Maternal Health Equity Workshop: From Story to Data to Action
Bianca’s Story: Turning Pain into Action

Bianca Dickerson-Williams, JD
Founder and Executive Director, Fighting Our Injustices for Women of Color
Fighting our Injustices for Women of Color
Background

This story needs to be heard to save the lives of Black women, Black children, and Black men.


Background of my story. (Misconception about who and the stereotypes)

Law Enforcement and Lawyer.

Factual Details of the case and what happened in the delivery room.
Statistics: Department of Health and Center for Disease Control

The United States has the highest maternal mortality rate amongst developed countries.

Black infants are twice as likely to die in their first year of life than White infants.

In California, Black women are six times more likely to die due to complications within the first year of pregnancy than White women.

In the United States, Black women are three times more likely to die from pregnancy complications.

Black women have a 70% increased risk for severe maternal morbidity.

Indigenous women have the next highest morbidity and maternal health issues.
Police brutality and healthcare brutality

Why do they conceal (insurance, credibility, stain on their career, medical license suspended or revoked).

Doctors are more important than us and they have been put on a pedestal by society.

Police body cameras versus no other checks and balances for physicians. Their story is the story regardless if it is true or not.

Abuse of power: “whatever it is they say happened, that is what happened”
Corruption

Concealment of issues due to a lack of documentation, altering the truth, or omitting the truth.

Altered medical records, deletions, and omissions/failure to truthfully document in police reports and medical records

Doctors cover for other doctors just like police cover for other police. It is a systemic abuse of power.
Call to action

Mommibus Legislation

Women of Color in Policy Coalition

U.S. Congress (Making laws to stop the problem)

Legal Matters

The White House

Documentary “Bianca’s Story: Black Moms At Risk”

Youtube and Fightingourinjustices.com
The AAMC Center for Health Justice Welcomes You to the Maternal Health Equity Workshop: From Story to Data to Action

Philip M. Alberti, PhD
Senior Director, Health Equity Research & Policy
Founding Director, AAMC Center for Health Justice

Carla S. Alvarado, PhD, MPH
Director of Research, AAMC Center for Health Justice
Agenda

10:20 a.m.  From Story to Data: Understanding Natural Language Processing
10:50 a.m.  From Data to Research: NLP & Maternal Health Use Cases
11:45 a.m.  Break
12:00 p.m.  Beyond Buzzwords: Reimagining the Default Settings of Technology & Society
12:35 p.m.  Exercise: Principles of Trustworthy NLP use in Maternal Health
2:00 p.m.  From Data to Action: What Public Health, Hospitals and Health Systems Can Do
2:50 p.m.  Drawing Change Workshop Summary

*All times are ET
Introduction to NLP

*from story to data to action*

Maria Antoniak, PhD
Allen Institute for AI

*AAMC Maternal Health Equity Workshop, May 18th 2023*
Welcome!

During this session, we’re going to cover the basics of NLP: methods, models, and challenges.

NLP comes with both big risks and exciting potential.

You can better understand how to weigh these risks and benefits if you know how NLP works “under the hood.”

I’m excited to learn with you today!
Our goals

1. Overview of different methods and workflows
2. Breadth not depth; the other sessions will go deeper
3. Understand the assumptions underlying these models
4. Awareness of how text turns into numbers used by NLP systems
About me

I’ve been working in NLP for a decade both in academia and industry.

Right now, I’m a postdoc at the Allen Institute for AI. We’re a research nonprofit, and our goal is to study “AI for the common good.”

**I have a lot of worries about AI but also a lot of hope.** I want to open up AI so that as many different people and perspectives are included as possible.

Teaching you the basics of NLP is part of that goal! **We need your expertise.**
Overview
What is natural language processing?

NLP, or computational linguistics, uses computational methods to study human language.

This can include analyzing human language and generating human language.

Methods can include statistics, machine learning, linguistics, and programming.

Examples:

- Google Search
- Alexa
- Email spam filters
- Autocomplete
- ChatGPT
Why are people excited about NLP?

1. Human language is fascinating!

2. We can do so many different things using NLP!
   - question answering
   - measuring biases
   - extracting information
   - ... and more!

   *To see lots of exciting examples, make sure you attend all of the talks today!*

3. Recent advances makes more things possible than most of us expected!
Why are people wary or critical of NLP?

Models can encode biases that are difficult to measure or correct.

Many systems rely on poorly documented training data.

Models are very large and training them takes up a lot of resources.

The processes leading to results can be hard to interpret.

For more on this topic, make sure you attend Dr. Ruha Benjamin’s talk later today!
From story to data
Computers don’t understand words, but they do understand numbers!

How can we turn words into numbers?
“You shall know a word by the company it keeps.”

The distributional hypothesis, popularized by Firth (1957).
“You shall know a word by the company it keeps.”

In other words: Which words often appear together?

We can learn about the meaning of a word by studying its usage.

In NLP, we often use a word’s usage patterns as a proxy for its semantic relationships to other words.
Definition Alert

**Vector**
A list of numbers

8 32 0 0 3 99

**Matrix**
A list of lists

8 32 0 0 3 99
74 0 5 45 0 0
3 26 0 5 0 92
0 0 16 0 64 23
Turning words into numbers

Use the distributional hypothesis!

Create a **matrix** where each cell contains the number of times the words occur in the same sentence.

<table>
<thead>
<tr>
<th></th>
<th>computer</th>
<th>technology</th>
<th>maternal</th>
<th>baby</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>computer</strong></td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>technology</strong></td>
<td>5</td>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>maternal</strong></td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td><strong>baby</strong></td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>13</td>
</tr>
</tbody>
</table>
Turning words into numbers

Each row of the matrix is a vector that represents the word.

\[
\text{baby} = [10, 5, 2, 0]
\]

<table>
<thead>
<tr>
<th></th>
<th>computer</th>
<th>technology</th>
<th>maternal</th>
<th>baby</th>
</tr>
</thead>
<tbody>
<tr>
<td>computer</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>technology</td>
<td>5</td>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>maternal</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>baby</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>13</td>
</tr>
</tbody>
</table>
What can we do with those numbers?

We can plot the words in a graph.

And then we can measure distances and relationships between words!

\[
baby = [10, 5, 2, 0]
\]
What can we do with those numbers?

For whole texts, we can average all the word vectors to get a single vector for the text.

Now we can measure distances and relationships between the texts.
Word Vectors (or Word Embeddings)

There are many different ways to create these word vectors.

Latent semantic analysis (LSA) is a simple method from 2004.

In 2013, **word2vec** was introduced, which used a neural method to discover the set of vectors.
Where do we get all those word counts?

These models take advantage of **big datasets** and **pretraining** on a large dataset and applying that model to a smaller task.

Datasets can include:

- Wikipedia
- Reddit
- books
- EHR
- scientific publications
Contextualized vectors

Word2vec and similar methods produce static vectors, where each word is represented by a single vector.

Newer methods create a new vector for each time a word is used in a dataset. These are contextualized vectors.

These are some of the methods enabling new tools like ChatGPT.
Static Vectors: one point per vocabulary item
Contextualized Vectors: one point per instance of a word in context!
made rhyme an art

the nature of philosophie

against nature can persuade

imparting science, or celestial truth

what lessons science waits to teach

what he calls religion

theirs is a mixt religion

when art was sacred

made rhyme an art

thou art to me a fly

art thou not prone
“Extempore Effusion upon the Death of James Hogg”

By William Wordsworth

When first, descending from the moorlands,
I saw the Stream of Yarrow glide
Along a bare and open valley,
The Ettrick Shepherd was my guide.

