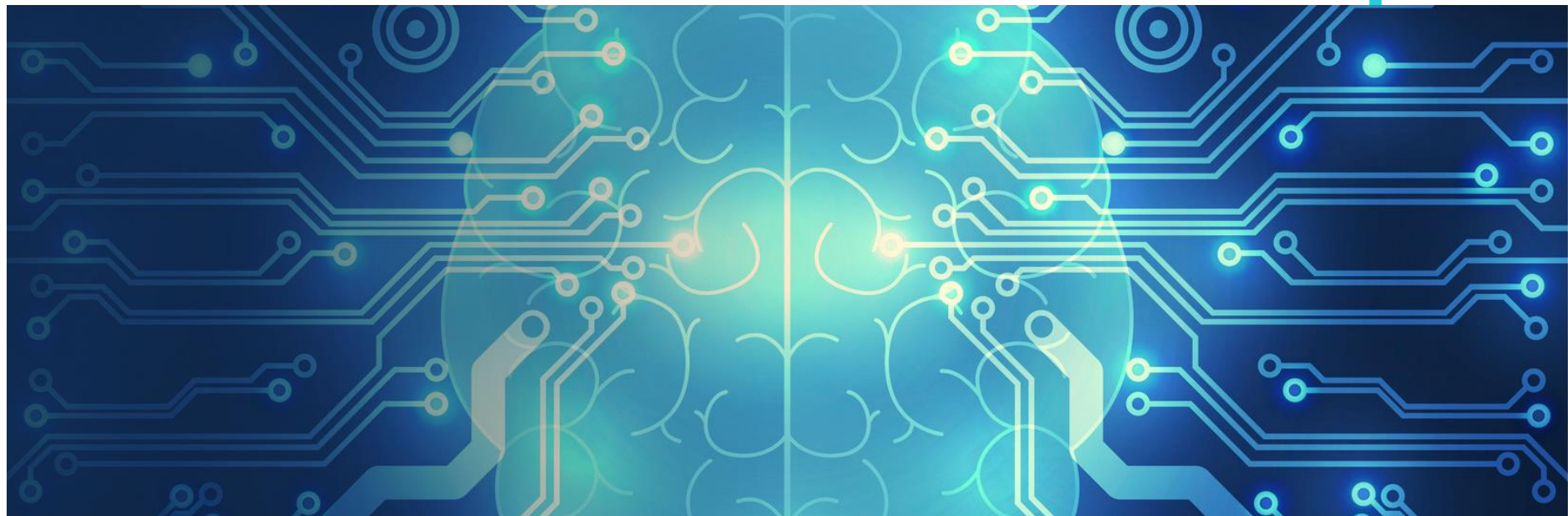


# Kaggle Toxic comment classification

**Presenter: Kiamehr Rezaee**  
**Instructor: Dr. Pilehvar**

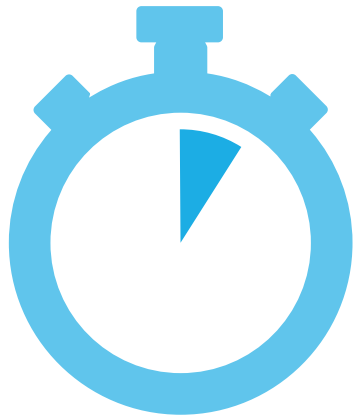


دانشگاه علم و صنعت ایران May, 30, 2020



# Road Map

**The problem**



**5 min**

What is the problem?

