UTHealth
The University of Texas
Health Science Center at Houston
School of Biomedical Informatics

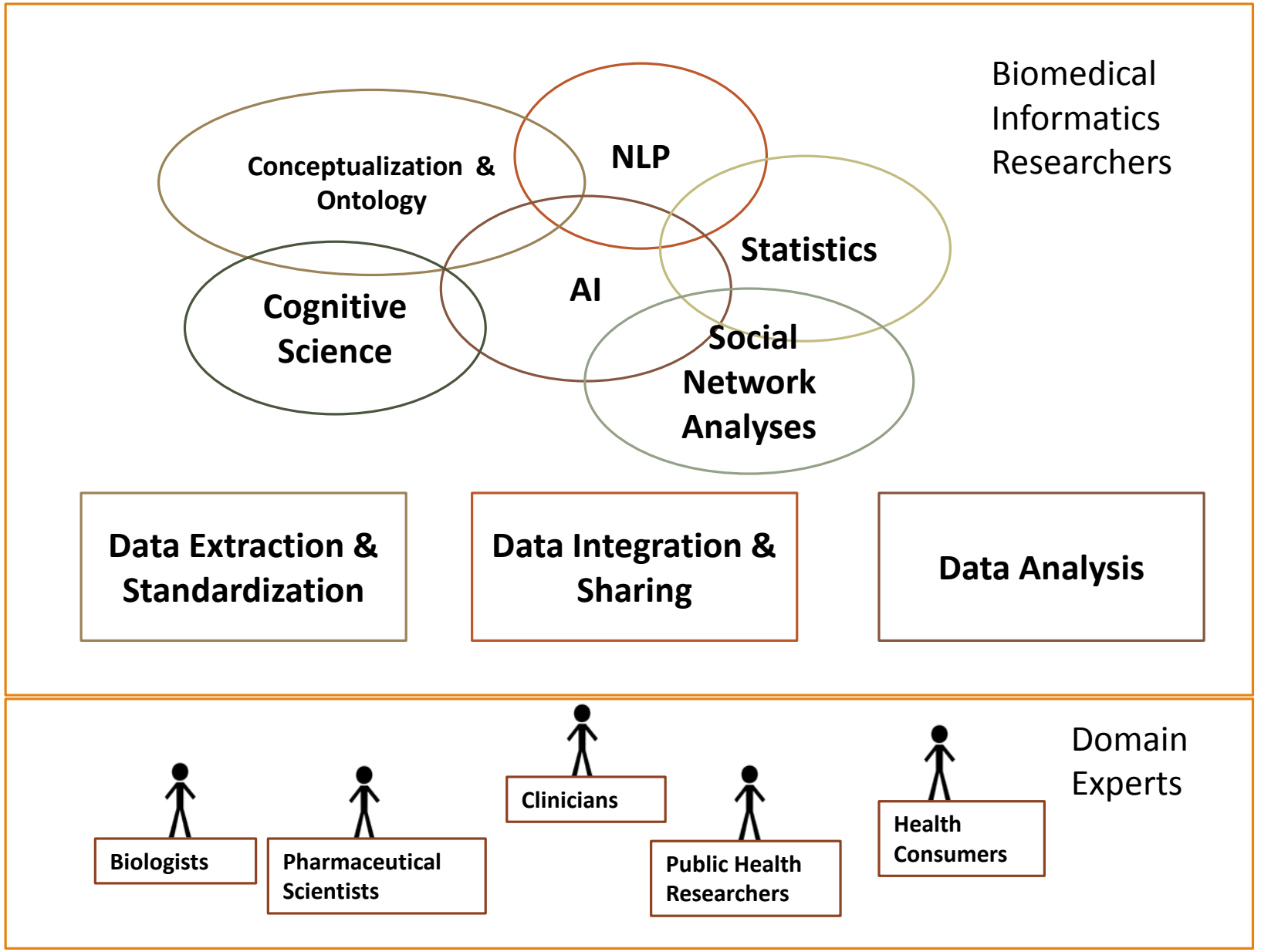
UTHealth
The University of Texas
Health Science Center at Houston
School of Public Health

Modeling Influential Neighbors in Social Networks of Young MSM: An Application of Graph Attention Networks to Predict and Track HIV Infection

