

# Identifying Abusive Head Trauma in Young Children Presenting to Emergency Medical Services using a Large Language Model



<sup>1</sup> University of Colorado School of Medicine, Aurora, CO

<sup>2</sup> Oklahoma State University, Stillwater, OK

Kathleen Adelgais, MD MPH<sup>1</sup>, Allison Broad, MSIV<sup>1</sup>, Fattah Muhammad Tahabi,MS<sup>2</sup>, Denise Abdoo PhD CPNP<sup>1</sup>, Zhan Zhang PhD<sup>3</sup>, Xiao Luo PhD<sup>2</sup>

# **BACKGROUND**

Abusive Head Trauma (AHT) is a leading cause of death and disability in young children

Identification of characteristics presenting to Emergency Medical Services are often limited to structured data fields.

Large Language Models (LLM) enhanced with expert information may enhance detection of complex diagnoses such as AHT

### **OBJECTIVE**

To compare AI and LLM efficacy to detect AHT in young children in EMS narrative documentation against standard NLP and ML methods

#### **METHODS**

Data Source: ESO Health Data Exchange

(2018-2019)

**Population:** Patients <36 months, ICD10 dx of AHT

NLP pipeline extraction:

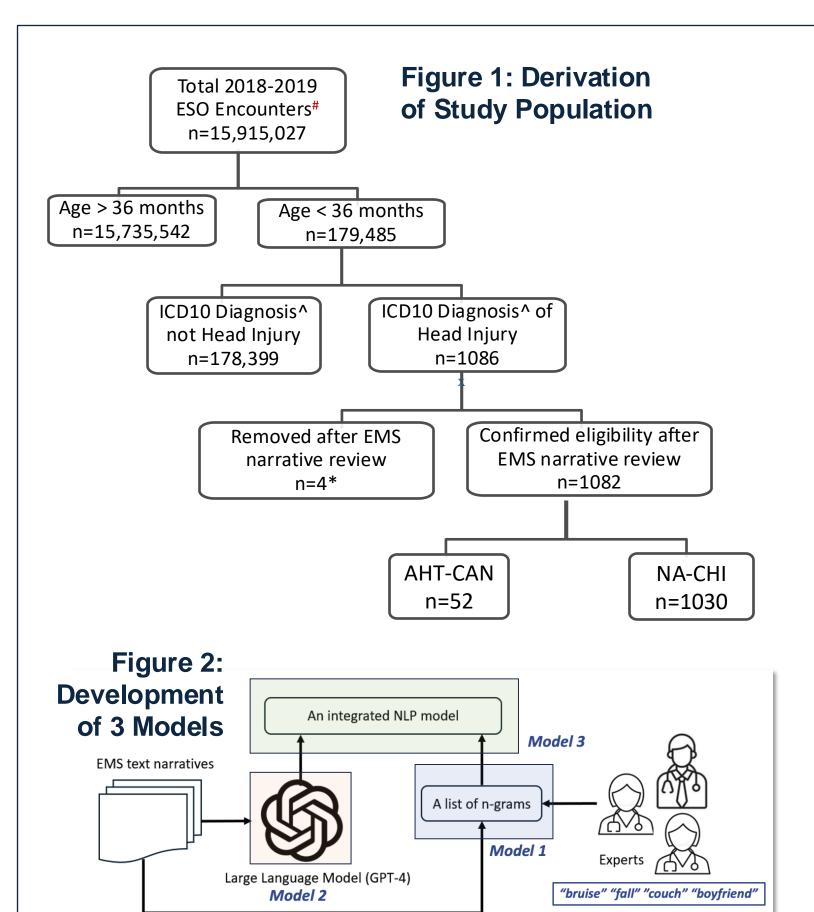
- Demographics
- Mechanism of Injury
- Ineligible encounters

Development of 3 models to detect AHT

#### Outcomes: Comparison of 3 models

- Test Characteristics
- Area Under the Receiver Operator Curve
- Area Under the Precision Recall Curve