**Main Challenge**



**5 min**

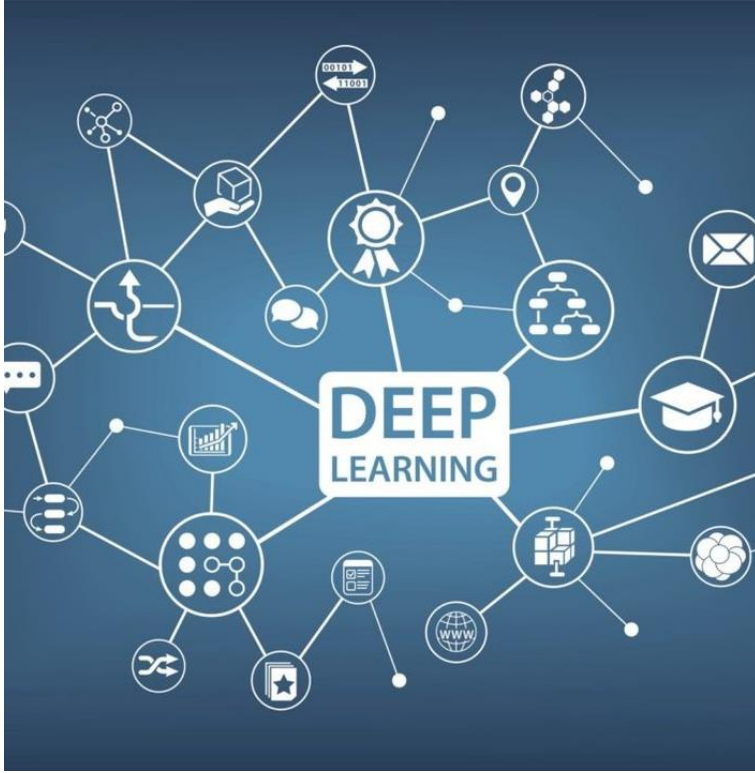
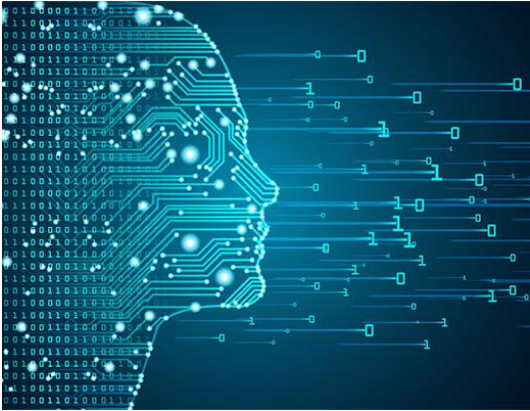
What is the main challenge?

**Implementation**

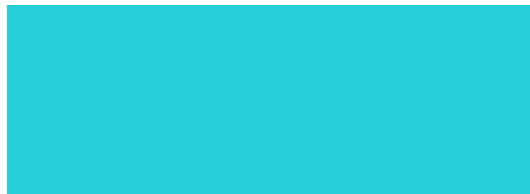
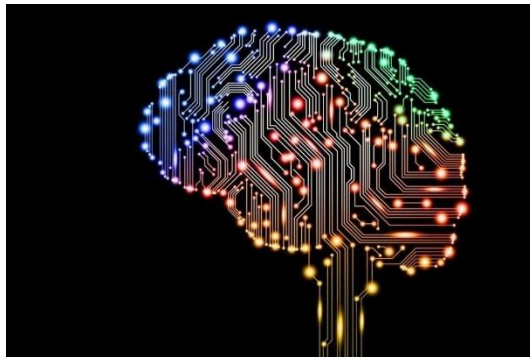


**5 min**

Implementation details



**Problem**  
What is the problem?



# Problem

## Definitions

Toxicity is defined as anything rude, disrespectful or otherwise likely to make someone leave a discussion

It only takes one toxic comment to sour an online discussion

A main area of focus is machine learning models that can identify toxicity in online conversations

we could have a safer, more collaborative internet

Definition  
Toxic  
Comments

# Problem

Definition

Kaggle Challenge

kaggle™

An online community of data scientists and machine learning practitioners

- Enter competitions to solve data science challenges
- Find and publish data sets
- Explore and build models in a web-based data-science environment

# Problem

Definition

Kaggle Challenge

Jigsaw Multilingual Toxic Comment Classification



Given a document



Classify it as  
toxic or non-toxic

# Problem

Definition

Kaggle Challenge

## Jigsaw Multilingual Toxic Comment Classification



You are stupid!

I went home.



# Problem

Definition

Kaggle Challenge

## Jigsaw Multilingual Toxic Comment Classification



You are stupid!

I went home.





# Problem

Definition

Kaggle Challenge

## Jigsaw Multilingual Toxic Comment Classification



You are stupid!

I went home.



