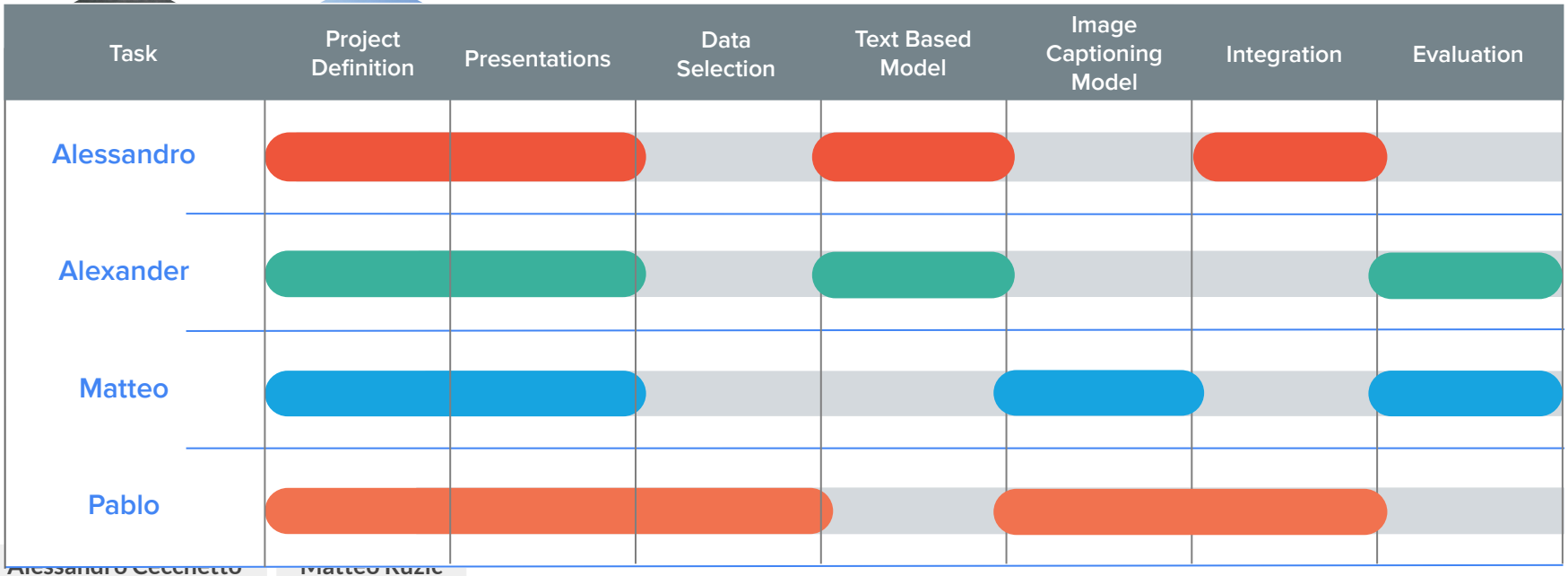
Early Emergency Detection using Social Media

DEIB - Dipartimento di Elettronica, Informazione e Bioingegneria
Supervised by Prof. Barbara Pernici and Carlo Bono

