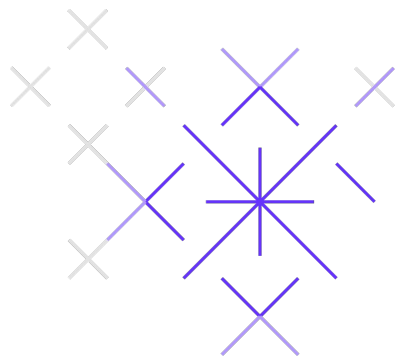
Sudhakar Velamoor

VP of Engineering
COTA



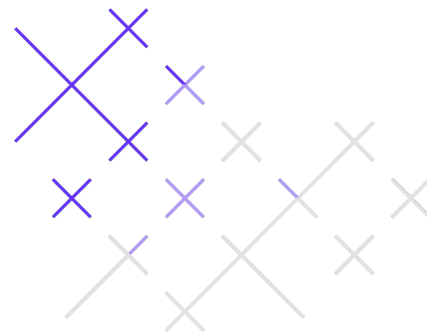
Brooke Gruman

VP of Product
COTA



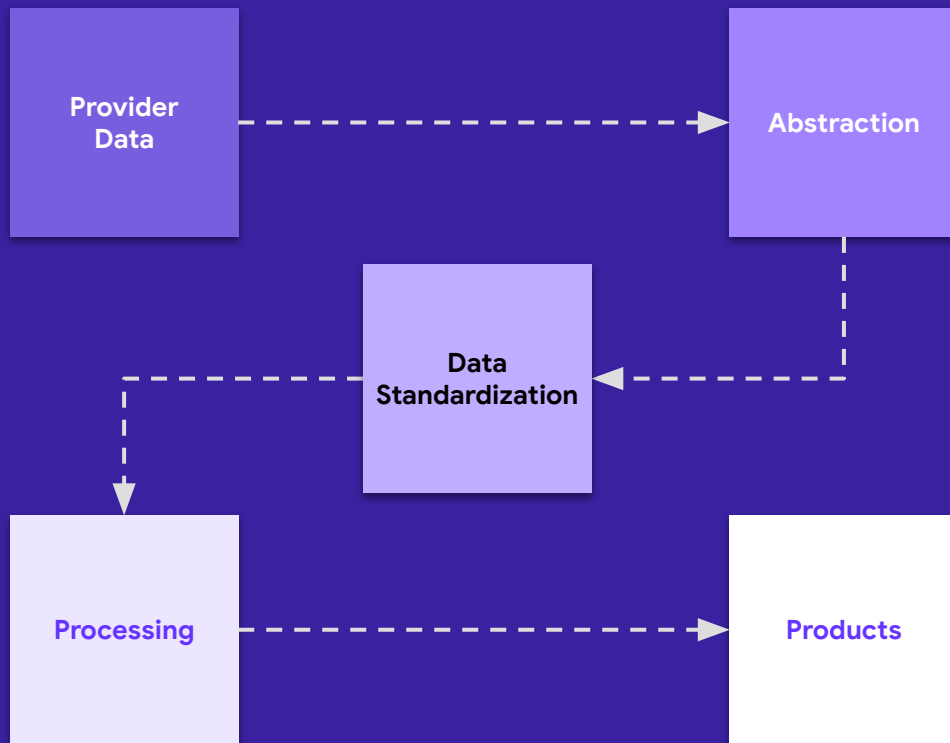
Bringing clarity to cancer

At COTA, we are working to ensure that
everyone touched by cancer
has a clear path to the right care.



How we do it

COTA organizes fragmented, often hidden data from the real world, through tech-assisted and enhanced human abstraction.