# Problem

Definition

Kaggle Challenge



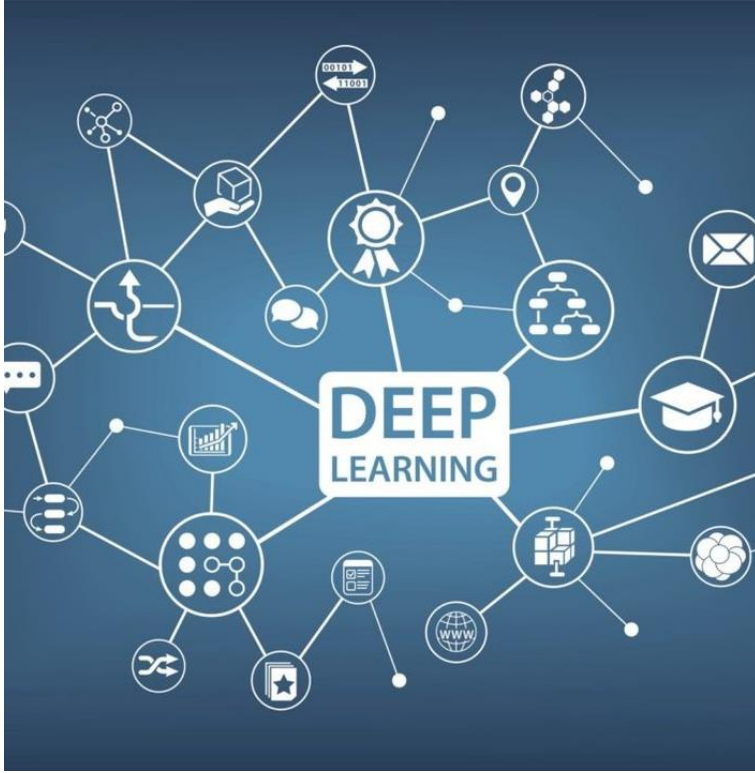
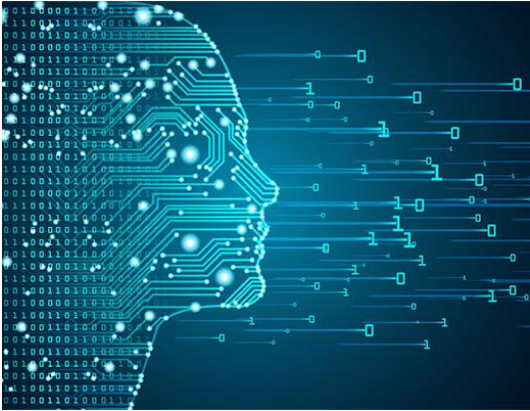
Multilingual

Classification



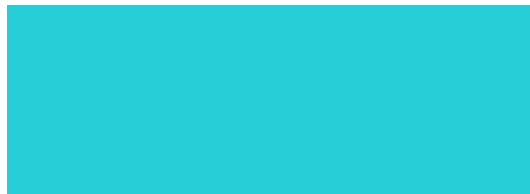
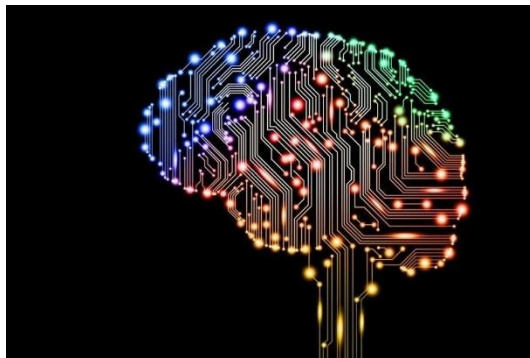
Multilingual BERT





# Main Challenge

What is the main challenges?



