Leveraging YouTube Video Analytics for Health Literacy: An Augmented-Intelligence Machine Learning Pipeline to Retrieve and Recommend Videos on Chronic and Infectious Diseases

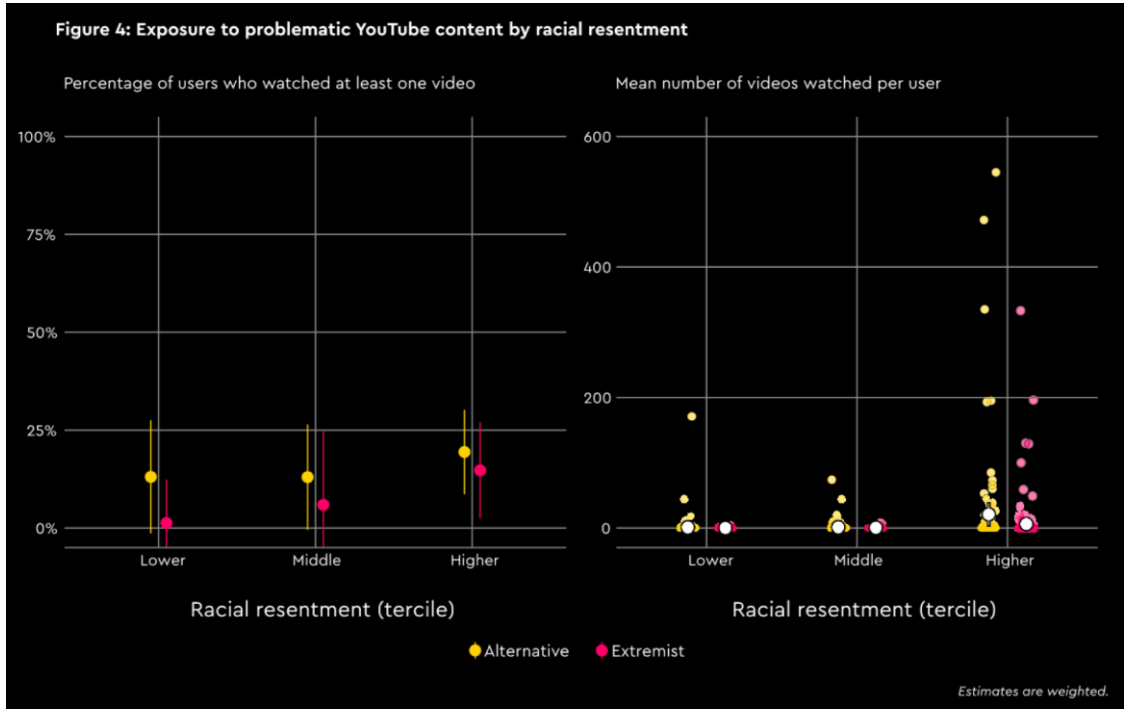
Rema Padman, PhD, rpadman@cmu.edu, Carnegie Mellon University, Pittsburgh, PA; Xiao Liu, PhD, Xiao.Liu.10@asu.edu, Arizona State University, Tempe, AZ; Anjana Susarla, PhD, asusarla@broad.msu.edu, Michigan State University, Lansing, MI.

Goal: Develop an automated, scalable and generalizable recommendation system to identify and retrieve YouTube videos that have medically-relevant, understandable, actionable and accurate content for improving ***health education, patient engagement and public health literacy***