Multidisciplinary Project, Spring Semester 2023

Team and Collaboration



MSc ICT Engineering, Business & Innovation

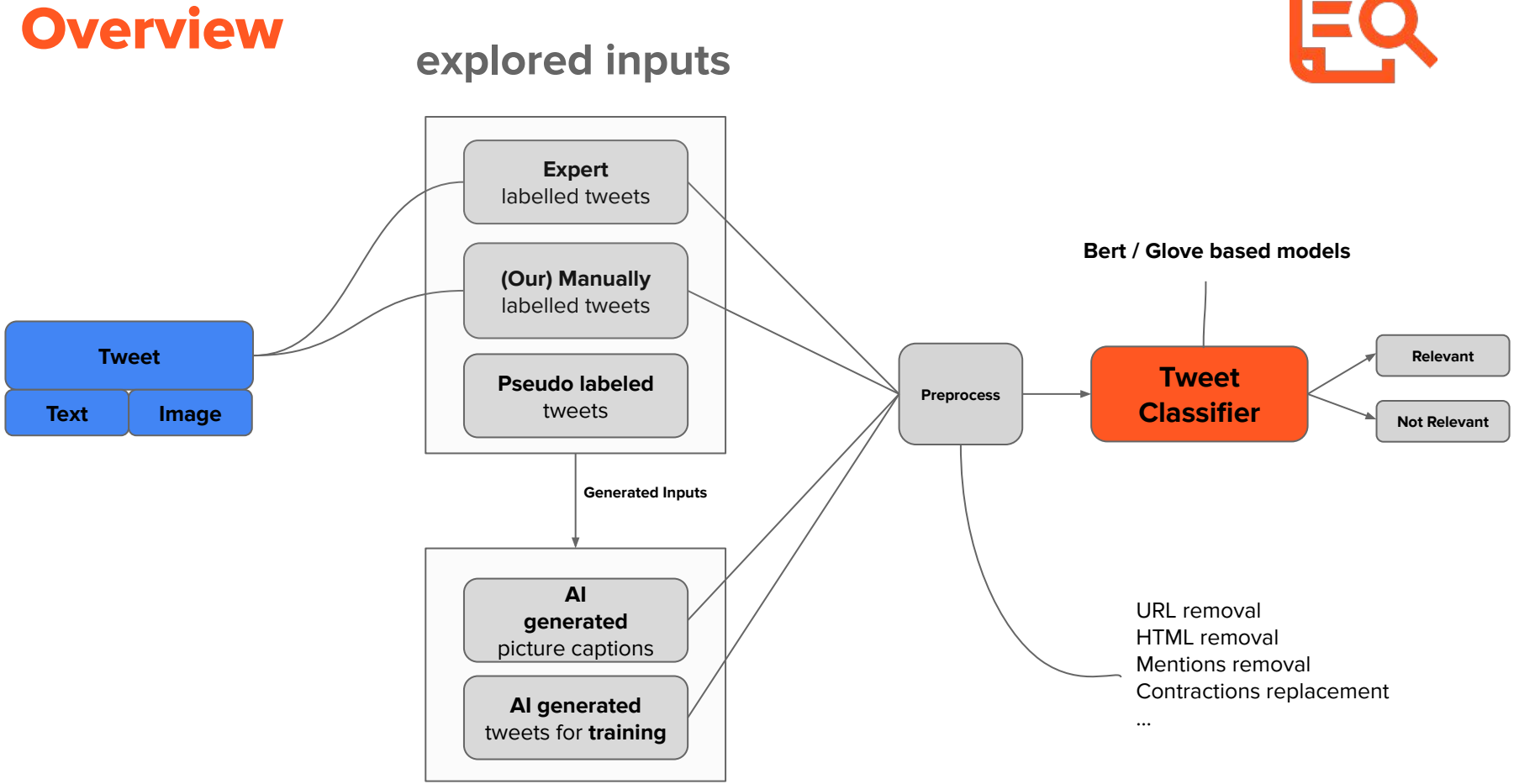
Recap: Context and Goal

- **Early Detection and Rapid Response (EDRR)** to natural disasters is a **laborious task**
- Importance of **rapid detection** for early response
- Tweets can provide a **low-latency** source of first-hand information
- Numerous possibilities to make use of them for improval of situation
  - **Detect beginning** of a catastrophe
  - Follow situation without being on site
  - **Locate hotspots**, even without geo-data

Development Process and Challenges

- **Challenges and Problems** faced during planning and development phase:
  - Complicated data acquisition and quality
  - Models very prone for being deluded
- Scope and shape of project developed during semester
  - More research-oriented approach, using additional data sources
  - Performing exhaustive analysis of different models and data handlings
  - Dropping idea of end-user interface

Overview


explored inputs
Expert
labelled tweets
(Our) Manually
labelled tweets
Pseudo labeled
tweets
Tweet
Text
Image
Generated Inputs
AI
generated
picture captions
AI generated
tweets for training
Bert / Glove based models
Preprocess
Tweet
Classifier
Relevant
Not Relevant
URL removal
HTML removal
Mentions removal
Contractions replacement
...

Implementation Details (1/2)

Explored inputs

Expert
labelled tweets

- **CrisisMMD** Dataset
- Strictly related to **Hurricane Harvey**
- **3991** Tweets
- Tweet is relevant if useful for **“Humanitarian aid”**

AI generated
picture captions

- **Captions** of Tweets’ images
- Done with **BILP Vision-LLM**
- High quality descriptions
- Expensive model: **80GB of RAM** required

(Our) Manually
labelled tweets

- **Double-agreement** labels
- Not always easy to get the “humanitarian aid” definition
- **Distribution shift w.r.t CrisisMMD** worsened model performance