# Challenges

Transfer Learning

English  
Training Set

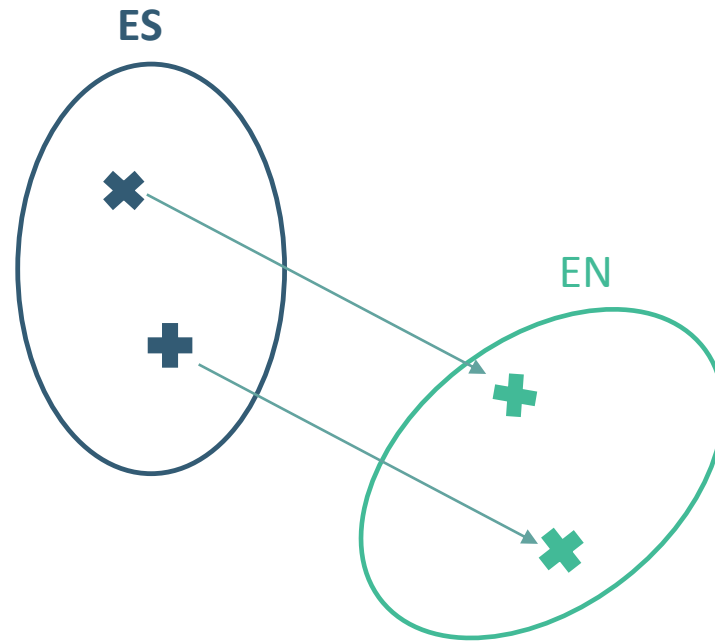
Non-English  
Test Set



# Challenges

Transfer Learning

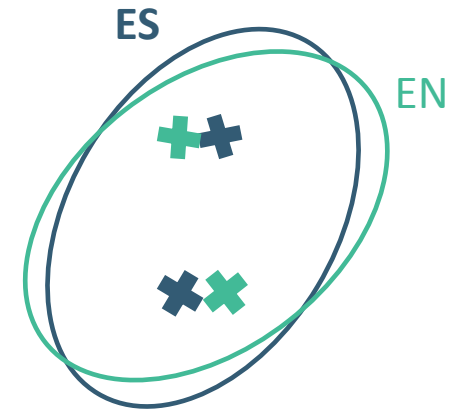
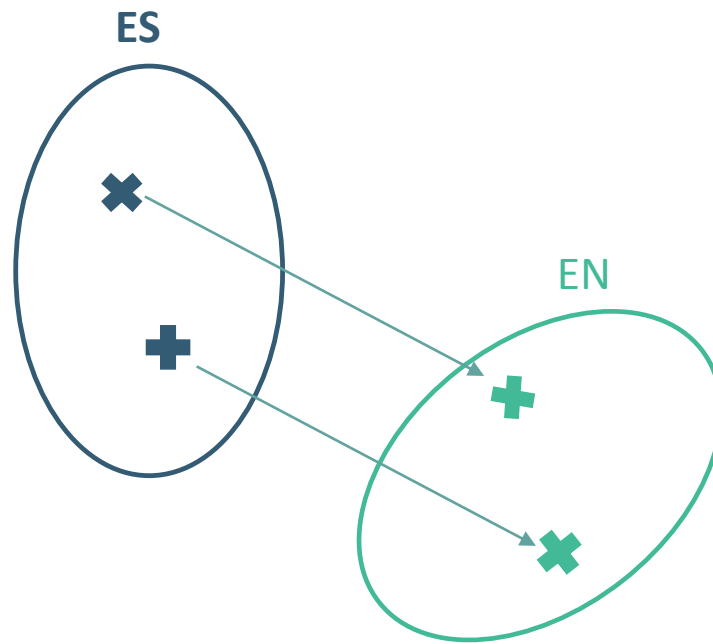
Embedding Space  
Alignment