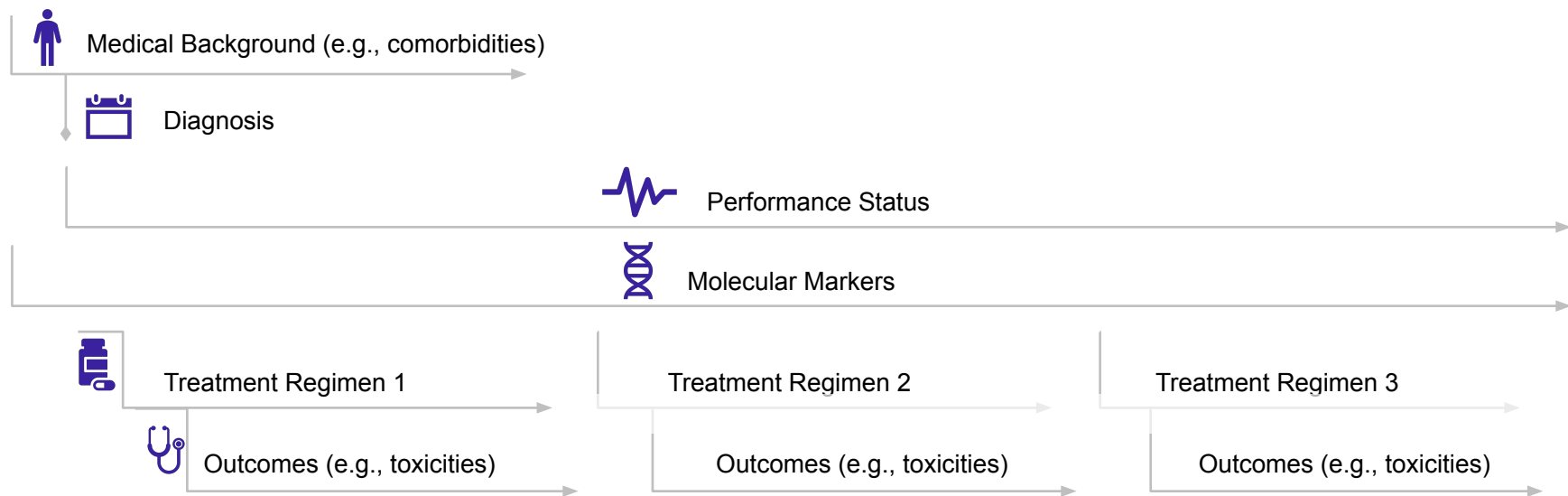


The COTA platform provides a clinically comprehensive and longitudinal patient journey for each specific oncology cohort






LONGITUDINAL PATIENT JOURNEY

Illustrative

COTA collects a range of data elements including...



RWD Clinical Data Abstraction Challenges

Challenge	Options			
 Manual Clinical Data Abstraction High clinical accuracy	 Structured EMR Data Fragmented, non-standard, maintenance heavy	 Search / Text Processing Rules based, somewhat unpredictable	 Generic Healthcare NLP Models Useful in some contexts	 Domain specific Pretrained Models Great for a start, but needs tuning

NLP Use cases explored



Unstructured Text

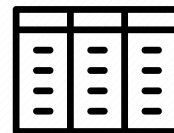
Highlighting relevant text for abstraction