When last along its banks I wandered,
Through groves that had begun to shed
when first, descending from the moor lands, I saw the stream of the arrow glide along a bare and open valley. The etrick shepherd was my guide. When last along its banks I wandered, through groves that had begun to shed their golden leaves upon the pathways, my steps the border-mind the minstrel led. The mighty minstrel breathe no longer, I mid mo undring ruins low lies; and death upon the braess of the arrow, has closed the shepherd-poet's eyes: nor has the rolling year twice measured, from sign to sign, its steed's ast course, since every mortal power of Cole's ridge was frozen at its marvelous source; the rapacious one, of the godlike forehead, the heaven-eyed creature sleeps in earth: and lamb, the frisky and gentle, has vanished from his lonely hearth. Like clouds that rake the mountain, summit, or waves that own no curb, how fast has brother followed brother, from sunshine to the sunless land! Yet I, whose lids from infant slumber were earlier raised, remain to hear a timid voice, that asks in whispers: "Who next will drop and disappear?" Our haughty life is crowned with darkness, like London with its own black wreath; on which with thee, o crab! Forth- looking, I gazed from Hampstead's breezy heath. As if but yesterday departed, thou too art gone before; but why, o'er ripe fruit, seasonably gathered, should frail survivors have a sigh? Mo urn rather for that holy spirit, sweet as the spring, as ocean deep; for her who, er her summer faded, has sunk into a breathless sleep. No more of old romantic sorrow's, for slaughtered youth or love-lovern maid! With sharpe grief is the arrow stream, and et etrick mo urn with her their poet dead.
From data to model
Supervised vs Unsupervised Machine Learning

**Supervised**: texts have labels, model learns patterns from **texts + labels**

<table>
<thead>
<tr>
<th>DATA</th>
<th>LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>The final score between the Patriots and the Seahawks...</td>
<td>Sports</td>
</tr>
<tr>
<td>The politician’s fiscal policies are untenable...</td>
<td>Op-Ed</td>
</tr>
<tr>
<td>Brian Jones was injured during the final inning...</td>
<td>Sports</td>
</tr>
</tbody>
</table>

**Unsupervised**: texts do not have labels, model learns patterns from **text only**

<table>
<thead>
<tr>
<th>DATA</th>
<th>LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>The final score between the Patriots and the Seahawks...</td>
<td>?</td>
</tr>
<tr>
<td>The politician’s fiscal policies are untenable...</td>
<td>?</td>
</tr>
<tr>
<td>Brian Jones was injured during the final inning...</td>
<td>?</td>
</tr>
</tbody>
</table>
Supervised Methods: Examples

Let’s take a look at some supervised machine learning models!

These models assume that you have labels for your data, and you want to learn how to apply those labels to new data.
Supervised Methods: Examples

We have a dataset of texts and labels.

We divide that dataset into

- a **training** set (used to teach our model)
- a **test** set (used to test whether our model works on new data)
Learn from your neighbors

**Intuition:** assign the most common label amongst an input’s $k$ nearest neighbors

**Assumptions:** similar inputs have similar outputs
Draw a line

**Intuition:** draw a line separating classes of data

**Assumptions:** data is linearly separable

\[ y = wx + b \]
Draw a line and use it as your label probability

**Intuitions**
Predict the **probability** that an input belongs to a class with a sigmoid curve
We transform or “squish” linear regression into a [0,1] range

\[
sig(t) = \frac{1}{1 + e^{-t}}
\]

- \( p = 1 \) label is very likely
- \( p = 0 \) label is unlikely
- \( p = 0.5 \) label is difficult to decide
Draw a line that maximizes the distance between the groups

**Intuition:** find the line separating the classes that has the **maximum margin**

**Assumption:** support vectors are the most useful data points
The bias-variance tradeoff

**Bias**
- under-fit
- makes inaccurate assumptions
- model is too simple

**Variance**
- over-fit
- sensitive to noise
- model is too complex
Loss Function and Regularization

Cost Function

\[ J(w) = \frac{1}{m} \sum_{i=0}^{m} \left| (wx_i + b) - y_i \right| + \lambda \sum_{j=0}^{n} \left| w_j \right| \]

- \( m \) = number of documents
- \( y_i \) = true label
- \( wx_i + b \) = predicted label
- \( n \) = number of features
- \( w \) = feature weight
- \( \lambda \) = strength of regularization
- \( \lambda \) = strength of regularization

Loss Function

punish the model for labeling documents incorrectly

Regularization

punish the model for being too complex
Unsupervised Methods: Examples

Let’s take a look at some unsupervised machine learning models!

These models assume that all you have is your data (no labels), and you want to explore that data and see what patterns and structure you can find.
Find clusters

Intuitions:

- assign each data point to the nearest cluster centroid
- centroids should be the mean of all the points in the cluster
Measure relationships between words

France and Italy are quite similar
\[ \theta \text{ is close to } 0^\circ \]
\[ \cos(\theta) = 1 \]

Ball and crocodile are not similar
\[ \theta \text{ is close to } 90^\circ \]
\[ \cos(\theta) \approx 0 \]

The two vectors are similar but opposite
the first one encodes (city - country)
while the second one encodes (country - city)
\[ \theta \text{ is close to } 180^\circ \]
\[ \cos(\theta) \approx -1 \]

Discover and track topics in a dataset
Introduction to language modeling
What is a language model?

“a model that assigns a probability to sequences of words”

(Jurafsky & Martin, *Speech and Language Processing*)

**Given** a sequence of words, can we **predict** the next sequence of words?

We rely on some text **dataset** to estimate these probabilities.
A simple language model

\[ P(\text{“question”}|\text{“to be or not to be, that is the”}) = ? \]
A simple language model

\[ P(\text{"question"}|\text{"to be or not to be, that is the"}) = ? \]

**Unigrams**: count how many times “question” follows “the” in our dataset.

**Bigrams**: count how many times “question” follows “is the” in our dataset.

**Trigrams**: count how many times “question” follows “that is the” in our dataset.

Etc.
What are “large language models”?  

Also referred to as pretrained models and foundation models.

These models rely on vast amounts of pretraining data. While some performance gains come from the model architectures, a lot is coming from just the sheer amount of pretraining data.

Common pretraining sources include web scrapes, Wikipedia, Reddit, scientific publications, and books.

Examples: ChatGPT, Claude, BERT, Galactica, BLOOM
Ethics of large language models
Selected readings on NLP ethics

**BOOKS**
- *Data Feminism* by Catherine D’Ignazio & Lauren F. Klein
- *Race After Technology* by Ruha Benjamin
- *Sorting Things Out* by C. Bowker and Susan Leigh Star
- *Automating Inequality* by Virginia Eubanks

**PAPERS**
- Ethical Machine Learning in Health Care (Chen et al., 2021)
- Datasheets for Datasets (Gebru et al., 2021)
- The Values Encoded in Machine Learning Research (Birhane et al., 2022)
- A Survey of Race, Racism, and Anti-Racism in NLP (Field et al., 2021)
- Language (Technology) is Power: A Critical Survey of “Bias” in NLP (Blodgett et al., 2020)
Some risks to keep in mind

1. Lack of interpretability

2. Giant datasets that are very difficult to document

3. Poor representation and quality for non-English languages

4. Toxicity/bias that is baked into and even enhanced by the models
Our own sensemaking biases

<table>
<thead>
<tr>
<th>Property</th>
<th>Human-Human Context</th>
<th>Human-Machine Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity Construction</td>
<td>Sensemaking is a question about who I am as indicated by the discovery of how and what I think.</td>
<td>Given multiple explanations, people will internalize the one(s) that support their identity in positive ways.</td>
</tr>
<tr>
<td>Social</td>
<td>What I say and single out and conclude are determined by who socialized me and how I was socialized, and by the audience I anticipate will audit the conclusions I reach.</td>
<td>Differences in micro- and macro-social contexts affect the effectiveness of explanations.</td>
</tr>
<tr>
<td>Retrospective</td>
<td>To learn what I think, I look back over what I said earlier.</td>
<td>Providing explanations before people can reflect on the model and its predictions negatively affects sensemaking.</td>
</tr>
<tr>
<td>Enactive</td>
<td>I create the object to be seen and inspected when I say or do something.</td>
<td>The order in which explanations are seen affects how people understand a model and its predictions.</td>
</tr>
<tr>
<td>Ongoing</td>
<td>Understanding is spread across time and competes for attention with other ongoing projects, by which time my interests may already have changed.</td>
<td>The valence and magnitude of emotion caused by an interruption during the process of understanding explanations from interpretability tools change what is understood.</td>
</tr>
<tr>
<td>Focused on Extracted Cues</td>
<td>The 'what' that I single out and embellish is only a small portion of the original utterance, that becomes salient because of context and personal dispositions.</td>
<td>Highlighting different parts of explanations can lead to varying understanding of the underlying data and model.</td>
</tr>
<tr>
<td>Plausibility over Accuracy</td>
<td>I need to know enough about what I think to get on with my projects, but no more, which means sufficiency and plausibility take precedence over accuracy.</td>
<td>Given plausible explanations for a prediction, people are not inclined to search for the accurate one amongst these.</td>
</tr>
</tbody>
</table>

Table 1: An overview of the seven properties of sensemaking, their description in the human-human context, and our proposed claims for the human-machine context grounded in each property.

“Interpreting Interpretability: Understanding Data Scientists’ Use of Interpretability Tools for Machine Learning” (Kaur et al., 2020)

“Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory” (Kaur et al., 2022)
From model to action
Prompting and finetuning

Start with a base model trained on a bunch of random data.