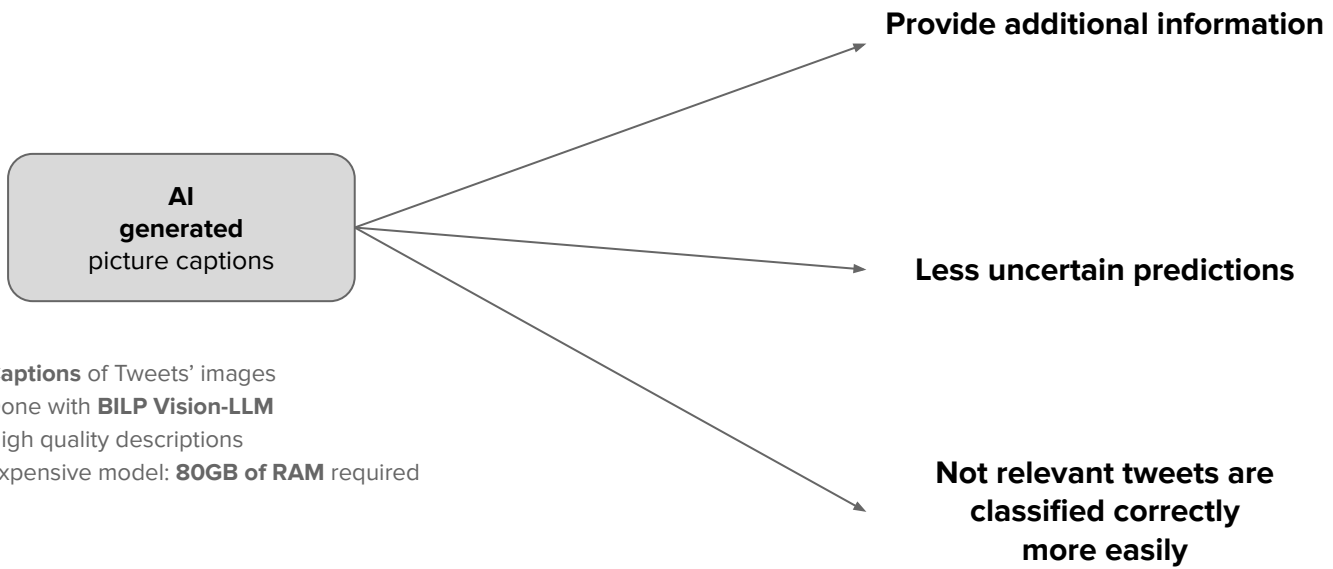
AI generated
tweets for **training**

- **ChatGPT** based generation
- Data augmentation of training set
- **Didn’t solve** class imbalance
- Worsened model performance

Pseudo labeled
tweets

- Aimed to **enlarge** training set
- **Didn’t solve** class imbalance
- A lot of **false positives**
- Worsened model performance

Implementation Details (2/3)



- **Captions** of Tweets' images
- Done with **BILP Vision-LLM**
- High quality descriptions
- Expensive model: **80GB of RAM** required

Implementation Details (3/3)

Tweet Classifier - explored models

Bert-based models

- **Bert (base uncased)** based classifier
- **roBERTa (disaster tweets fine-tuned)** based classifier
- **High performance** & computational requirements

Glove-based models

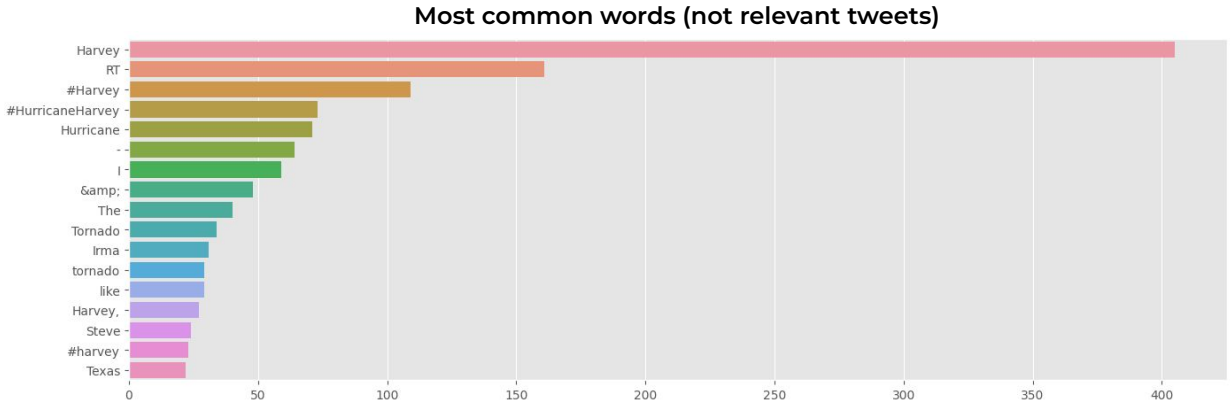
- **Glove (27B Twitter**) based classifier
- Better for a **fast and inexpensive** training & inference