CUI TAO, PHD, FACMI

THE UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER AT HOUSTON

An Ecosystem for Data Centered Informatics Research


Biomedical Informatics Researchers
Conceptualization & Ontology
NLP
Statistics
AI
Cognitive Science
Social Network Analyses
Data Extraction & Standardization
Data Integration & Sharing
Data Analysis
Biologists
Pharmaceutical Scientists
Clinicians
Public Health Researchers
Health Consumers
Domain Experts

Three major steps

Data Extraction (Text-mining and Natural Language Processing)

Data Standardization & Integration (Ontology and Semantic Technologies)

Data Processing and Analyses (Machine learning, Statistical, social network analysis)

AI-assisted HIV prevention/intervention

Background

1981 to 2015 in the U.S.

HIV infected: > 1.8 million
Died: 650,000

2015 in the U.S.

Men who have sex with men (MSM) in high risk of HIV infection: 30.7%

2016 in the U.S.

Young MSM (20-29) in new infections: 37.2%

HIV is costly with about $29.5 billion spent in 2014 and $32.8 billion in 2017 in the US.

Between 2010 and 2016, the annual number of HIV infections decreased but the number of YMSM remained stable or increased.

Background

- President Trump announced his Administration's goal to end the HIV epidemic in the United States within 10 years
- Intervention: Early identification of newly HIV infected cases is essential
  - 162,500 (14%) persons whose infections had not been diagnosed
  - Many persons with HIV are tested late in the course of infection
  - Late testing results in missed opportunities for prevention and treatment of HIV and increased risk for transmission to their partners
- Prevention: PrEP: (pre-exposure prophylaxis) is a daily medication taken to prevent HIV infection
  - Public perception, awareness, misinformation, lack of education
  - High risk population recruitment

AI-assisted HIV prevention/intervention

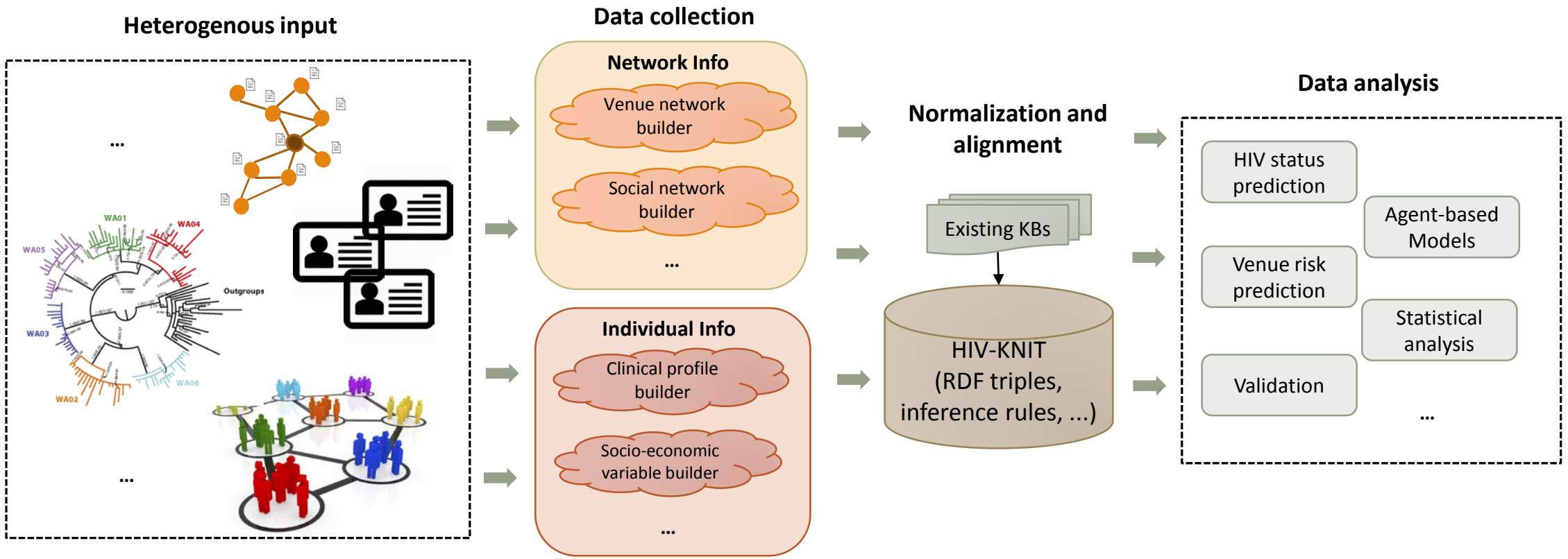
Deep learning based HIV infection risk prediction

Public perception analyses using social media

Electronic methods to recruit persons at high risk of HIV infections

AI-based personalized education and intervetion

Deep learning based HIV infection risk prediction



Data

Young Men's Affiliation Project (YMAP), includes individuals of YMSM, aged 16 to 29, in Houston and Chicago.