# Challenges

Transfer Learning

Embedding Space Alignment



# Challenges

Transfer Learning

Embedding Space  
Alignment

Contextual Alignment  
Pairs



**multilingual encyclopedic dictionary**

16 million entries

284 languages

very large network of semantic relations

# Challenges

Transfer Learning

Embedding Space  
Alignment

Contextual Alignment  
Pairs

Finding the transformation  
matrix using singular value  
decomposition method



Extracting BERT CLS  
token embeddings of  
these definitions as  
aligned pairs

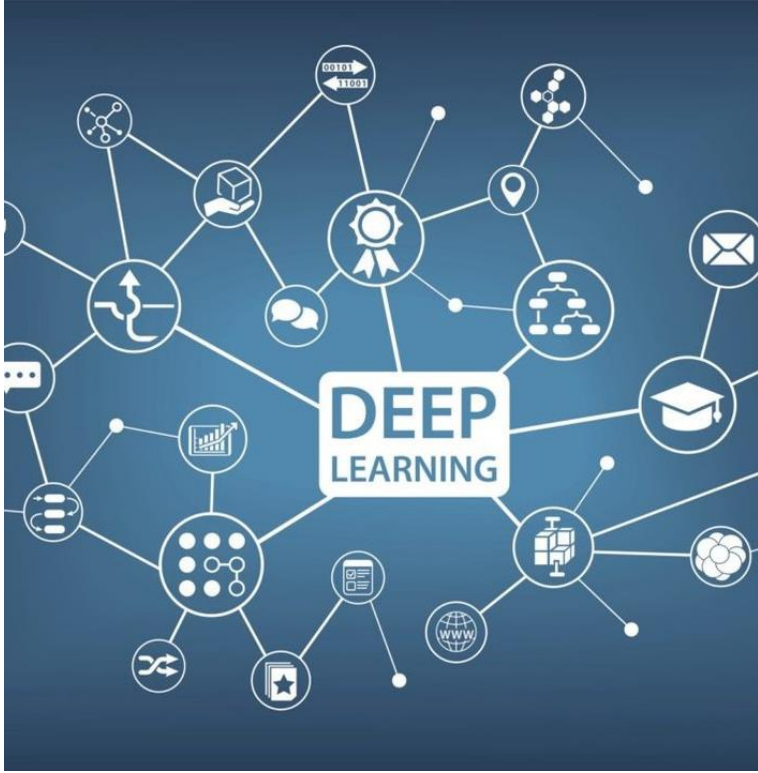
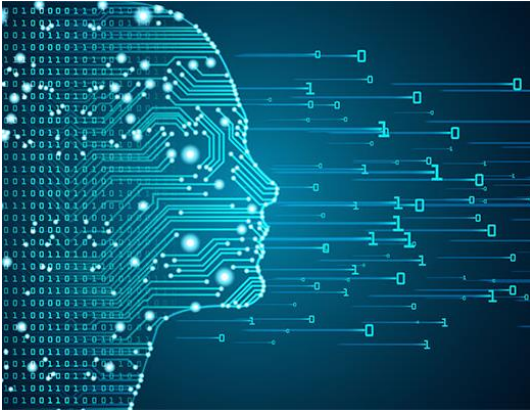
Finding the definition of  
these words using  
BabelNet in ES and EN



Gathering a list of 3000  
frequent English words

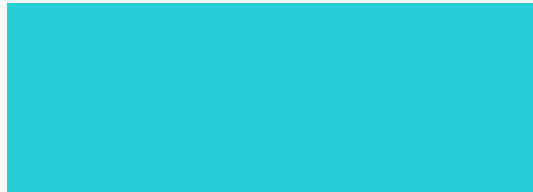
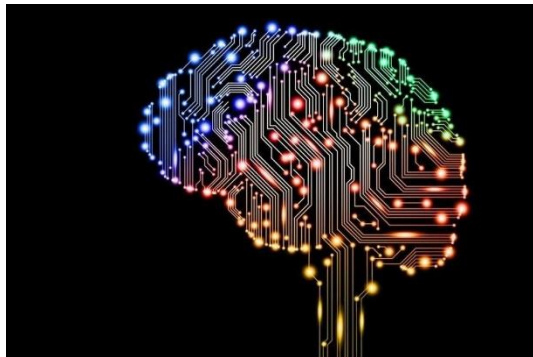