# RESULTS



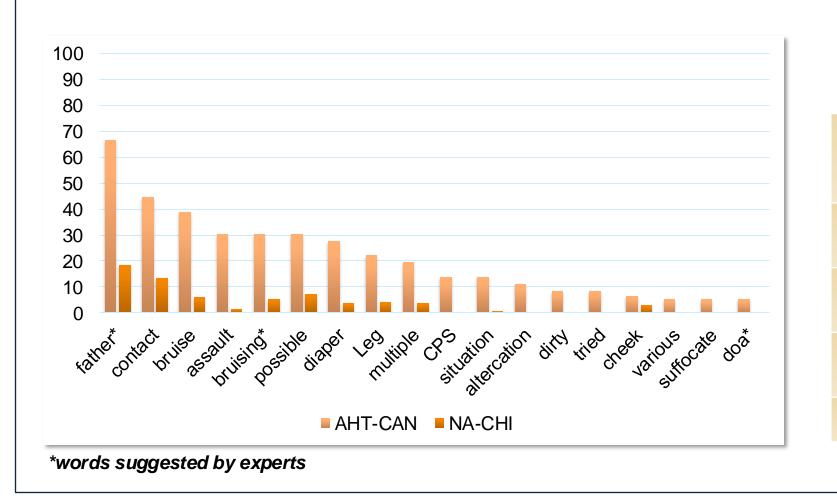
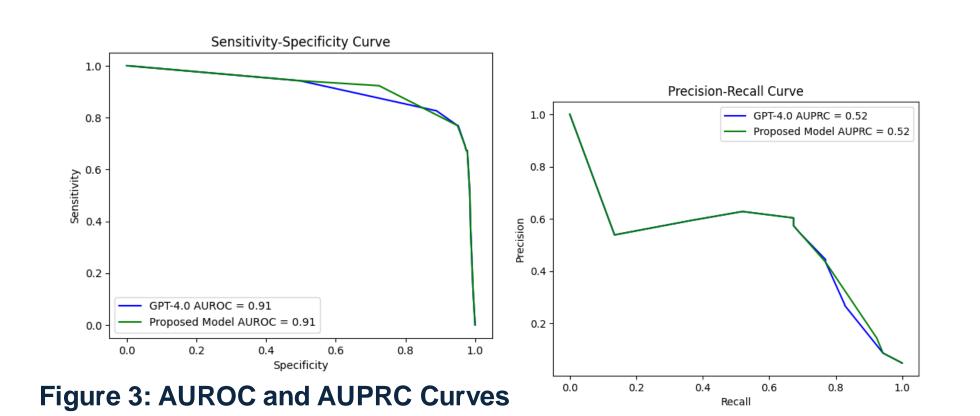


Table 1. Demographics of Study Popul		
Demographics	AHT-CAN (n=52)	NA-CHI (n=1030)
Age, months	7	12
Male Sex, n (%)	57.69%	54.27%
Race		
White	34.62%	49.03%
Hispanic or Latino	7.69%	9.32%
First GCS, median	13	14
Altered LOC/Altered Mental Status	9.62%	1.95%
Advanced Life Support	50.00%	55.48%
Emergent/Lights/Sirens	82.69%	88.33%
Admitted or Transferred	57.60%	25.40%
Hospital LOS, days, mean	10.9	3.7



**Table 2: Test Characteristics by Model** 

Table 1: Demographics of Study Population

Metrics	GPT-40	N-gram only	LLM w n- grams	•	missed cases:  MVC with intoxicated driver
Sensitivity	0.69	0.53	0.92	•	6-month-old dropped by father
Specificity	0.97	0.62	0.72	•	Infant went 'limp' BRUE-type event
PPV	0.54	0.06	0.14		288 false
NPV	0.98	0.96	0.99		positive cases

# **DISCUSSION**

- LLM with n-grams model algorithm performed best (AUROC 0.91, specificity 72 %, and sensitivity 92%)
- Words alone are not sufficient to detect AHT in EMS narratives (specificity 62 %, and sensitivity 53%)
- Need to mitigate risk of implicit bias in documentation given high false positive rate
- Future work includes feasibility testing of embedding machine learning into ePCR software and impact on EMS clinician decision making

#### CONCLUSIONS

- All augmented NLP has high sensitivity and specificity to detect AHT in EMS free-text narratives
- Machine learning using natural language processing may help identify sentinel injury detection in EMS encounters of young children

#### CORRESPONDENCE

Kathleen Adelgais, MD MPH
Pediatric Emergency Medicine
kathleen.adelgais@childrenscolorado.org

<sup>&</sup>lt;sup>3</sup> Pace University, New York, NY