Data in this study came from 378 YMAP baseline participants in Houston collected between 2014 and 2016 using the respondent-driven sampling (RDS) method.

Our data includes combined social network information from peer referral and social/sexual partners, HIV testing results and survey data (sociodemographic, sexual behavior).

Motivation

Social network drives the individual HIV risks.

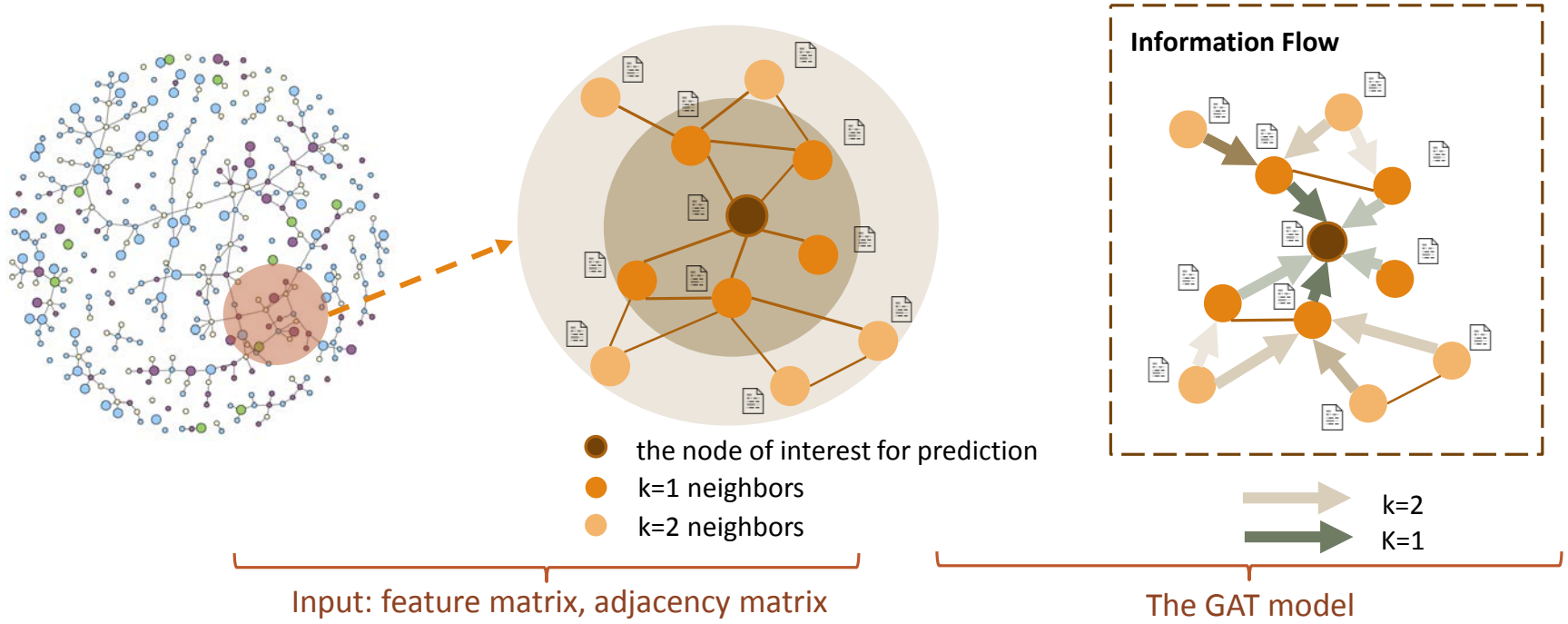
To make predictions upon the risk of HIV infection on each individual and perform early preventions.

Assumption

- Apart from individual-level variables, social network context also contributes significantly on HIV risks.
- Network neighborhood also provide important information for the prediction and influential neighbors should be identified.