Add task-specific heads that produce specific output patterns

You can do this via

- **Finetuning**: adjusting the model to a small amount of target data
- **Prompting**: provide the first part of your output as a “hint” of where to go
Quantitative + qualitative methods

Quantitative methods like those used in NLP are great at finding patterns and averages.

But they can also be used to find outliers and interesting cases.

And those results can lead us back to specific stories and the direct words of people seeking care.
Thank you!

I’ll see you all again later today when we’ll do an interactive exercise where you’ll get to try out a language model for yourself!
From Data to Research: NLP & Maternal Health Use Cases
Meet the Speakers

Allan Fong, MS
Senior Research Scientist and Data Scientist, MedStar Health

Angela D. Thomas, DrPH, MPH, MBA
Vice President, Healthcare Delivery Research, MedStar Health

Mark Clapp, MD, MPH
Maternal-Fetal Medicine Specialist, Physician Investigator, Massachusetts General Hospital

Anna Wexler, PhD
Assistant Professor of Medical Ethics and Health Policy, University of Pennsylvania Perelman School of Medicine

Amulya Yadav, PhD
Assistant Professor, PNC Technologies Career Development and Associate Director, Center for Socially Responsible Artificial Intelligence, Penn State University
The Role of NLP in Addressing Maternal Harm through a Patient Safety and Equity Lens

May 18, 2023

Angela D. Thomas, DrPH, Vice President, Healthcare Delivery Research, MedStar Health Research Institute
Allan Fong, MS, Senior Research and Data Scientist, MedStar Health Research Institute
Agenda

1. Safe Babies Safe Moms Maternal Taxonomy Study
   Background
   Disparities in Maternal Harm
   Addressing Disparities through a Patient Safety & Health Equity Lens
   Aims
2. Deeper Dive on NLP Aim
3. Future Directions
SBSM Maternal Taxonomy Supplemental Study
Note on Inclusivity

Not all birthing individuals identify as women or mothers. Throughout this presentation, you may hear the terms women and mothers to most accurately reflect the current state of the literature, but our approach is meant to be inclusive of all birthing individuals.
About Safe Babies Safe Moms

- April 2020 – March 2025
- $30M initiative
- Largest philanthropic donation in MedStar history

Reduce disparities in maternal & infant mortality in Washington, D.C.
Key Literature Facts

• Maternal harm is a major crisis that disproportionately affects Black women
• Education is not a protective factor
• ~80% of maternal deaths and ~90% of severe maternal morbidity events are preventable
• Most frequent preventable factors are provider-related and/or system-related
Figure 1: The Safety Spectrum

Aviation Safety: Constantly monitors and investigates ALL

Patient Safety: Increasingly better at monitoring and investigating

Maternal: Grossly limited
Figure 2: Patient Safety Taxonomy Adapted from the MERP Index for Categorizing Medication Errors

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Unsafe Condition (Non-Event)</td>
</tr>
<tr>
<td>B1</td>
<td>Near miss-No Harm - Didn’t Reach Patient/Caught by Chance</td>
</tr>
<tr>
<td>B2</td>
<td>Near miss-No Harm - Didn’t Reach Patient because of Active Recovery by Caregivers</td>
</tr>
<tr>
<td>C</td>
<td>No Harm - Reached Patient - No Monitoring Required</td>
</tr>
<tr>
<td>D</td>
<td>No Harm - Reached Patient - Monitoring Required</td>
</tr>
<tr>
<td>E</td>
<td>Harm - Temporary Harm - Intervention Needed</td>
</tr>
<tr>
<td>F</td>
<td>Harm - Temporary Harm - Hospitalization Needed</td>
</tr>
<tr>
<td>G</td>
<td>Harm - Permanent Harm</td>
</tr>
<tr>
<td>H</td>
<td>Harm - Permanent Harm - Intervention Required to Sustain Life</td>
</tr>
<tr>
<td>I</td>
<td>Death</td>
</tr>
</tbody>
</table>
More Key Literature Facts

• Racism, discrimination, and implicit bias contribute to adverse health outcomes that include adverse maternal outcomes

• Giving Voice to Mothers study found that women of color are more likely to experience mistreatment during the perinatal period
  – Shouted at, scolded, threatened, ignored, or receiving no response to requests for help

• Stereotypes stigmatizing Black motherhood including assumptions about being single, low income, and having multiple children
What doesn’t get documented…

breath. Due to her history of pulmonary embolisms (Williams underwent emergency treatment for a life-threatening embolism in 2011), the tennis star quickly alerted a nurse about her symptoms.

But the response wasn’t what she expected. Vogue writer Rob Haskell explains:

She walked out of the hospital room so her mother wouldn’t worry and told the nearest nurse, between gasps, that she needed a CT scan with contrast and IV heparin (a blood thinner) right away. The nurse thought her pain medicine might be making her confused. But Serena insisted, and soon enough a doctor was performing an ultrasound of her legs. “I was like, a Doppler? I told you, I need a CT scan and a heparin drip,” she remembers telling the team. The ultrasound revealed nothing, so they sent her for the CT, and sure enough, several small blood clots had settled in her lungs. Minutes later she was on the drip. “I was like, listen to Dr. Williams!”

• If she died…”pre-existing conditions” (not bias)
Story after story after story...

There was the new mother in Nebraska with a history of hypertension who couldn't get her doctors to believe she was having a heart attack until she had another one. The young Florida mother-to-be whose breathing problems were blamed on obesity when in fact her lungs were filling with fluid and her heart was failing. The Arizona mother whose anesthesiologist assumed she smoked marijuana because of the way she did her hair. The Chicago-area businesswoman with a high-risk pregnancy who was so upset at her doctor's attitude that she changed OB/GYNs in her seventh month, only to suffer a fatal postpartum stroke.
Kira Johnson

Charles Johnson (her widower) and their sons
Including the birthing individual’s voice can uncover the role of racism, discrimination, and implicit bias across the full spectrum of maternal harm.
Opportunities to Systematically Remove Bias

The Central Tenant of Human Factors

“We don’t redesign humans; We redesign the system within which humans work.”

MedStar Health Research Institute
Key Innovations

• Expertise in using clinical informatics techniques to identify “signals” in the EHR of unsafe care delivery
  – Identify “signals” in the electronic record for unsafe maternal care delivery

• Data science expertise in analyzing clinic notes in the EHR for tone and sentiment in preventable patient safety events
  – Apply to clinic notes for maternal care to uncover the role of differential treatment and biases
Aviation Safety: Constantly monitors and investigates ALL

Patient Safety: Increasingly better at monitoring and investigating

Maternal Taxonomy Study

Maternal: Grossly limited
Aim 1 (Phase 1)

1. Identifying common themes and signals that contribute to unsafe conditions, hazards, near misses, and other maternal injuries by:
   - Incorporating birthing individual voice – quantitative and qualitative interviews
   - Analysis of patient complaints
   - Common signals in EHR
   - Voluntary occurrence reporting system analysis
   - NLP Analysis of EHR Notes
Aim 2 (Phase 2)

2. Developing a full maternal safety spectrum taxonomy by identifying how the common themes and signals contribute to a progression from:
   – Safe to unsafe conditions
   – Unsafe conditions to hazards
   – Hazards to near misses
   – Near misses to maternal harm
   – Maternal harm to severe maternal morbidity
   – Severe maternal morbidity to maternal mortality
Aim 3 (Phase 2)

3. **Developing a toolkit** for ongoing surveillance of common themes and signals, with mitigation strategies when they are identified.
Local & National Interdisciplinary Expertise

- Maternal qualitative survey experts
- Health equity experts
- Obstetrics experts
- Midwifery experts
- Human factors experts
- Patient safety experts
- Data science experts
- Informatics experts
- Biostatistics experts
Aim 1 (Phase 1)

1. Identifying common themes and signals that contribute to unsafe conditions, hazards, near misses, and other maternal injuries by:
   – Incorporating birthing individual voice – quantitative and qualitative interviews
   – Analysis of patient complaints
   – Common signals in EHR
   – Voluntary occurrence reporting system analysis
   – NLP Analysis of EHR Notes
A Deeper Dive

Using NLP - Biases in EHR Notes
Method - Motivation

Negative Patient Descriptors: Documenting Racial Bias In The Electronic Health Record

Michael Sun, Tomasz Oliwa, Monica E. Peek, and Elizabeth L. Tung

Published: January 19, 2022

Abstract

Little is known about how racism and bias may be communicated in the medical record. This study used machine learning to analyze electronic health records (EHRs) from an urban academic medical center and to investigate whether providers’ use of negative patient descriptors varied by patient race or ethnicity. We analyzed a sample of 40,113 history and physical notes from January 2019–October 2020 across 18,459 patients.