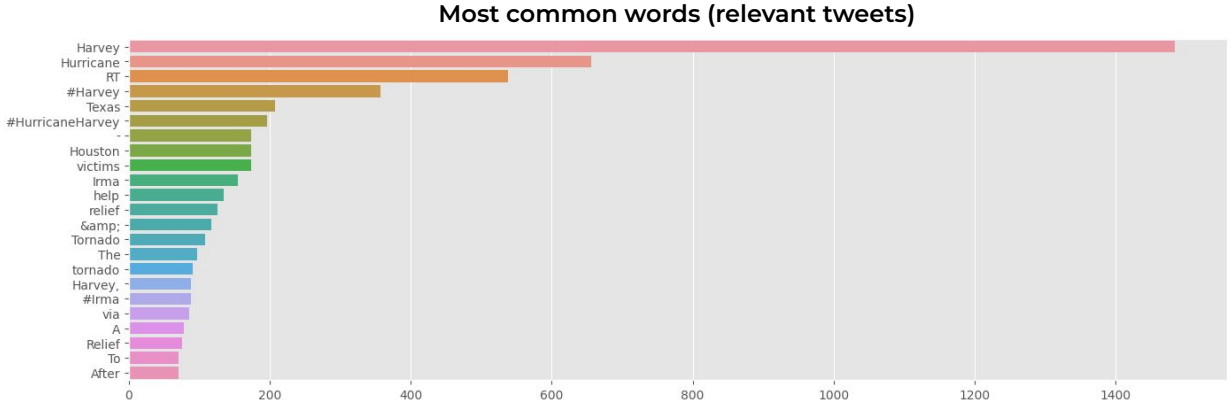
Results and Evaluation

- **Pros:**
    - The model performs very well at detecting relevant tweets since they are the majority in the training data set and they have a very small vocabulary
- **Cons:**
    - The model has difficulty detecting “not relevant” tweet since they are less and they have a very similar vocabulary to the relevant ones

The main cause of the similarity in the vocabulary is that there is no clear definition of relevant

The Vocabulary Problem



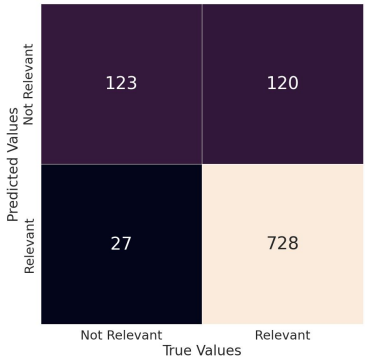


Results Table

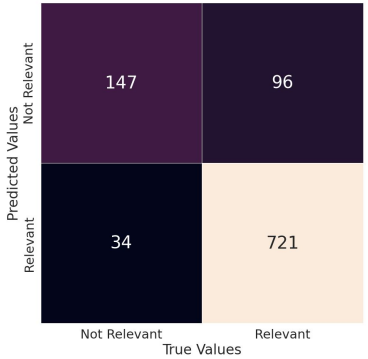
Model	Precision (Not Relevant)	Precision (Relevant)	Recall (Not Relevant)	Recall (Relevant)	Accuracy
Glove Twitter	0.5015	**0.9461**	0.7482	0.8561	0.8386
Glove Twitter (with image captions)	0.6049	0.9350	**0.7500**	0.8802	0.8547
Bert Base	0.7252	0.8639	0.5432	0.9337	0.8386
Bert Base (with image captions)	0.7849	0.8805	0.6000	0.9470	0.8627
Bert Twitter	**0.8200**	0.8584	0.5061	**0.9642**	0.8527
Bert Twitter (with image captions)	0.8125	0.8824	0.6049	0.9549	**0.8697**
Bert Base (with image captions & synthetic tweets)	0.7538	0.8804	0.6049	0.9364	0.8557
Bert Twitter (with image captions & synthetic tweets)	0.7790	0.8751	0.5802	0.9470	0.8577