Method

Graph Attention Network (GAT)



Input

- Graph: adjacency matrix
- Individual features: socio-demographic variables (race/ethnicity, age, etc.), sexual behaviors (condom use, number of recent sex partners, etc.), graph characteristics (diagonal, centrality, etc.)

Result

Evaluation metric: Area under the ROC Curve (AUC)

Baselines: Logistic Regression (LR), Random Forest (RF), Graph Convolutional Network (GCN) and Multi-layer Perceptron (MLP). Five independent realizations of train-test splits.

GAT outperforms baselines.

	AUC	Std (±)
LR	72.17	4.59
RF	75.59	3.73
GCN	57.34	5.84
MLP	67.50	6.04
GAT	80.6	6.56

Conclusion

We illustrated the capability of the GATs to increase the prediction of HIV and tracking the propagation of the disease in social network of YMSM.

Our study demonstrated the importance of focusing on the most influential neighbors on better predicting individual's HIV status.

Our study findings inform effective network-based interventions through identifying most influential neighbors.

Next step: Investigating the scalability of the machine learning model across cities.

AI-assisted HIV infection prevention/intervention

Deep learning based HIV infection risk prediction

Public perception analyses using social media

Electronic methods to recruit persons at high risk of HIV infections

AI-based education

?

Age group

Race & ethnicity

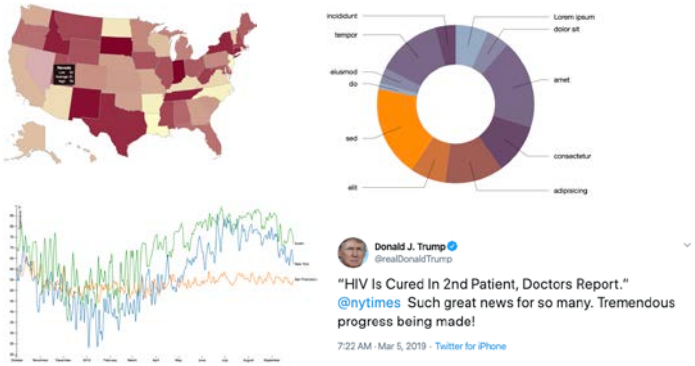
?

Gender

Location

Sex orientation

User attributes inference

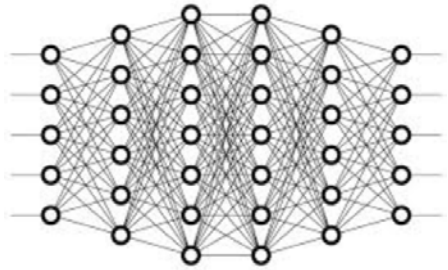


Real-time surveillance

Sentiment analysis

Health Belief Model

Misinformation identification



Deep learning-based content analysis

Personalized intervention through tailored message

AI-assisted HIV prevention/intervention

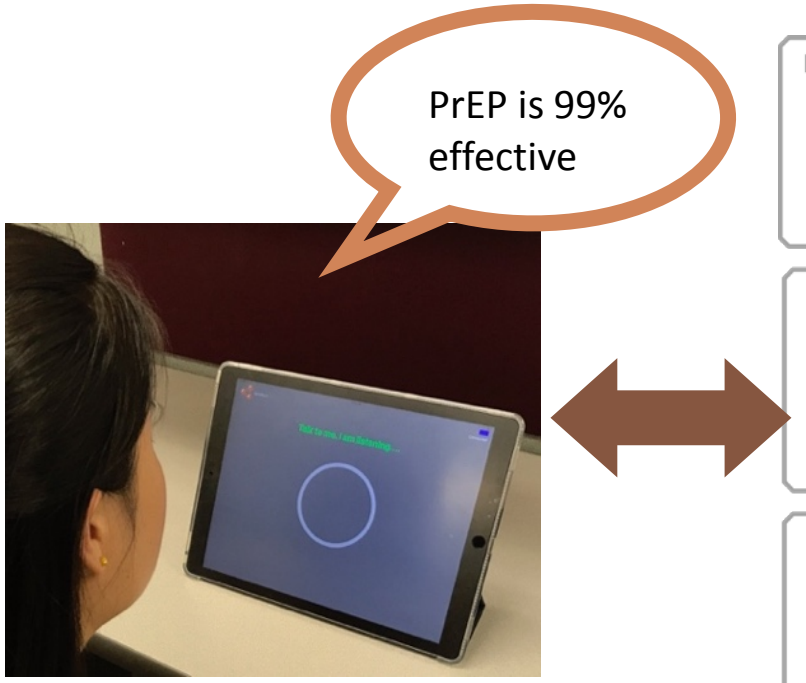
Deep learning based HIV infection risk prediction