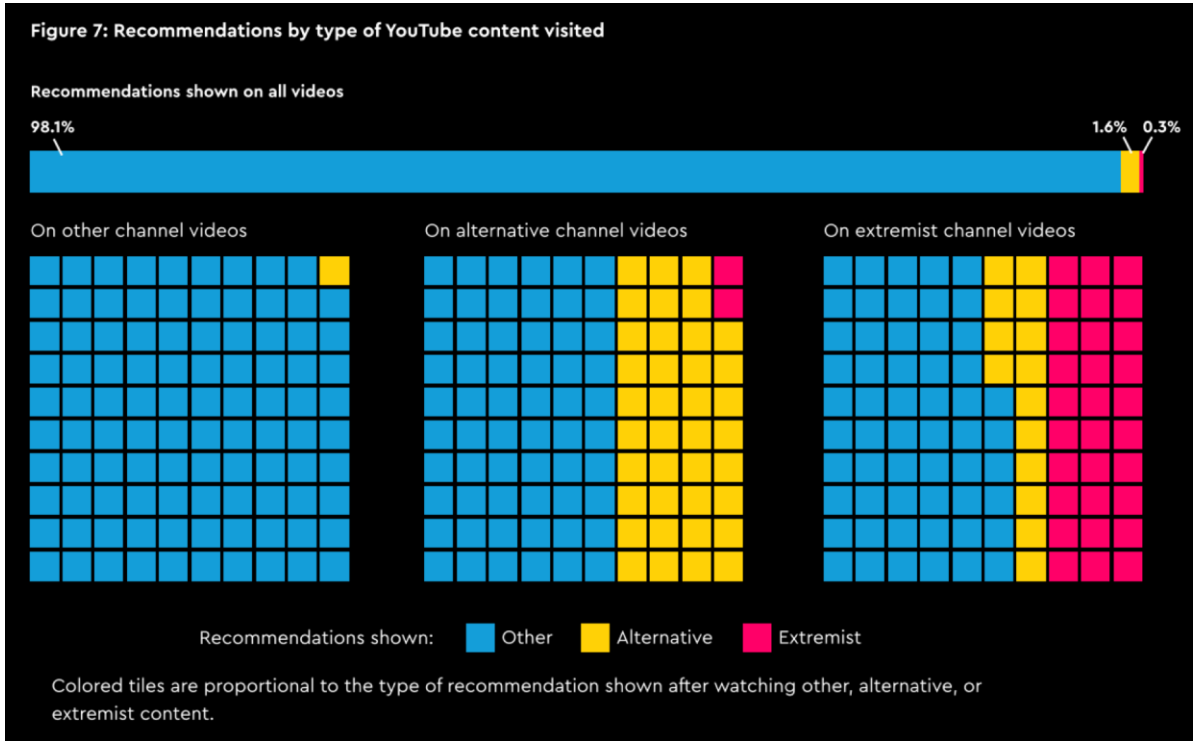
Health literacy is critical for improving health outcomes and reducing healthcare costs – low health literacy is a serious societal challenge!

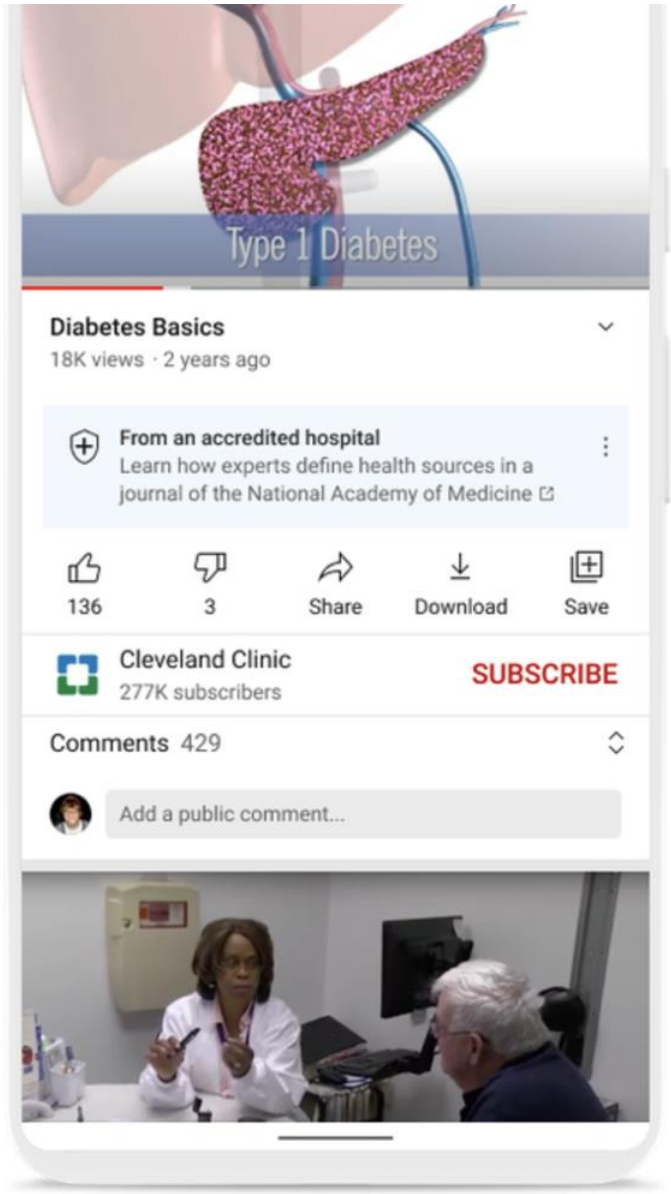
- Leverage the vast repository of computable biomedical and human-centered knowledge in health-related videos on the YouTube platform
- Use AI, optimization, machine learning and natural language processing to identify, retrieve and curate videos that clinicians can prescribe and public can access efficiently

