




Natural Language Processing



Alec Chapman
MIMI34MD
June 2020



About Me

- **Education:** M.S. Biomedical Informatics, University of Utah
- **Current job:** Data Architect at University of Utah / US Department of Veterans Affairs
- **Professional Interests:** Data Science for epidemiological and clinical applications



Overview of Clinical NLP

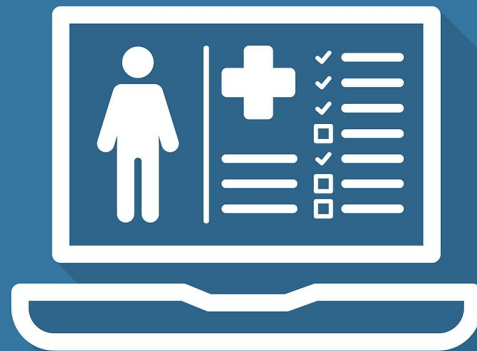
What is NLP?

Subdomain of Artificial Intelligence (AI)

Extract information from unstructured text

Limitations of structured data

- Incomplete
- Lack detail
- Can be inaccurate



Valuable information contained in free-text notes

- Rich, detailed narratives
- Treatment information and context

However, this data is **unstructured**

Admission Date: **01/01/2019**
01/02/2019

Discharge Date:

Date of Birth: 05/31/1947

Sex: **F** **MRN: 12345**

Attending: Dr. Brooke Shields, MD

Chief Complaint: **Chest pain, sob**

History of Present Illness:

Briefly this is a **72 yo f** w/ **h/o Asthma** and **high cholesterol**, complains of **sob** and **chest pain**. Symptoms **consistent w/ PE**. **D-Dimer** came back **positive**. CTPA confirmed diagnosis of **pulmonary embolism** in **right lung**. administered **80 units/kg IV heparin**, clotting back to normal. Prescribed **warfarin** for px.

Past Medical History: **Asthma, HCL, stroke**

Physical Exam: ...

Signed electronically by: DR. Brooke Shields MD 01/02/2019
11:34 AM

Real-World Applications

- **Research:**

- Cohort selection for clinical studies
- Phenotyping

- **Epidemiology:**

- Biosurveillance
- Case detection
- Clinical + social media data

- **Administrative**

- Billing

How I Use NLP

COVID-19 Surveillance

- Operational surveillance in US Department of Veterans Affairs (VA)
- Case detection of patients without lab evidence
 - "The patient tested positive for **COVID-19** at an outside laboratory."
 - "Ms. X called to inform us that she was diagnosed with **novel coronavirus**."
- Used to identify $\sim\frac{1}{3}$ of cases in national VA
- Manuscript (OpenReview): https://openreview.net/pdf?id=ZQ_HvBxcdCv

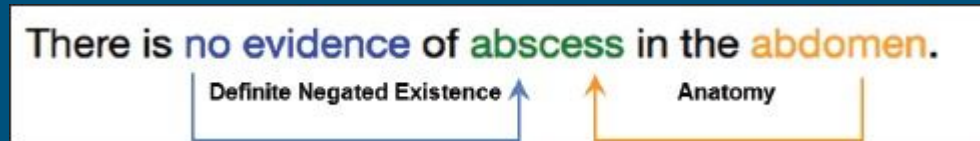
Patient presents for possible UNCERTAIN COVID-19 CORONAVIRUS . His wife OTHER_EXPERIENCER recently tested positive
DEFINITE_POSITIVE_EXISTENCE for novel coronavirus CORONAVIRUS .

Veteran Housing Status

- Identify whether a patient is stably housed
 - "The patient is **currently homeless**."
 - "She received an **eviction notice** yesterday."
 - "He **got the keys** to her **new apartment** and is doing well."
- Socioeconomic data not reliably stored in EHR
- Evaluate effectiveness of housing programs

Evidence of SSI from Radiology Reports

- Identify whether a patient developed infection following a procedure
- Connect to body location
- Patient safety and quality of care
- Manuscript: <https://pubmed.ncbi.nlm.nih.gov/29854116/>



Workshop



Clinical NLP with spaCy