Confusion matrices

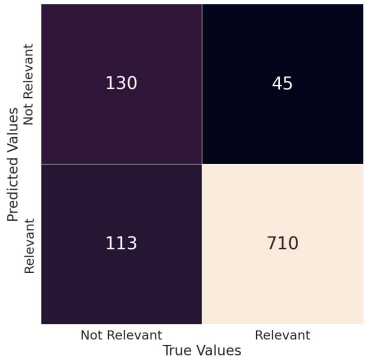
Bert Twitter



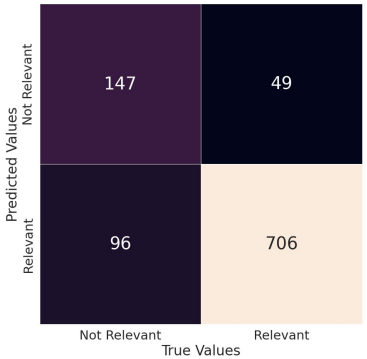
Bert Twitter (with image captions)



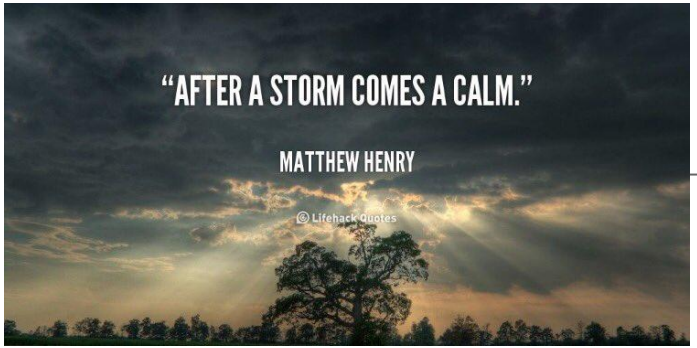
Glove Twitter



Glove Twitter (with image captions)



AI Captions: an example



BILP Vision-LLM

“The image attached to the tweet depicts a stormy sky with a tree in the foreground and the words after a storm comes a calm”

Learnings

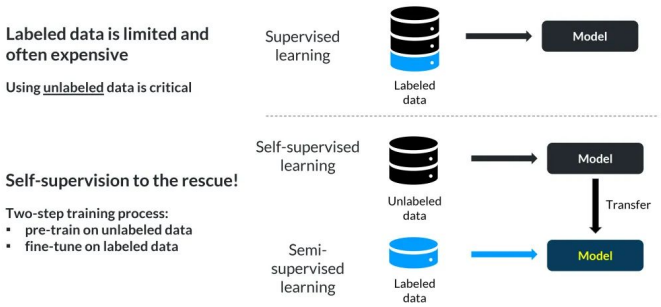
- Manual labelling must be conducted following a **strict policy**

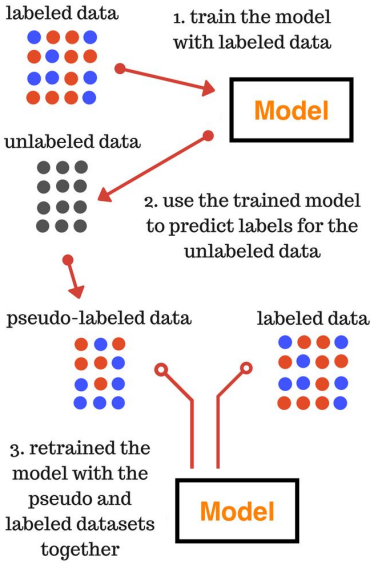
- Our LLMs slightly decreased performances due to interpretation of edge cases

- A wider, **more heterogeneous** dataset is crucial for a real world application

Possible improvement

- Use models with bigger and more heterogeneous dataset
- Try multi language model
- Try pseudo-labeling with less biased dataset
- Use other metadata
- Time series pipelining