Identifying credible information on YouTube

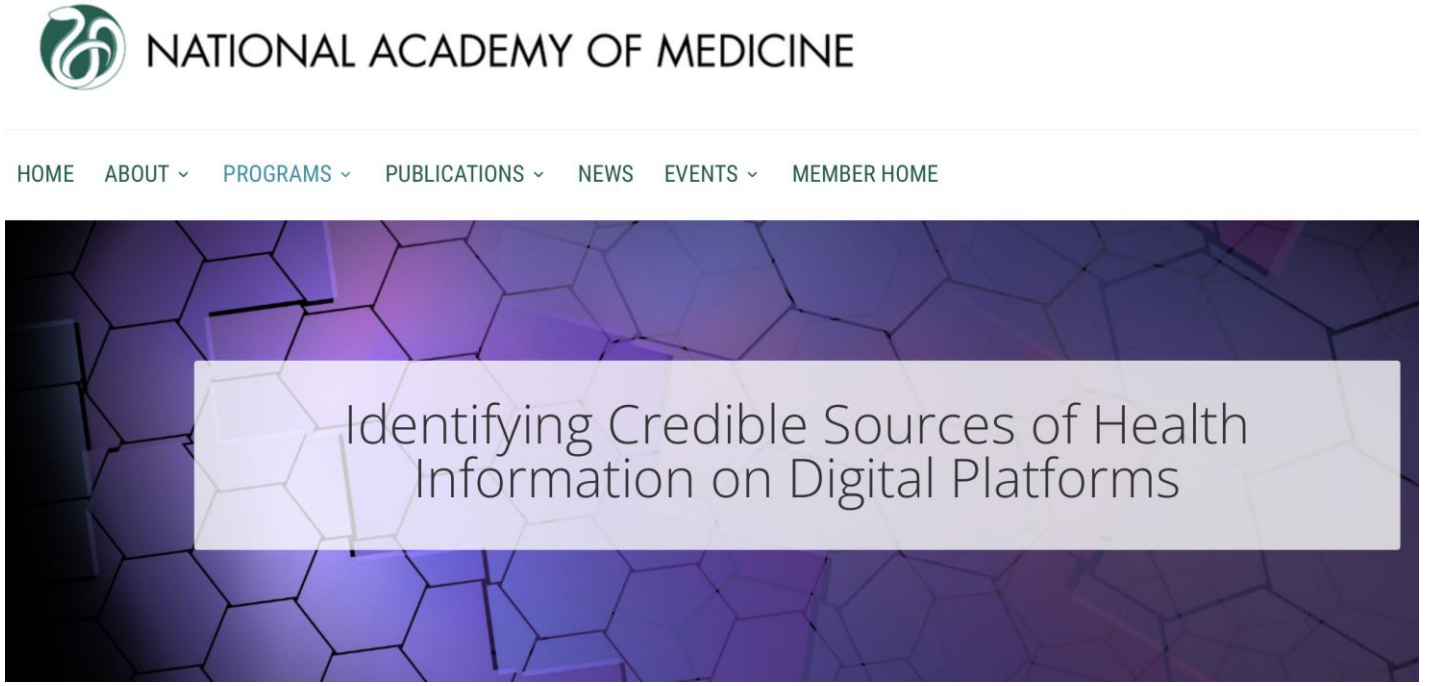