Public perception analyses using social media

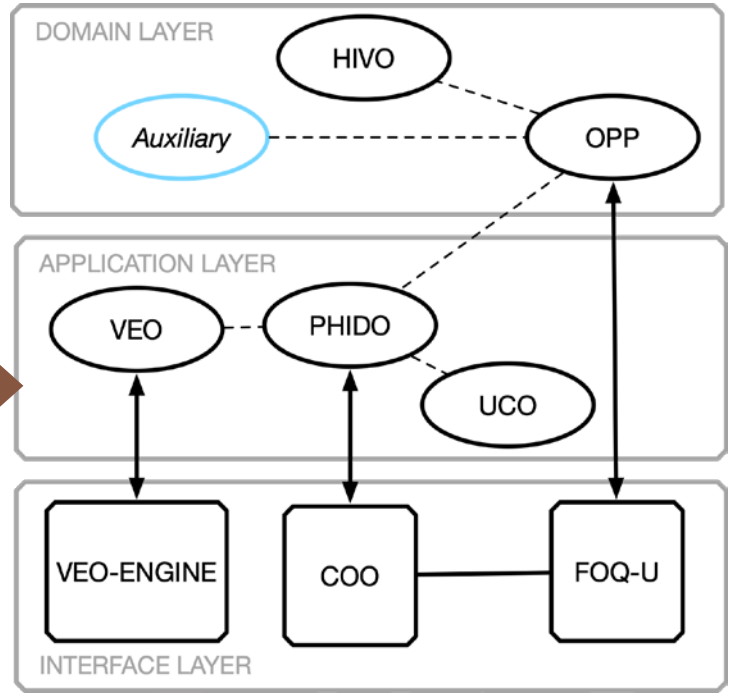
Electronic methods to recruit persons at high risk of HIV infections

AI-based personalized education

Mobile Health Dialogue System



Natural language interface



Ontology-based software architecture

Ontology-based software architecture

Ontologies

-HIV Ontology (HIVO)
-Ontology of PrEP & PEP (OPP)
-Patient dialogue ontology (PHIDO)
-Visualized Emotion (VEO)
-User Content Ontology (UCO)

Software Component

-VEO-Engine
-Conversational Ontology Operator (COO)
-Question-answering component (FOQ-U)

…. AND THERE'S MORE

Acknowledgement

Dynamic learning for post-vaccine event prediction using temporal information in VAERS. (R01AI130460, Contact PI Tao)

Metadata applications on informed content to facilitate biorepository data regulation and sharing. (U01HG009454, PI Tao)

Patient Medical History Representation, Extraction, and Inference from EHR Data. Funded by NIH/NLM. (R01LM011829, PI Tao)

AI-aided Personalization on Dual Antiplatelet Therapy Duration for Patients Underwent Coronary Stent Implantation Funded by AHA (19GPSGC35180031, PI: Tao)

YMAP (R01MH100021, PI: Fujimoto/Schneider).

caCDE-QA: A quality assurance platform for cancer study common data elements. Funded by Mayo Clinic/NIH/NCI. (U01 CA180940, UTHealth PI Tao), Biocaddie: Biomedical And Healthcare Data Discovery And Indexing Engine Center (U24AI117966, PI Ohno-Machado)

Advancing Cancer Pharmacoepidemiology Research through EHRs and Informatics (U24CA194215, Contact PI Xu)

Interactive machine learning methods for clinical natural language processing (2R01LM01068, PI Xu)

Modeling & Analysis Of Clinical Care For HIT Improvement. Funded by University Of Washington/AHRQ. (R01 HS021233, UTHealth PI Tao)

Cancer Prevention Research Institute of Texas (CPRIT) Training Grant #RP160015

SHARPc and SHAPRn funded by ONC

WHO
WHAT
WHERE
WHEN
WHY
HOW
QUESTIONS
ANSWERS