

#### VITAL STRATEGIES + FRAMENET BRASIL

Leveraging data linkage and frame-based textual analysis for the identification of candidate cases prone to suffer from **GBV** in territories

## DATA PROBLEM

The high percentage of Gender-Based Violence (GBV) underreporting in Brazilian municipalities due to:

victims' fear of talking about the violence experienced;

health professionals' difficulties in identifying violence episodes



lack of integration in Brazilian public health information systems







lack of instruments to support Health Teams

VITAL STRATEGIES + FRAMENET BRASIL

# Data Retrieval Pipeline

Data for going beyond stereotypes





Brazilian Public Health Systems store data of different types about people living in a given territory. Part of this data is parameterized, part takes the form of open text fields.

## THE SCENARIO



#### **STAGE 1 Data Collection**





## **STAGE 3** Frame Analysis

Semantic frames for gender-based violence case identification in medical records; Innovative knowledge about genderbased elements sufficient to structure alert systems to support medical professionals in identifying cases of violence that the victim does not refer; Innovative inputs for prioritizing health strategies in preventing genderbased violence



## **OUTCOMES**

Strengthening of the public health surveillance strategies for GBV



The municipality may require additional time to make the data available, tempering with the timeline;





available, tempering with the timeline;

variables may show low completeness;

The data may be sparse considering the municipality population and the

the output of the analyses, impairing

# Risks



The municipality may require additional time to make the data available, tempering with the timeline;

The quality of the data for some variables may show low completeness;

The data may be sparse considering the municipality population and the % using the public health system;

The quality of the data may harm the output of the analyses, impairing the usability of the final tool;

The inferences provided by the AI may put victims or candidate victims in embarrassing or vulnerable situations.

# Risks

## Mitigation Strategies

Reordering some steps in the project timeline once we already
have parts of the necessary; datasets

Adding more than one municipality to the project.

Adding more than one municipality to the project once there are national patterns, but the data quality can differ in each region.

Adding more than one municipality to the project. If they don't have medical record, it can be used for augmenting the original dataset

5 Elaboration of strategies for the anonymization of women with a minimum loss of potential in the final tool;

Conduction of focus groups comprising healthcare and social services professionals



## **OVERVIEW**

#### **TIMELINE & MILESTONES**

MAR. 23	APR. 23	MAY. 23	<b>JUN 23</b>	JUL. 23	AUG. 23

#### ACTIVITIES

Data collection – municipality will make data available through a cloud environment

Cleaning and organization of databases

Focus groups to gather qualitative information and guarantee data usage ethics

Linkage execution

#### **Tactical Roadmap Revision**

Completion of the paired database

**OUTPUT 1:** Dataset from the linkage of public health systems of one or two Brazilian municipalities

Sample data analysis and annotation for identification of GBV-related frames and lexical items associated with them, as well as for flagging potential instances of PII in the open text fields

#### **TIMELINE & MILESTONES**

MAY. 23	JUN. 23	JUL. 23	AUG 23	SEP. 23	OCT. 23

#### ACTIVITIES

Adaptation of the FrameNet Brasil database for the representation of the GBV lexicon

**OUTPUT 2:** Descriptive analysis of the linked dataset focused on the GBV phenomenon

**Mid-Point with Program Lead** 

#### **Interim Grant Report**

**OUTPUT 3:** Descriptive analysis of the linked dataset focused on the data quality and the services provided

Annotation of text data with frames and frame elements for building the gold standard dataset

Assessment of ML methods and techniques for suitability to the problem and dataset

Test-implementation of selected algorithms

Ablation studies for defining the role of different data structures in the results

#### **TIMELINE & MILESTONES**

	OCT. 23	NOV. 23	DEC. 23		
_					

#### **ACTIVITIES**

**OUTPUT 4:** Revised and expanded model of the lexicon associated with the domain of GBV, defined in terms of computationally implemented semantic frames

#### Analysis

Final implementation of selected algorithm

**OUTPUT 5:** Extendable, general framework for GBV risk classification of territories, based on public health records and their respective links and semantic information

**OUTPUT 6:** Territorialized data visualization dashboard to instruct public policy - to be available for the local partner surveillance health department

Project wrap-up

#### **Final Insight Report**

**Final Activities:** Technical report; Blog #2 Final Grant Report with the code library

# Thank you!