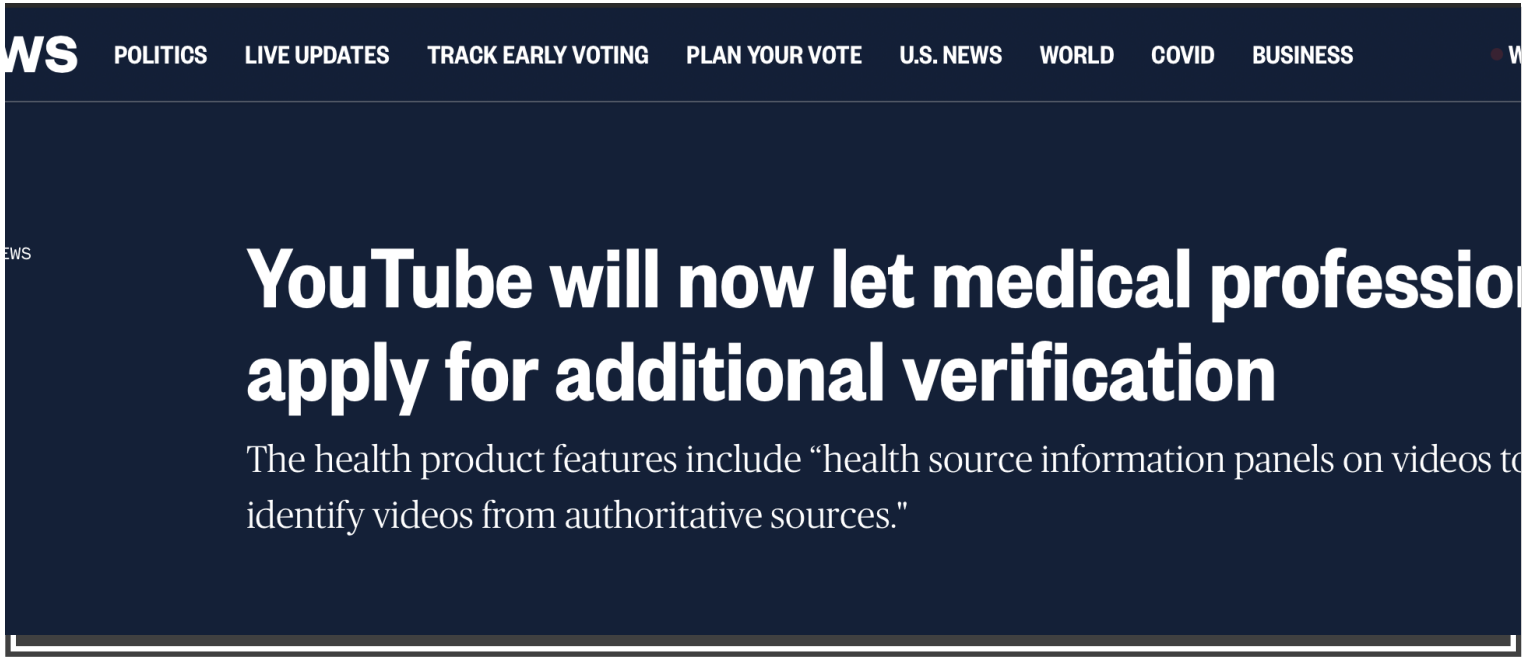
(Chen, Nyhan, Reifler, Robertson, and Wilson 2021)