- Sentences containing a negative descriptor
- Compared with White patients, Black patients had 2.54 times the odds of having at least one negative descriptor in the history and physical notes
Method

- **Data Source**
  - Jan. 1, 2016 to March 31, 2020
  - History and physical, review of system, history of present illness, physical examination, triage notes
Method

- **Data Source**
  - Jan. 1, 2016 to March 31, 2020
  - History and physical, review of system, history of present illness, physical examination, triage notes

- **Negative Descriptors**
  - Nonadherent, aggressive, agitated, angry, challenging, combative, noncompliant, confront, noncooperative, defensive, exaggerate, hysterical, unpleasant, refuse, and resist
  - Descriptor and root
  - Query expansion using W2V (Why use NLP)
Method

• Data Source
  – Jan. 1, 2016 to March 31, 2020
  – History and physical, review of system, history of present illness, physical examination, triage notes

• Negative Descriptors
  – Nonadherent, aggressive, agitated, angry, challenging, combative, noncompliant, confront, noncooperative, defensive, exaggerate, hysterical, unpleasant, refuse, and resist
  – Descriptor and root
  – Query expansion using W2V (Why use NLP)
Method

• Data Source
  – Jan. 1, 2016 to March 31, 2020
  – History and physical, review of system, history of present illness, physical examination, triage notes

• Negative Descriptors
  – Nonadherent, aggressive, agitated, angry, challenging, combative, noncompliant, confront, noncooperative, defensive, exaggerate, hysterical, unpleasant, refuse, and resist
  – Descriptor and root
  – Query expansion using W2V (Why use NLP)
Method

• Data Source
  – Jan. 1, 2016 to March 31, 2020
  – History and physical, review of system, history of present illness, physical examination, triage notes

• Negative Descriptors
  – Nonadherent, aggressive, agitated, angry, challenging, combative, noncompliant, confront, noncooperative, defensive, exaggerate, hysterical, unpleasant, refuse, and resist
  – Descriptor and root
  – Query expansion using W2V (Why use NLP)
Method

• ‘aggressive’
  – “displaying irritable mood and aggression towards staff…”
• angry
  – “she felt as if the previous nurse was laughing at her and this made her very angry…”
• exaggerated
  – “her exaggerated blame on the hospital staff…”
• defensive
  – “she became defensive and upset and said it was because of…”
Method

- Annotations
- Train model (Why use NLP)
- Applied model to unseen data
- Analyze results at patient level with demographics
## Model Results

Performance metrics of the negative language classifiers

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Train Data</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Under Receiver Operating Characteristics Curve</td>
<td>0.99</td>
<td>0.89</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.81</td>
<td>0.61</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>F-1 Score</td>
<td>0.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Area Under Precision-Recall Curve</td>
<td>0.94</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Discussion

• Analysis is still underway
• Key differences align with previous literature on documentation bias
• Why use NLP
  – Data driven way to tailor search terms
    • Our healthcare system specific
  – Too many sentences to manually review
    • Limited time and resources
Future Directions

NLP & Operations
Discussion - NLP

• Negative language/racial bias
• Positive language
  – ‘a pleasant 26 yo female patient’
• Self-awareness
  – Changes in documentation tone (ie. between patients, open notes access, burn-out)
Discussion - Operations

• Aim 4 – Pilot the toolkit
• Ongoing surveillance efforts for accountability and change
  – Ongoing capture of the birthing individuals’ voices
  – Ongoing analysis of patient complaints
  – Ongoing surveillance of signals in the EHR
  – Ongoing analysis of voluntary occurrence reporting system analysis
  – Ongoing analysis of EHR Notes
• Opportunities for automation using NLP?
Thank you

It’s how we treat people.
NLP for Risk Prediction in Maternity Care

Mark Clapp, MD MPH
Massachusetts General Hospital
Harvard Medical School
The Last Person You’d Expect to Die in Childbirth

Nearly Dying In Childbirth: Why Preventable Complications Are Growing In U.S.

KATHERINE ELLISON, PROPUBLICA

NINA MARTIN, PROPUBLICA

Maternal mortality: An American crisis
Goals of Risk Prediction
Goals of a Clinical Risk Tool

Utilize EHR information to develop methods for personalized maternal risk stratification in obstetrics
Natural Language Processing
Ms. Smith presents for induction of labor for pre-eclampsia.
Model Derivation and Testing

Hospital A
(Model Development)

\[ A_{\text{train}} \]
Training Set
(75%)

\[ A_{\text{test}} \]
Test Set
(25%)

Hospital B
(External Validation)

\[ B_{\text{valid}} \]

Learn Model

Internal Testing

External Testing

Development and Testing of Model

Clapp et al. AJOG. 2022.
Model Performance

Area Under the Curve (AUC) values

0.72 for severe morbidity

0.76 for severe morbidity (excluding transfusion)

Clapp et al. AJOG. 2022.
Model Performance

Severe Morbidity

Non-Transfusion Severe Morbidity

Clapp et al. AJOG. 2022.
NLP Model vs. Manual Risk Assignment
NLP Model vs. Manual Risk Assignment

## Model Performance

<table>
<thead>
<tr>
<th>Model Characteristic</th>
<th>OB-CMI</th>
<th>NLP</th>
<th>OB-CMI+ NLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen Positive Rate</td>
<td>3.9%</td>
<td>4.2%</td>
<td>7.1%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>24.4%</td>
<td>28.7%</td>
<td>37.4%</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>17.6%</td>
<td>19.4%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Specificity</td>
<td>96.7%</td>
<td>96.5%</td>
<td>93.8%</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>97.8%</td>
<td>97.9%</td>
<td>98.1%</td>
</tr>
</tbody>
</table>

Other Findings

• Similar performance when using more advanced NLP modeling techniques
  • Bi- and tri-gram models
  • Including negation
  • Using pretrained dictionaries
  • Various machine learning models

• Similar results when predicting postpartum hemorrhage, one of the most common pregnancy complications
Next Steps

Prospective application in clinical practice
Challenges for NLP Applications

Translation to the bedside
- End-user buy-in
- Adapting clinical workflows
- Facile health IT infrastructure

Ensuring performance and equity in application
- Challenges in inequities in underlying data to derive models
- Variable model performance among population subgroups
- Uncertain generalizability
- Availability of model/technology to all populations
Special Thanks

Funders
• AAOGF/ABOG Research Scholar Award
• SMFM Bridge Award

Collaborators
• Anjali Kaimal, MD MAS
• Tom McCoy, MD
• Roy Perlis, MD MSc
• Jeff Ecker, MD
NLP for Risk Prediction in Maternity Care

Mark Clapp, MD MPH
Massachusetts General Hospital
Harvard Medical School

mark.clapp@mgh.harvard.edu  @markaclapp
Pregnancy in the Age of the Internet: A Content Analysis of Online Pregnancy Forums

AAMC Maternal Health Equity Workshop | May 18, 2023

Anna Wexler
Assistant Professor, Department of Medical Ethics & Health Policy
University of Pennsylvania Perelman School of Medicine
PREGNANCY AND DIGITAL HEALTH

➢ Pregnant individuals increasingly turn to the Internet & digital health during pregnancy (Wexler et al. 2020)

➢ Prior work has assessed how and why people utilize online health information during pregnancy: they find it empowering, entertaining, quick, reassuring (Sayakhot & Carolan-Olah 2016, for review)

➢ Yet little work has assessed what pregnant individuals generate (i.e., online content)
Community Groups

Join our community for baby name ideas, due date discussions, local birth and parent community groups, and more!