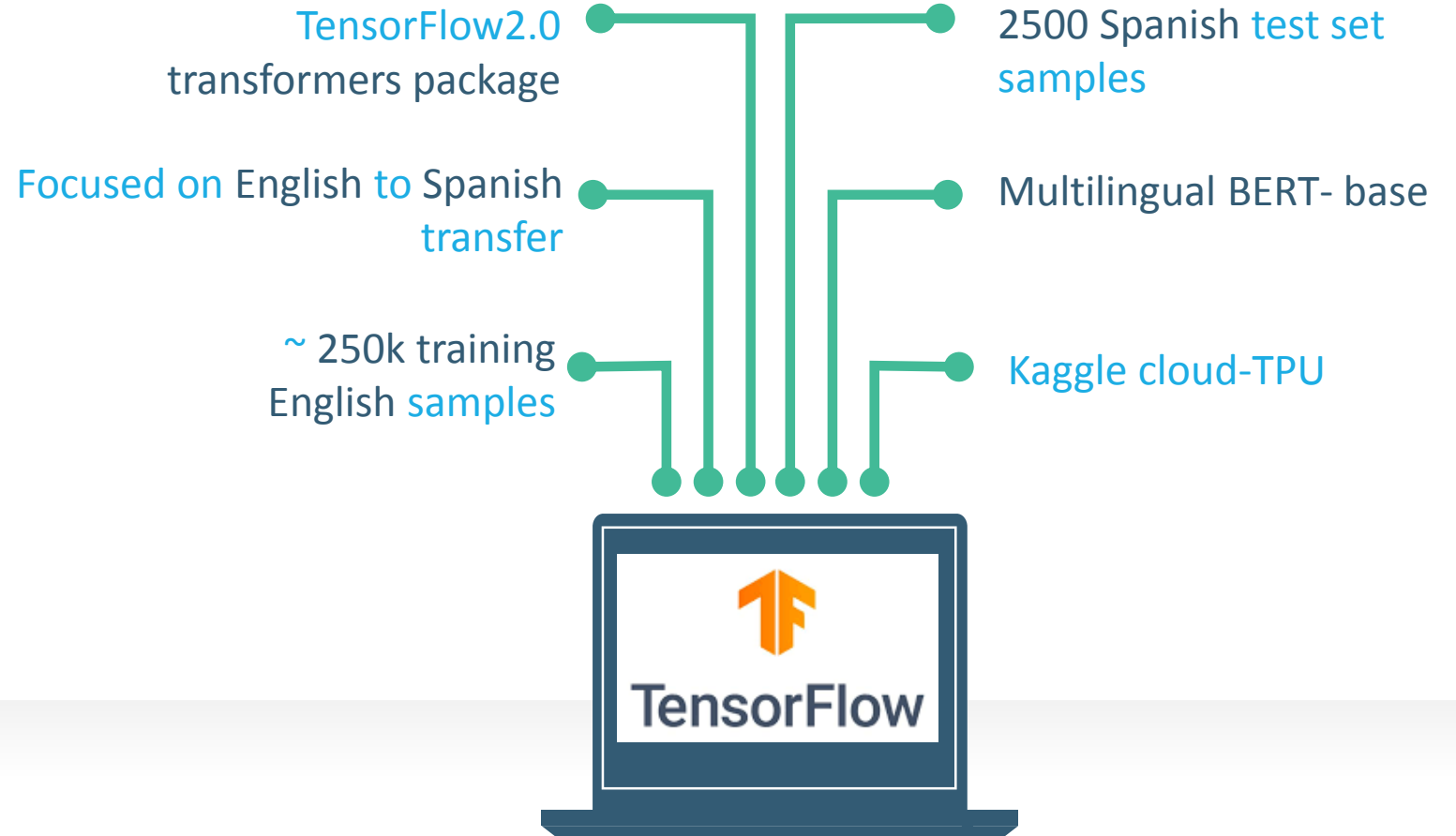
# Implementation

Implementation details



# Implementation

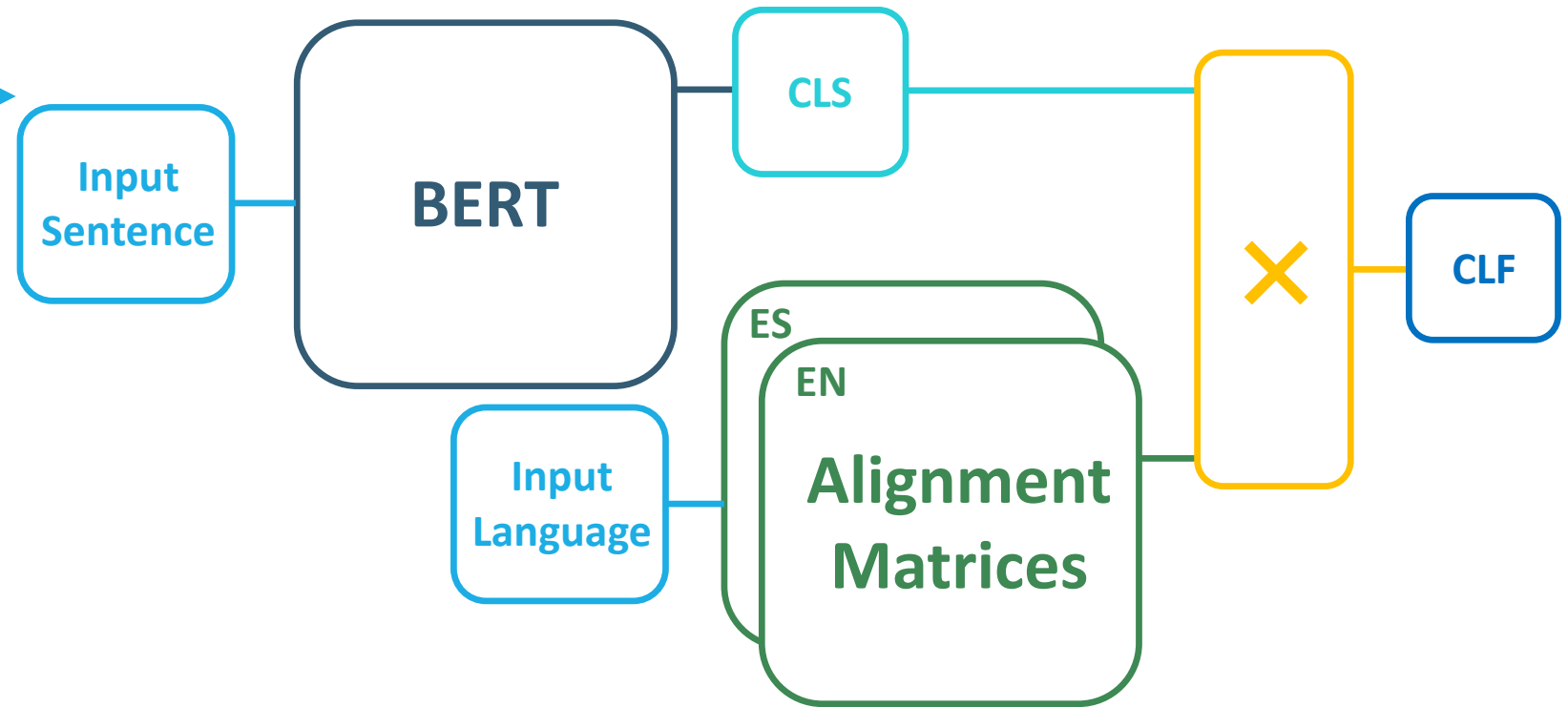