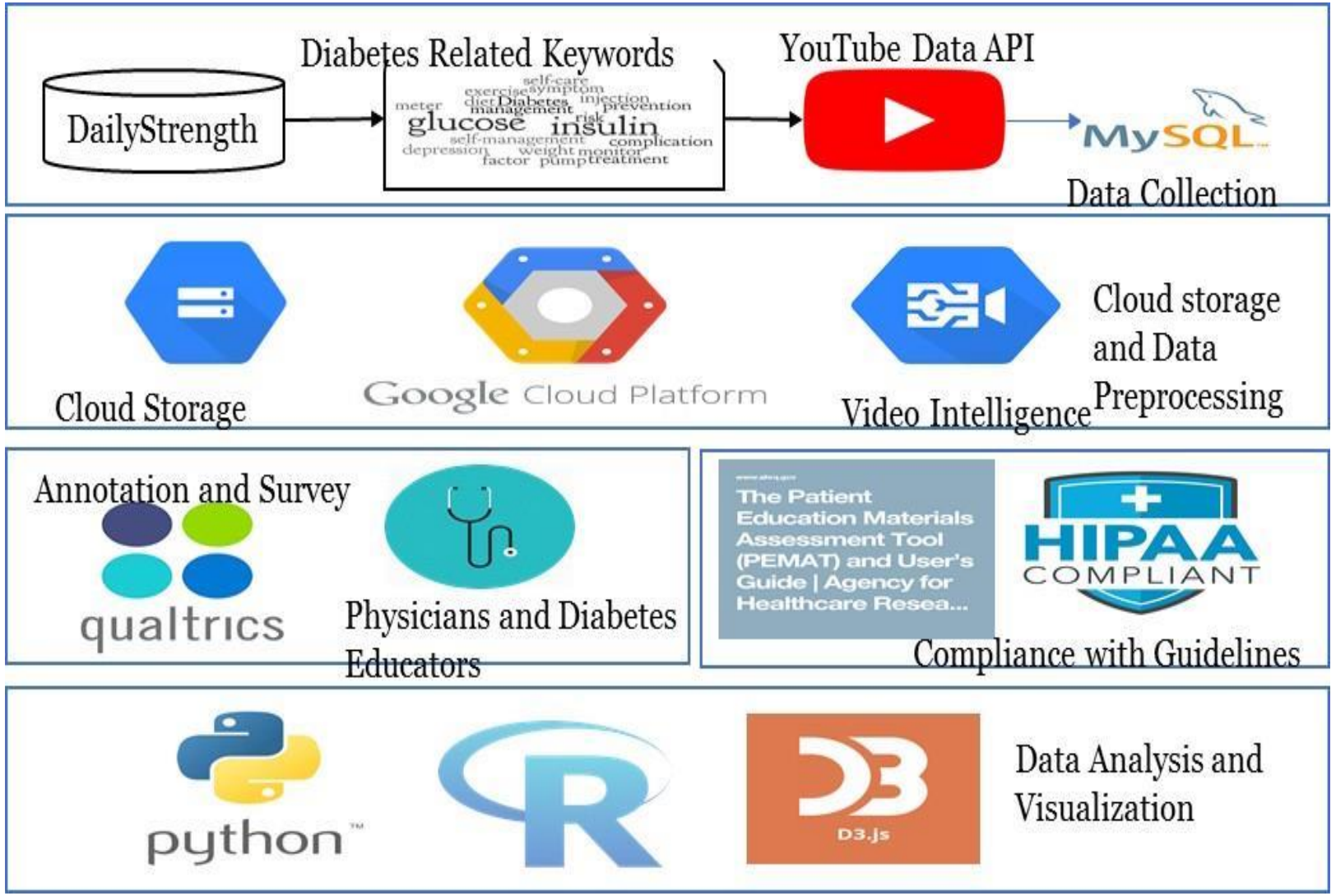


YouTube announced that health professionals can apply to have their accounts labeled an authoritative source. YouTube