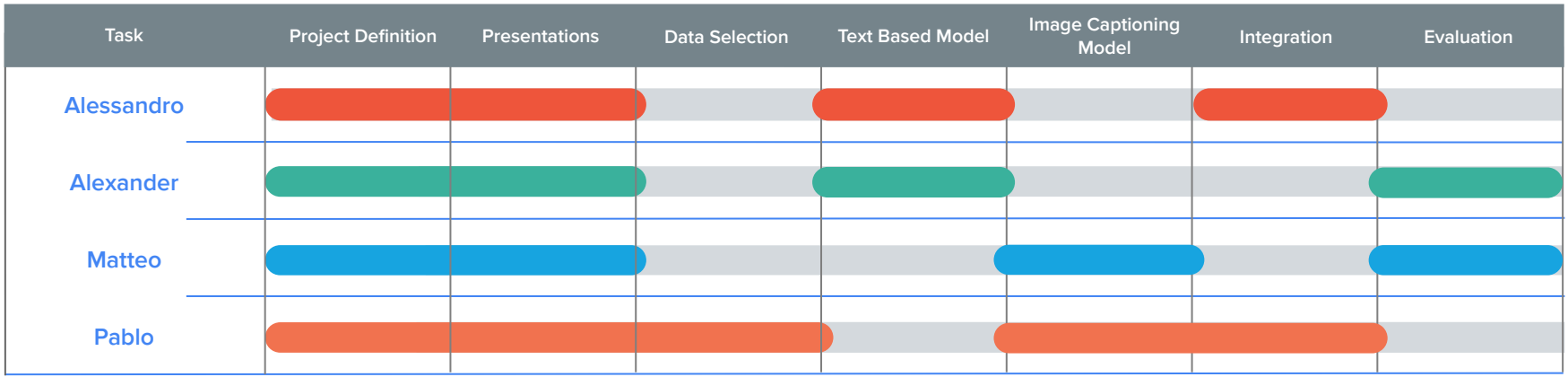




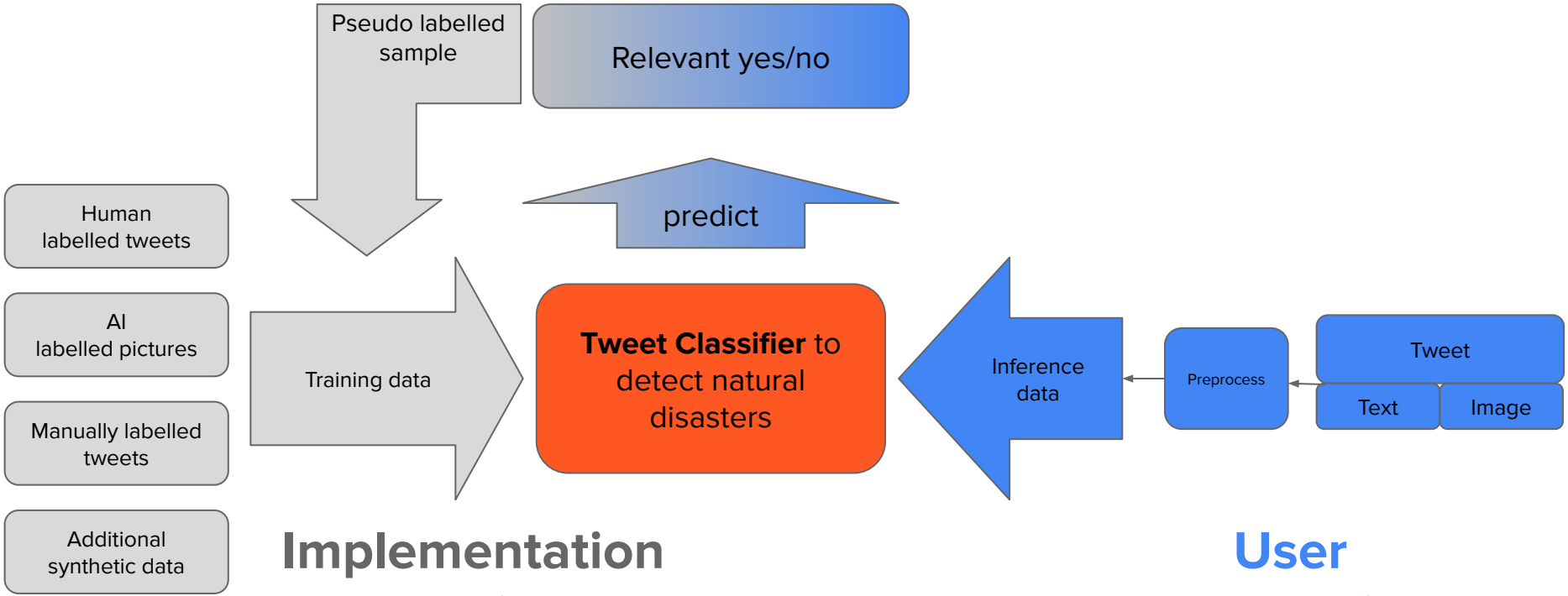
Thank you
for your attention!

Team Collaboration

- Constantly collaborating
    - Zoom for **meetings**, GDrive, Colab and GitHub for **code** and **data**
- Teamwork coordination **coherent with initial plans**



Shape of the project



Implementation Perspective

User Perspective