Find Moms Who Share Your Birth Month
<table>
<thead>
<tr>
<th>Month</th>
<th>Members</th>
<th>Discussions</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 2019 Babies</td>
<td>230K</td>
<td>59.4K</td>
</tr>
<tr>
<td>July 2019 Babies</td>
<td>214K</td>
<td>56.4K</td>
</tr>
<tr>
<td>June 2019 Babies</td>
<td>185K</td>
<td>52K</td>
</tr>
<tr>
<td>May 2019 Babies</td>
<td>181K</td>
<td>54.7K</td>
</tr>
<tr>
<td>April 2019 Babies</td>
<td>171K</td>
<td>58.8K</td>
</tr>
<tr>
<td>March 2019 Babies</td>
<td>172K</td>
<td>60K</td>
</tr>
<tr>
<td>February 2019 Babies</td>
<td>156K</td>
<td>54.3K</td>
</tr>
<tr>
<td>January 2019 Babies</td>
<td>167K</td>
<td>62K</td>
</tr>
<tr>
<td>December 2018 Babies</td>
<td>165K</td>
<td>63.5K</td>
</tr>
<tr>
<td>November 2018 Babies</td>
<td>157K</td>
<td>56.2K</td>
</tr>
<tr>
<td>October 2018 Babies</td>
<td>161K</td>
<td>60.5K</td>
</tr>
</tbody>
</table>

~500,000 total posts per birth club month across 3 trimesters + post-partum
Getting a used breast pump from S-I-L?
My sister in law is selling a lot of her baby stuff for cheap and so we are going to get much as we can. However, one of the items is a breast pump. I was all for it due to the cost savings (my insurance won't cover it) but now I'm reading on...
Created by WGrigg317
Last comment from lashawn/712 23 minutes ago

Flu shot or no?
Last time I was pregnant the doctor literally scared me into getting a flu shot...said me and baby could die if I didn't...this time around I don't know what to think or do? Anyone have an opinion on this? Thanks
Created by alisonmarie
Last comment from JFlickman 28 minutes ago

It doesn't hurt to be nice
Or to just keep scrolling.
Created by dcnemomma
Last comment from B864 31 minutes ago

Heading to the ER and crossing fingers
I was having back pain on sunday that was significant enough to make an apt in the middle of the night for Monday with my ob. I generally felt malaise and just rotten. She did a cervix check and urine test stating my urine had some white blood...
Created by ArtisanEGG
Last comment from B864 31 minutes ago

show me your nursery w/&#128557;:
23 weeks and its already almost all set up! Am I crazy? This is baby #5 and I don't think I've ever got it done this soon - it will be shared with his older sister who will be 19 months when he arrives. I still have alot of organizing to do...
Created by lemonitefresh
Last comment from Lisa2ndtimepregna 33 minutes ago

Jalen or Jaylen —HELP!
Help!! I need help deciding on the spelling for my little guy. We like both but can't decide. Thanks!
Created by JMS1023

Update after a scary situation
Hey! I got discharged from the hospital this afternoon and before I left spent time with Noah and he's doing good! His breathing tube got moved or clogged whatever happened he stopped breathing so they started pumping air with a bag and his...
Created by MHandy82518
Last comment from Amberlee/317 42 minutes ago

Weird gush of fluid
So first off I have called the dr. It was the first thing I did. I was at work just standing up and I felt a gush similar to heavy discharge. I went to the bathroom though and it had left quite a patch on my leggings. I will mention I've been...
Created by Hayleyukt
Last comment from DubMam 45 minutes ago

Group B strep in urine- concerning?
Just got my urine test results with the above diagnosis. Anyone else had such situation?
Created by Catherine12350 46 minutes ago

Debilitating Round Ligament Pain
Anyone else struggling with debilitating round ligament pain? I've been pretty much bed bound for three days now and I can't find any information online or with my doctor for how to cope. Baths, stretches, Tylenol, belly band, etc. all provide...
Created by katiemoo 47 minutes ago

HELP! Abdominal muscle pain!
I have been experiencing a slowly increasing pain in what feels like my abdominal muscle on the right side. I can push into my belly and find a very painful spot and it seems to get worse when I am active or working (working especially, I'm...
Created by Landria1986
Last comment from katiemoo 48 minutes ago

Iron supplements
My doctor told me I need to take an iron supplement in addition to my (prescription) prenatal that has extra iron in it. I asked her twice but I've never taken iron before and I'm still
Created by JMS1023
Goal: to better understand the topics that individuals post about on public online pregnancy forums using automatic methods of language analysis (specifically, topic modeling)

- to discover the extent to which individuals discuss health-related topics, and which topics appear most frequently

- to differentially characterize the topics that individuals discuss across the three trimesters and the postpartum period
THE TEAM

Graciela Gonzalez-Hernandez
Associate Professor of Informatics
Department of Biostatistics, Epidemiology, and Informatics

Karen O’Connor
Senior Data Analyst
Department of Biostatistics, Epidemiology, and Informatics

Davy Weissenbacher
Research Data Scientist
Department of Biostatistics, Epidemiology, and Informatics

Anahita Davoudi
Lead Statistician
Department of Biostatistics, Epidemiology, and Informatics

Holly Cummings
Assistant Professor of Clinical Obstetrics and Gynecology
Department of Obstetrics and Gynecology

Rebekah Choi
Project Manager
Department of Medical Ethics and Health Policy
Figure 1. Data collection flowchart.

All posts collected from 7 different WhatToExpect.com 2018 birth club months (N=3,337,891)

Total initial posts (n=305,183)

Total comments (n=3,072,091)
  - Average number of comments per initial post (SD): 10 (26)

  - Excluded due to file incompatibility (e.g., special characters), or scraping error (n=82)
  - Excluded due to time frame eligibility (n=42,863)

Total initial posts analyzed (n=262,238)

Average initial posts per birth club month (SD)
  - Initial posts: 37,463 (3,678)
  - Users (those who made initial posts): 7,410 (712)
1. Topic modeling method known as Latent Dirichlet Allocation (LDA), which discovers a “word cluster” that has a high probability of appearing together.

Example:

\[ 0.068 \times \text{"morning"} + 0.050 \times \text{"symptom"} + 0.040 \times \text{"nausea"} + 0.037 \times \text{"pregnancy"} + \\
0.035 \times \text{"throw"} + 0.033 \times \text{"trimester"} + 0.032 \times \text{"stomach"} + 0.030 \times \text{"sickness"} + 0.030 \times \text{"sick"} + \\
0.023 \times \text{"nauseous"} + 0.018 \times \text{"vomit"} + 0.017 \times \text{"pregnant"} + 0.017 \times \text{"eat"} + 0.016 \times \text{"experience"} \\
+ 0.014 \times \text{"wrong"} \]

2. Requires pre-processing data & setting parameters (50 topics, 200 iterations)
<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Wordcloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>Vital signs</td>
<td>blood pressure, heart rate, respiration, temperature</td>
</tr>
<tr>
<td></td>
<td>Blood work</td>
<td>blood, red blood cells, white blood cells, platelets</td>
</tr>
<tr>
<td></td>
<td>Mental health</td>
<td>memory, mood, stress, anxiety, depression</td>
</tr>
<tr>
<td></td>
<td>Allergies</td>
<td>pollen, dust, mold, pet dander, foods</td>
</tr>
<tr>
<td></td>
<td>Nutritional</td>
<td>nutrition, diet, vitamins, minerals</td>
</tr>
<tr>
<td></td>
<td>Benefits</td>
<td>exercise, weight loss, muscle gain, improved sleep</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>Pre-pregnancy</td>
<td>fertility, ovulation, prenatal care, weight gain</td>
</tr>
<tr>
<td></td>
<td>Early pregnancy</td>
<td>nausea, fatigue, mood swings, prenatal vitamins</td>
</tr>
<tr>
<td></td>
<td>Late pregnancy</td>
<td>back pain, sciatica, varicose veins, edema</td>
</tr>
<tr>
<td></td>
<td>Postpartum</td>
<td>nursing, lactation, postpartum depression, return to work</td>
</tr>
<tr>
<td></td>
<td>Newborn care</td>
<td>breastfeeding, jaundice, diaper rash, meconium</td>
</tr>
<tr>
<td></td>
<td>Infant care</td>
<td>immunizations, developmental milestones, feeding issues</td>
</tr>
<tr>
<td></td>
<td>Sibling care</td>
<td>sibling rivalry, sibling bonds, sibling care tips</td>
</tr>
<tr>
<td></td>
<td>Geriatric care</td>
<td>falls prevention, arthritis, memory loss, incontinence</td>
</tr>
<tr>
<td></td>
<td>Child care</td>
<td>child development, discipline, potty training, sensory processing</td>
</tr>
<tr>
<td></td>
<td>Adoption</td>
<td>adoption process, placement, legal aspects, post-adoption support</td>
</tr>
<tr>
<td></td>
<td>Infertility</td>
<td>infertility treatments, ovulation tracking, fertility medications</td>
</tr>
<tr>
<td></td>
<td>Reproductive</td>
<td>ovulation, menstrual cycle, sperm, eggs, hormones</td>
</tr>
<tr>
<td></td>
<td>Men's health</td>
<td>prostate health, testicular health, sexual health</td>
</tr>
<tr>
<td></td>
<td>Women's health</td>
<td>menopause, endometriosis, polycystic ovarian syndrome, pelvic floor disorders</td>
</tr>
</tbody>
</table>