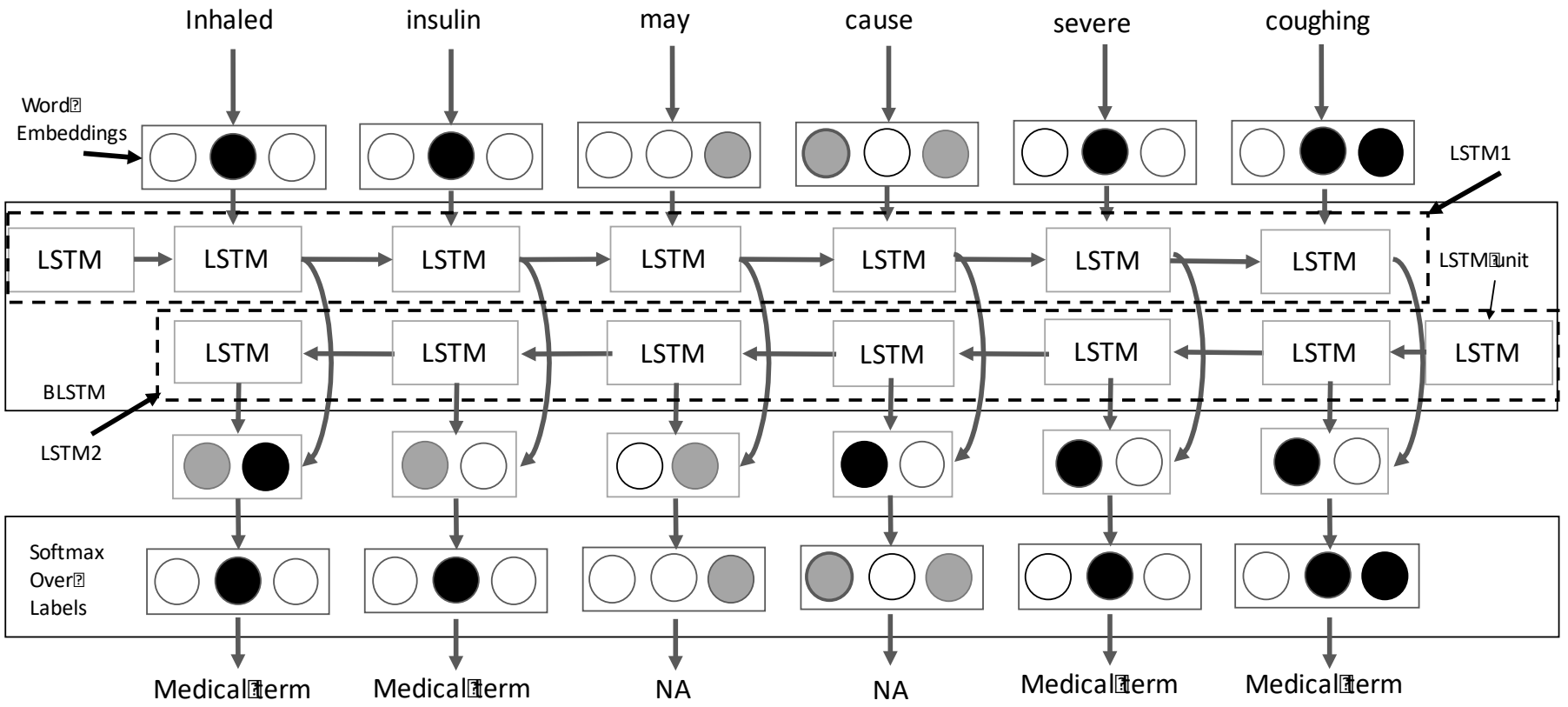




ML Pipeline


DailyStrength
Diabetes Related Keywords
glucose
insulin
YouTube Data API
MySQL
Data Collection
Cloud Storage
Google Cloud Platform
Video Intelligence
Cloud storage and Data Preprocessing
Annotation and Survey
qualtrics
Physicians and Diabetes Educators
The Patient Education Materials Assessment Tool (PEMAT) and User's Guide | Agency for Healthcare Resea...
HIPAA COMPLIANT
Compliance with Guidelines
python
D3.js
Data Analysis and Visualization

Identifying Medical Terminology in YouTube Video Description



- Medical terms
  - Disease
  - Treatment
  - Symptom
  - Condition
  - Procedure
  - Component/location
- Writing styles
  - Standard medical terminology
  - Consumer health vocabulary

Patient Educational Materials Assessment

- Evaluation of patient educational videos relied on the judgment of domain experts on several critical dimensions (Backinger et al. 2011)
  - Content understandability by end users (Ruppert et al. 2017)
  - The volume of medical information (Liu et al. 2019)
- Agency for Healthcare Research and Quality (AHRQ) proposed the Patient Education Materials Assessment Tool (PEMAT) (Shoemaker et al. 2014)
  - Evaluate and compare patient education materials in written, audio and video formats
  - PEMAT highlights the need to emphasize the understandability of patient educational materials

A video is understandable when

- Consumers of diverse backgrounds and varying levels of health literacy can process and explain key messages

Contributions to practice