Extraction of data for review / usage



Model Tuning & Active Learning

Annotation Lab
+ Feedback loop through Abstractors

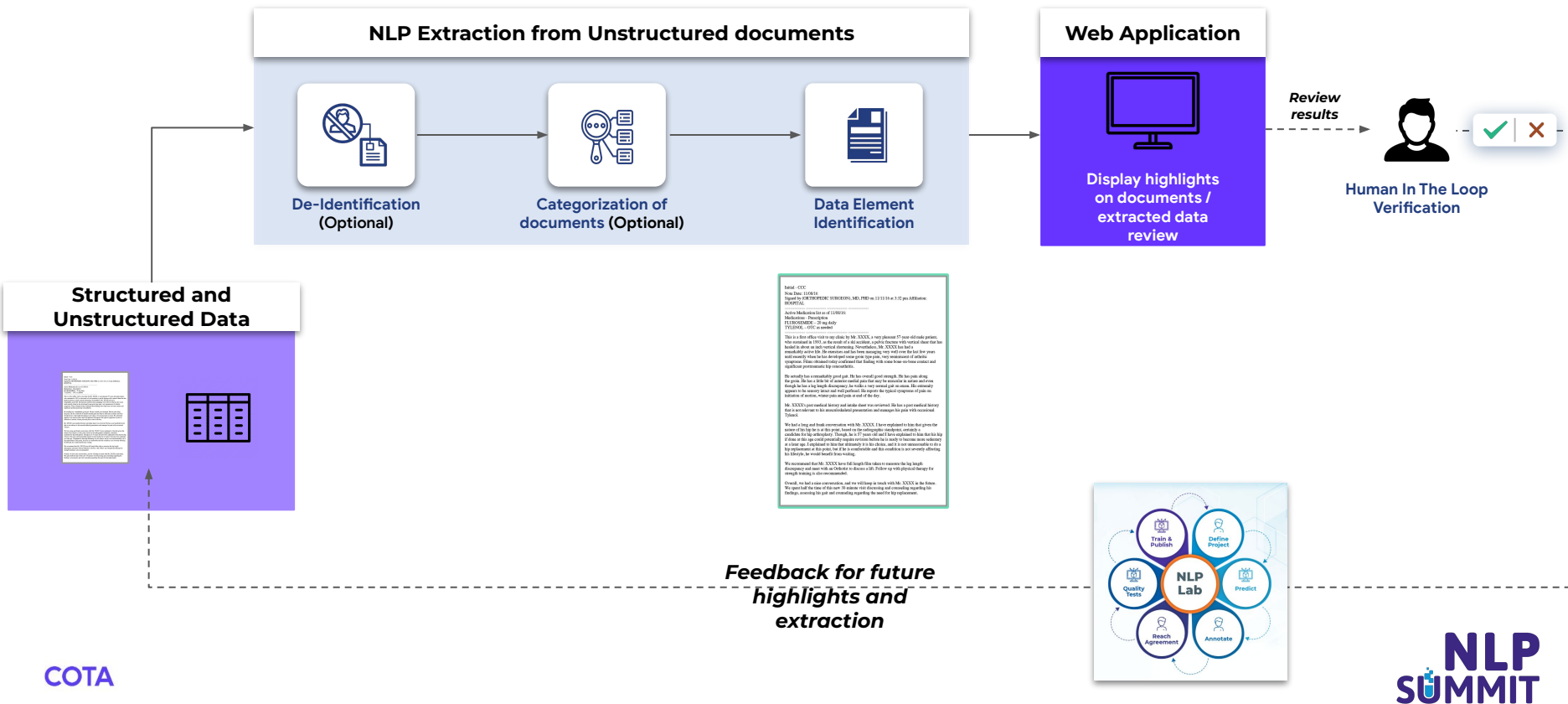


Structured Data

Extract normalized data for review from free text fields

De identify free text / structured text

Model Enhancements & Active Learning Process



Adverse Events

- Any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medical treatment or procedure. Also called a toxicity.

B-Symptoms

- Fever – Temperature $>38^{\circ}\text{C}$ ($>100.4^{\circ}\text{F}$)
- Weight loss – Unexplained loss of >10 percent of body weight over the past six months prior to diagnosis
- Sweats – The presence of drenching night sweats

Adverse Events

SUMMARY: In summary, Mr. [REDACTED] has CLL. He has been receiving fludarabine/rituximab. He has tolerated the first four cycles well. The red cell and platelet counts are normal. He remains fatigued.

The WBC has decreased; he has mild [REDACTED]. We will defer therapy today. I will see him again in 2 weeks; we plan to proceed with FR likely at a reduced dose when counts are adequate.

neutropenia
+ADVERSE EVENT

We again discussed the treatment plan and possible side effects/toxicity.

B-Symptoms

IMPRESSIONS:

1. Extensive marrow infiltration for which the differential diagnosis reported includes "myeloproliferative disorder (chronic anemia, myelofibrosis, leukemia or lymphoma) or marrow infiltrating metastatic disease"

2. [REDACTED] Bilateral lower extremity pain

Weight loss, chills, and night sweats\n3.
+BSYMPTOMS

B-Symptoms Model Details

- Detect Lymphoma, B-Symptoms, Dates of Assessment
- Initial accuracy - somewhat lower for practical use cases
- Clinical expertise required to train the model
- Additional training increased accuracy by 20%
- Increased accuracy can be achieved with more training and data points

Challenges

Side effects of treatment explained at length including cytopenias, neuropathy, fever, and fatigue.
+BSYMPTOMS +ADVERSE EVENT -ADVERSE EVENT

Hem/Onc Diagnosis Non-Hodgkin's lymphoma : Stage at diagnosis: IVB
+ADVERSE EVENT

Review Of Systems

General: As above.

Head, Eyes: Negative.

ENT: Negative.

Cardiovascular: Negative
+ADVERSE EVENT

Pulmonary: Negative
+ADVERSE EVENT

GI: Negative
-ADVERSE EVENT

GU: Negative
-ADVERSE EVENT

Musculoskeletal: Negative
+ADVERSE EVENT

Dermatologic: Negative
+ADVERSE EVENT

Neurologic: Negative
-ADVERSE EVENT

Extremities: Negative
-ADVERSE EVENT

Future Use: Outcomes

- An outcome is a measure of the overall response to systemic treatment
- Ranges from CR (Complete Response) to Progression/Relapse

5/1/2019 PET scan—favorable response to therapy. v PR decreased size and resolution of activity in mesenteric lymphadenopathy and decreased activity and bilateral lung hila. Mesenteric nodal mass 20 x 17 mm, SUV 1.6, previously 62 x 44 mm. Nodular airspace opacity in the lingula of the left lung stable in size, increase activity. Lingular nodule currently 19 x 11 mm, SUV 3.4.

Summary

Challenges / Learning:

- Challenges persist when trying to identify healthcare related data within unstructured text
 - Oncology is even more challenging given that treatments are personalized
- Creating completely new models from scratch to solve data extraction is possible but requires in-house expertise in understanding how best to operationalize the models
 - Data for training is often sparse in Oncology

Positives:

- NLP models can help with extraction of data at scale, with quality, as long as there's a good review / learning process
- Time savings for Human Abstraction can be realized in targeted data extraction scenarios by highlighting as well as review of extracted data.
- Models can be improved with data labeling and active learning
 - Tooling is essential for improving accuracy of models
- Multiple models and pipelines are often needed, but there are lots of choices out there.



Thank you!

NLP SUMMIT HEALTHCARE

www.nlpsummit.org

Presented by