---

**Figure 3. Distribution of posts by label and rank per time period**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Time period</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First trimester</td>
<td>150,000</td>
</tr>
<tr>
<td>2</td>
<td>Second trimester</td>
<td>100,000</td>
</tr>
<tr>
<td>3</td>
<td>Third trimester</td>
<td>50,000</td>
</tr>
<tr>
<td>4</td>
<td>Postpartum</td>
<td>25,000</td>
</tr>
</tbody>
</table>

---

**Figure 4. Distribution of posts by label and rank per time period**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Time period</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First trimester</td>
<td>120,000</td>
</tr>
<tr>
<td>2</td>
<td>Second trimester</td>
<td>80,000</td>
</tr>
<tr>
<td>3</td>
<td>Third trimester</td>
<td>40,000</td>
</tr>
<tr>
<td>4</td>
<td>Postpartum</td>
<td>20,000</td>
</tr>
</tbody>
</table>

---

**Figure 5. Distribution of posts by label and rank per time period**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Time period</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First trimester</td>
<td>90,000</td>
</tr>
<tr>
<td>2</td>
<td>Second trimester</td>
<td>60,000</td>
</tr>
<tr>
<td>3</td>
<td>Third trimester</td>
<td>30,000</td>
</tr>
<tr>
<td>4</td>
<td>Postpartum</td>
<td>15,000</td>
</tr>
</tbody>
</table>
Figure 2. Distribution of posts by general category

- Maternal Health (45%) n=116,693
- Baby (29%) n=76,317
- People/Relationships (10%) n=26,412
- Product Recommendations (6%) n=16,615
- Misc (4%) n=9,675
- Pregnancy Confirmation (3%) n=8,695
- Excluded due to noise (3%) n=7,831
<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Topic Label</th>
<th>Total Posts</th>
<th>Word Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal Health</td>
<td>Physical Symptoms</td>
<td>Pain (mostly abdomen/back/pelvic)</td>
<td>13,196</td>
<td>pain, back, sit, hurt, experience, walk, pressure, stomach, stand, pelvic, help</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Morning sickness/nausea</td>
<td>9,751</td>
<td>morning, symptom, nausea, pregnancy, throw, trimester, stomach, sickness, sick, nauseous, vomit, pregnant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bleeding/spotting/cramping</td>
<td>9,357</td>
<td>cramp, bleed, period, spot, blood, discharge, normal, light, experience, red, brown, wipe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Misc health symptoms (e.g., sore/painful breasts, constipation, hemorrhoids)</td>
<td>5,488</td>
<td>weird, pee, hurt, sore, boob, happen, leak, normal, burn, tmi, heartbeat, notice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anxiety/depression; medication questions</td>
<td>4,798</td>
<td>anxiety, pregnancy, help, postpartum, tear, pregnant, infection, issue, doctor, experience, effect, medication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cold/flu symptoms</td>
<td>4,661</td>
<td>sick, cold, cough, flu, fever, shoot, nose, breathe, throat, catch, ear, help</td>
</tr>
<tr>
<td>Labor</td>
<td>Labor (induction/dilation)</td>
<td></td>
<td>12,062</td>
<td>induce, cm, dilate, labor, push, check, induction, cervix, due, epidural, doctor, bear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labor signs (contractions, am I in labor)</td>
<td>6,629</td>
<td>contraction, labor, water, break, min, braxton, apart, sign, hick, painful, hospital, strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anxiety about delivery</td>
<td>5,930</td>
<td>birth, hospital, nervous, scare, delivery, deliver, plan, experience, labor, natural, mom, prepare</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mucus plug (&amp; baby position/breech q’s)</td>
<td>1,983</td>
<td>head, show, lose, turn, mucus, plug, bloody, position, breech, flip, flat, mucous</td>
</tr>
<tr>
<td>Body and mood changes</td>
<td>Diet &amp; food (cravings, aversions, safety)</td>
<td></td>
<td>6,578</td>
<td>eat, drink, food, water, meal, dinner, tea, crave, snack, coffee, lunch, ice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stretch marks &amp; baby bump</td>
<td>5,558</td>
<td>belly, bump, leg, stretch, mark, pop, button, itchy, oil, pregnancy, boil, itch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weight (gain/loss)</td>
<td>4,584</td>
<td>weight, gain, lb, pound, loss, pregnancy, weigh, eat, healthy, pregnant, pre, fat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Energy levels (fatigue) &amp; exercise</td>
<td>3,962</td>
<td>pregnancy, body, run, tire, pregnant, walk, struggle, toddler, energy, exhaust, remember, easy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sex (sex drive, sex after birth)</td>
<td>1,913</td>
<td>crazy, sex, drive, headache, husband, pregnant, breath, hubby, migraine, pregnancy, nut, insane</td>
</tr>
<tr>
<td>Pregnancy complications</td>
<td>Pre-eclampsia (BP/protein/urine) but also general doc apt</td>
<td></td>
<td>9,209</td>
<td>doctor, appointment, blood, ob, dr, check, pressure, office, apt, back, nurse, concern</td>
</tr>
<tr>
<td></td>
<td>Placenta disorders during pregnancy (&amp; hypnobirthing/natural birth q’s)</td>
<td></td>
<td>2,132</td>
<td>hear, curious, experience, opinion, placenta, wonder, vs, whether, mum, class, love, mom</td>
</tr>
<tr>
<td>Topic</td>
<td>Score</td>
<td>Tags</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparing for baby</td>
<td></td>
<td>excited, baby, love, sweet, meet, beautiful, amaze, healthy, son, team, bless</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing baby name</td>
<td></td>
<td>name, middle, girl, boy, love, husband, pick, help, decide, suggestion, choose, opinion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First ultrasound picture &amp; early gender prediction</td>
<td></td>
<td>picture, pic, foot, swell, ramzi, twin, dream, photo, ultrasound, theory, rub, skull</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender reveal</td>
<td></td>
<td>guess, gender, fun, reveal, csection, surprise, ultrasound, wife, scan, idea, tale, correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby’s health (in utero)</td>
<td></td>
<td>pregnancy, miscarriage, heartbeat, ultrasound, level, pregnant, hcg, sac, nervous, progesterone, hear, pray</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concerns about ultrasound results</td>
<td></td>
<td>ultrasound, scan, measure, anatomy, doctor, tech, small growth, experience, result, back, fluid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby movement</td>
<td></td>
<td>move, kick, movement, tummy, roll, active, count, back, crawl, normal, lay, ride</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby heart rate</td>
<td></td>
<td>heart, rate, heat, glass, cup, wine, program, sippy, babys, proud, bpm, star</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring for newborn</td>
<td></td>
<td>sleep, night, wake, nap, fall, asleep, back, bed, awake, help, eat, fee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby sleep routine</td>
<td></td>
<td>milk, breast, feed, nurse, bottle, supply, nipple, fee, formula, latch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pumping &amp; breastfeeding</td>
<td></td>
<td>pump, breastfeed, milk, breast, feed, nurse, bottle, supply, nipple, fee, formula, latch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby feeding routines</td>
<td></td>
<td>bottle, formula, oz, eat, feed, spit, gas, solid, reflux, fee, burp, switch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby sleep arrangements</td>
<td></td>
<td>room, crib, play, sleep, swaddle, bassinet, rock, nursery, transition, arm, bed, pack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin problems (mostly baby, some mom)</td>
<td></td>
<td>change, diaper, face, skin, rash, app, iron, help, wet, acne, red, cream</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby teething &amp; thumb sucking</td>
<td></td>
<td>hand, mouth, suck, tooth, book, teethe, finger, pull, toy, pacifier, cut, lip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel movements (some pet related)</td>
<td></td>
<td>poop, dog, video, cat, pooping, potty, green, train, pooped, ride, fresh, normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misc</td>
<td></td>
<td>eye, open, white, color, blue, black, dark, pink, light, wids, fill, contact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Topic</td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People/relationships</td>
<td>Husband (related to household chores, work, kids)</td>
<td>13,478</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family/friends</td>
<td>12,934</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product recommendations</td>
<td>Baby maternity clothes &amp; baby items</td>
<td>7,325</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baby transport options (carseats, strollers)</td>
<td>6,343</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bra recommendations</td>
<td>2,947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pregnancy confirmation</td>
<td>Pregnancy test (also glucose test)</td>
<td>4,643</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimating due date &amp; trying to conceive</td>
<td>4,052</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misc</td>
<td>Emotional moms &amp; crying/fussy babies</td>
<td>4,225</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Misc temperature (hot weather, hot flashes, dressing baby for weather)</td>
<td>1,430</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Back to work &amp; childcare plans</td>
<td>1,262</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Six topics excluded due to noise</strong></td>
<td>7,831</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>TOTAL</strong></td>
<td>262,238</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- T1 = First trimester
- T2 = Second trimester
- T3 = Third trimester
- PP = Postpartum period
More than just emotional support or product recommendations, pregnant individuals are talking extensively about their health on online pregnancy forums.

Areas of greatest health information-seeking:
- Miscarriage
- Labor
- Newborn care

Pregnant individuals report that online information affects their health-care decision-making (Lagan 2010)… yet as many as 70-75% of them do not speak to health care providers about information retrieved from the Internet (Fredriksen et al 2016; Larsson 2009; Gao et al. 2013)
LIMITATIONS

➢ No demographic information about who participates in these forums
➢ Topic modeling limitations (i.e., posts that were very general likely did not arise in topic list)

FUTURE QUESTIONS

➢ Who participates in online pregnancy forums and who does not?
➢ What are pregnant individuals discussing online that they are not bringing up in their OB’s office?
➢ How is online health information changing the traditional patient-physician relationship in the maternal health context?
Thank you!

Anna Wexler
awex@pennmedicine.upenn.edu

This project was supported by the Office of the Director, NIH, under Award Number DP5OD026420
Harnessing NLP for Assisting Tele-Triage of Expectant & New Mothers in Kenya

Amulya Yadav
Penn State University
The State of Maternal Health Care in Kenya

Overview:

• Kenya has some of the highest rates of maternal and newborn mortality
  o 14 times higher than the US
  o 100 times higher than some countries in Western Europe
  o below the international minimum to deliver essential health services

• Why is this happening?
  o Mothers live in rural areas quite far away from hospitals or clinic
  o They lack access to timely information about their medical symptoms
Overview:

- Kenya has some of the highest rates of maternal and newborn mortality
  - 14 times higher than the US
  - 100 times higher than some countries in Western Europe
  - below the international minimum to deliver essential health services

- Why is this happening?
  - Mothers live in rural areas quite far away from hospitals or clinic
  - They lack access to timely information about their medical symptoms
    - Do my current medical symptoms warrant a trip to the doctor?
Enter Jacaranda Health

- They deliver low-cost and sustainable solutions to improve the quality of care for mothers and newborns.
- Currently serving more than 3 million mothers and babies annually.
A digital health platform that connects mothers with lifesaving advice and referral to care.
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

PennState
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

Users ask questions about pregnancy
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

Users ask questions about pregnancy

High urgency? (helpdesk agents)
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

Users ask questions about pregnancy

Send direct answers

High urgency? (helpdesk agents)

Low Urgency
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

Users ask questions about pregnancy

Send direct answers

Connect mothers to hospitals

High urgency? (helpdesk agents)

Low Urgency

High Urgency
A digital health platform that connects mothers with lifesaving advice and referral to care.

Users register for PROMPTS

Users ask questions about pregnancy

Quality care are provided

Send direct answers

Connect mothers to hospitals

Low Urgency

High Urgency

High urgency? (helpdesk agents)
A digital health platform that connects mothers with lifesaving advice and referral to care.

2 million enrolled into PROMPTS so far

0.35 million users enroll to PROMPTS every month

1.1 million messages received every month
A digital health platform that connects mothers with lifesaving advice and referral to care.

- **2 million** enrolled into PROMPTS so far
- **0.35 million** users enroll to PROMPTS every month
- **1.1 million** messages received every month

**Medical tips are sent to users**

- **High urgency?** (helpdesk agents)
- **Quality care** are provided
- **Users register for PROMPTS**
- **Users ask questions** about pregnancy
- **Send direct answers**
- **Refer the case to nearby hospitals**

2 million enrolled into PROMPTS so far

<table>
<thead>
<tr>
<th>Questions to Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Coronavirus and pregnancy</td>
</tr>
<tr>
<td>B. What if I am pregnant and become infected?</td>
</tr>
<tr>
<td>C. Coronavirus and miscarriage</td>
</tr>
<tr>
<td>D. Coronavirus and infants</td>
</tr>
<tr>
<td>E. Should I go to my clinic?</td>
</tr>
<tr>
<td>F. What if I am a healthcare worker and pregnant?</td>
</tr>
<tr>
<td>G. Information on Health Facilities in my County</td>
</tr>
<tr>
<td>X. I don't need any more information at this time</td>
</tr>
</tbody>
</table>

1.1 million messages received every month
PROMPTS

A digital health platform that connects mothers with lifesaving advice and referral to care.

- **2 million** enrolled into PROMPTS so far
- **0.35 million** users enroll to PROMPTS every month
- **1.1 million** messages received every month

Answering messages is not straightforward for helpdesk agents.
A digital health platform that connects mothers with lifesaving advice and referral to care.

How can we use NLP to improve the health care situation in Kenya and provide a better service experience?
Problem Statement

Goal:
- Automate the task of predicting emergency levels of a user’s medical condition based on their SMS messages

Contribution:
- Propose an NLP framework, TRIM-AI, to classify cases into different emergency levels
- Utilize a special Large Language Model (LLM) and continue pretraining to deal with the code-mixed text
- Output a sorted batch of incoming SMS messages (in decreasing order of risk score)
A Quick Intro on LLMs
A Quick Intro on LLMs: What do they do?

Can you please come here?

History → Word being predicted

Bard AI

Meta
A Quick Intro on LLMs: Building Classifiers

Meta LLAMA Model
(An Example of an LLM)

Classification Network

Predictions
Can we use an LLM based classifier to predict the severity of a mother’s medical condition using the content of their SMS messages?
**Code-mixing phenomenon:**

Phrases or words which belong to different languages appear in one sentence.

- Blue: English
- Red: Swahili

**Example 1:**

Code-mixed message: My son alimeza a coin since sunday na ajaenda haja is it risk ama itatoka tu

Translated English message: My son has swallowed a coin since sunday and that is going to need is it risky or it will just go away

**Example 2:**

Code-mixed message: Je heartburn inasababishwa na nini? sababu hata nikule nini lazma nikuwe nayo

Translated English message: What causes heartburn? Because even if I eat what I have to have
Most SMS messages received by the PROMPTS platform are code-mixed between English & Swahili.

Example 1:

mixed message:

Kanda haja is it risikidha?

Translated English message:

Do you need help?

Example 2:

mixed message:

Nigumu hapa ni moja. Nimechukua kila kitu

Translated English message:

I have one here. I need everything

Because even if I eat well...

How do we use LLMs with code-mixed data?
**Dataset Statistic:**

Code-mixed SMS messages within PROMPTS are stored across two different repositories:

1. **SalesForce dataset (Unlabeled)**
   - The total number of text messages is: 939,819

2. **Freshdesk dataset (Labeled with the correct emergency level)**
   - The total number of text messages is: 107,717
   - The total number of intent (emergency level) type is: 58
TRIM-AI: Model Architecture

1. Pre-processing (Tokenization)
2. Continual pre-training
3. Multilingual pre-trained model (XLM-ROBERTa-base)
4. XLM-ROBERTa-JH

Unlabeled Code-Mixed SMS
Text from Salesforce Repository

Labeled Code-Mixed SMS Text from FreshDesk Repository

Encoder Module

Decoder Module

Message Embedding

Input

Fully-connected

Dropout

Fully-connected

Output

Classification Network

“Intent” Probability Distribution

Emergency risk score

Intent to Risk Score Mapping

“baby_cry”: 5.0, “baby_breathing”: 0.2, “baby_cough”: 0.15, “baby_sneeze”: 0.15

PennState
TRIM-AI: Model Architecture (preprocessing module)
TRIM-AI: Model Architecture (XLM-ROBERTa-JH)
TRIM-AI: Model Architecture (classification network)
### Baseline comparison:

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Weighted Precision</th>
<th>Weighted Recall</th>
<th>Weighted $F_1$-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical NN (FastText as the encoder layer)</td>
<td>0.678</td>
<td>0.668</td>
<td>0.669</td>
</tr>
<tr>
<td>Hierarchical NN (TextRNN as the encoder layer)</td>
<td>0.682</td>
<td>0.673</td>
<td>0.67</td>
</tr>
<tr>
<td>Hierarchical NN (TextCNN as the encoder layer)</td>
<td>0.638</td>
<td>0.629</td>
<td>0.626</td>
</tr>
<tr>
<td>Hierarchical NN (RCNN as the encoder layer)</td>
<td>0.679</td>
<td>0.667</td>
<td>0.663</td>
</tr>
<tr>
<td>TRIM-AI (monolingual ROBERTa-base)</td>
<td>0.730</td>
<td>0.729</td>
<td>0.728</td>
</tr>
<tr>
<td>TRIM-AI (m-BERT)</td>
<td>0.727</td>
<td>0.726</td>
<td>0.725</td>
</tr>
<tr>
<td>TRIM-AI (XLM-ROBERTa-base)</td>
<td>0.736</td>
<td>0.735</td>
<td>0.735</td>
</tr>
<tr>
<td>Vertex AI model</td>
<td>0.765</td>
<td>0.598</td>
<td>0.671</td>
</tr>
<tr>
<td><strong>TRIM-AI</strong></td>
<td><strong>0.775</strong></td>
<td><strong>0.775</strong></td>
<td><strong>0.774</strong></td>
</tr>
</tbody>
</table>
Performance Evaluation (A/B Test)

PROMPTS receives messages

~7000 messages
Performance Evaluation (A/B Test)

Option A

Vertex-AI Pipeline

PROMPTS receives messages

~7000 messages
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
- Vertex-AI Pipeline

Option B
- TRIM-AI Pipeline

~7000 messages
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
- Vertex-AI Pipeline

Option B
- TRIM-AI Pipeline

Predict & detect confidence level

~7000 messages
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
Vertex-AI Pipeline

Option B
TRIM-AI Pipeline

Predict & detect confidence level

< 0.75
Sent back to helpdesk

Agents solve it manually

~7000 messages
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
- Vertex-AI Pipeline

Option B
- TRIM-AI Pipeline

Predict & detect confidence level

- < 0.75: Sent back to helpdesk
- > 0.75: Trigger automated response

~7000 messages

Agents solve it manually

~7000 messages
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
- Vertex-AI Pipeline

Option B
- TRIM-AI Pipeline

Predict & detect confidence level

< 0.75
- Sent back to helpdesk

> 0.75
- Trigger automated response

If response answers your question?
- Yes
- Collect percentage of Yes and analyze

~7000 messages

Agents solve it manually

88% Yes

Collect percentage of Yes and analyze

Yes

If response answers your question?

Collect percentage of Yes and analyze

Yes
Performance Evaluation (A/B Test)

PROMPTS receives messages

Option A
Vertex-AI Pipeline

Option B
TRIM-AI Pipeline

Predict & detect confidence level

< 0.75
Sent back to helpdesk

> 0.75
Trigger automated response

~7000 messages

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?

No
Sent back to helpdesk

Agents solve it manually

88% Yes
Collect percentage of Yes and analyze

Yes
If response answers your question?
Performance Evaluation (A/B Test)

Comparison of A/B test between TRIM_Ai and Vertex

- **TRIM-AI** fully operational since July 2022
- **17% more accurate** at identifying high risk medical conditions
- **12% reduction** in helpdesk workload
Performance Evaluation (A/B Test)

TRIM-AI fully operational since July 2022

17% more accurate at identifying high risk medical conditions

12% reduction in helpdesk workload

90% cost savings in model management costs
Summary

01. We propose TRIM-AI, an NLP-based framework for automated assessment of a pregnant woman’s medical condition based on code-mixed messages.

02. TRIM-AI utilizes multi-lingual pre-training and continual pre-training, achieves a weighted F1 score of 0.774.

03. Real world A/B test shows that TRIM-AI is highly effective in generating high-quality predictions based on incoming SMS messages. Right now, TRIM-AI has been deployed inside PROMPTS and reduces the helpdesk workload by ~12%.
Thank You & Questions?

Wenbo Zhang & Hangzhi Guo  
(PhD Students)

Sathy Rajasekharan  
(Exec Director, Jacaranda Health)

Jay Patel  
(Tech Director, Jacaranda Health)
Beyond Buzzwords: Reimagining the Default Settings of Technology & Society

Ruha Benjamin, PhD
Professor, Department of African American Studies, Princeton University and Founding Director, Ida B. Wells Just Data Lab

Moderated by Karey M. Sutton, PhD
Scientific Director, Health Equity Research, MedStar Health Research Institute
Beyond Buzzwords: Reimagining the Default Settings of Technology & Society

Ruha Benjamin
Princeton University
Ida B. Wells Just Data Lab

AAMC Center for Health Justice
May 18, 2023
Not all speed is movement.
-Toni Cade Bambara
two stories
new vision
racism distorts how we see & how we are seen
Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites

Kelly M. Hoffman, Sophie Trawalter, Jordan R. Axt, and M. Norman Oliver
A thousand cuts: A 'Zainichi' Korean reporter's deep dive into microaggression in Japan
“Why I am in such demand as a research subject, when no one wants me as a patient?”

—From People’s Science
Racial bias in a medical algorithm favors white patients over sicker black patients

‘Significant Racial Bias’ Found in National Healthcare Algorithm Affecting Millions of People
the new jim code

; coded bias + imagined objectivity

; innovation that enables containment
Deep learning

Computational depth without historical or sociological depth?

SUPERFICIAL LEARNING
encoded inequity
discriminatory design

Spikes in digital tools?
obvious vs insidious

From the Government Gazette registration of reclassifications, 1938:

- 518 Coloured persons were reclassified as White persons
- 1 White person was reclassified as a Coloured person
obvious vs insidious

i.e. subtle but harmful
Wyden, Booker and Clarke Introduce Algorithmic Accountability Act of 2022 To Require New Transparency And Accountability For Automated Decision Systems

Legislation Requires Assessment of Critical Algorithms and New Public Disclosures; Bill Endorsed by AI Experts and Advocates; Bill Will Set the Stage For Future Oversight by Agencies and Lawmakers

BLUEPRINT FOR AN AI BILL OF RIGHTS
MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE
OCTOBER 2022
36,369 observations from 4,172 patients.

“We didn't train the algorithm to predict what the doctor was going to say about the X-ray.

We trained it to predict what the patient was going to say about their own experience of pain in the knee.”
final proposition

if inequity is woven into the very fabric of society then each twist, coil, and code is a chance for us to weave new patterns, practices, politics its vastness will be its undoing once we accept that we are pattern makers.
Crowdsourcing Exercise: Designing Principles for NLP in Maternal Health

AAMC Center for Health Justice & Allen Institute for AI
Maria Antoniak, PhD
Allen Institute for AI
NLP, bias, reproductive healthcare

Carla S. Alvarado, PhD, MPH
AAMC Center for Health Justice
health policy, public health, intersectionality, health equity

Lucy Lu Wang, PhD
University of Washington
health informatics, NLP, accessibility

Irene Chen, PhD
UC Berkeley, UCSF
ethics of machine learning for health

Aakanksha Naik, PhD
Allen Institute for AI
medical NLP, understudied language

Our Team
Assisted by an amazing team of volunteer facilitators!

Affrille Degoma
Kathryn Brand
Jennifer Bretsch
Keith Krosinky
Dallas Peoples
Tracey Alcendor Robinson

Mary Heller
Mobie Nwaokomah
Mubarak Childs
Sherrie Reece
Ebonie Megibow
Our goal: guiding principles for NLP & maternal health

Your expertise and lived experience is invaluable.

We want to hear your hopes, fears, goals, critiques — all of it!

We share many of your concerns about the use of AI and NLP, and we want to build the best possible future, together with you.
What to expect from this session

An interactive exercise with a chatbot

A moderated discussion about your experiences and opinions about NLP and maternal health

A short survey asking for your perspectives
From Data to Action: What Public Health, Hospitals and Health Systems Can Do

Simon Linwood, MD, MBA
Janette Robinson Flint
Justin Schonfeld, PhD
Carlos Siordia, PhD, MS

Moderated by Karey M. Sutton, PhD
Scientific Director, Health Equity Research, MedStar Health Research Institute
Meet the Panelists

Simon Linwood, MD, MBA  
Chief Information Officer and Chief Information Officer, University of California Riverside Health and School of Medicine

Janette Robinson Flint  
Executive Director, Black Women for Wellness

Justin Schonfeld, PhD  
Research Scientist, Public Health Agency of Canada

Carlos Siordia, PhD, MS  
Lead Interdisciplinary Epidemiologist, Division of Violence Prevention, National Center for Injury Prevention and Control
Drawing Change Workshop Summary

Yolanda Liman
Drawing Change
Thank you for joining us!
Let’s Keep in Touch

Please use the hashtags:
#healthequity and #maternalhealth
And don't forget to tag us: @AAMCjustice

aamchealthjustice.org
healthjustice@aamc.org
@AAMCjustice

Sign up for the Center for Health Justice Newsletter