## Experimental Setting



# Implementation

Experimental Setting

Model Architecture

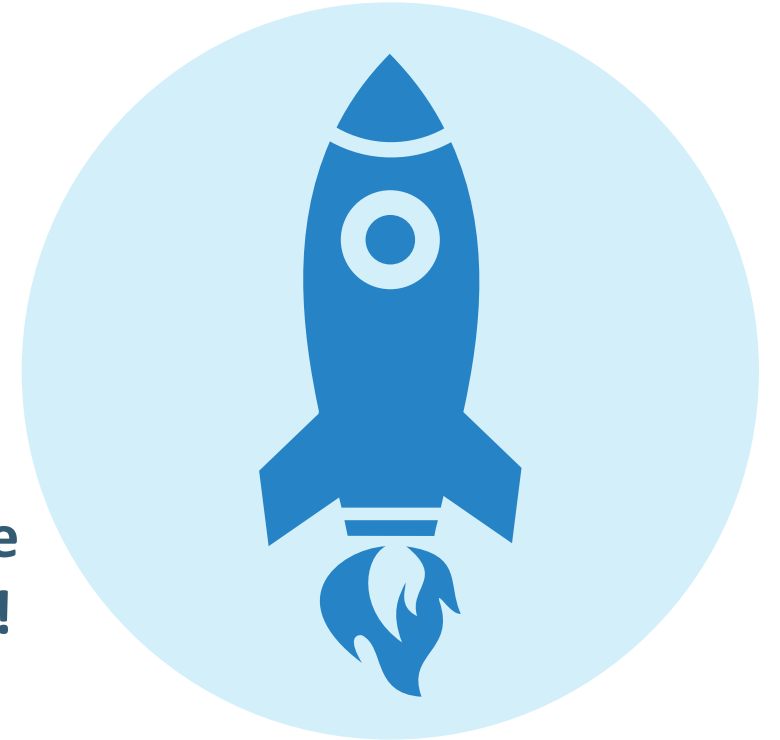
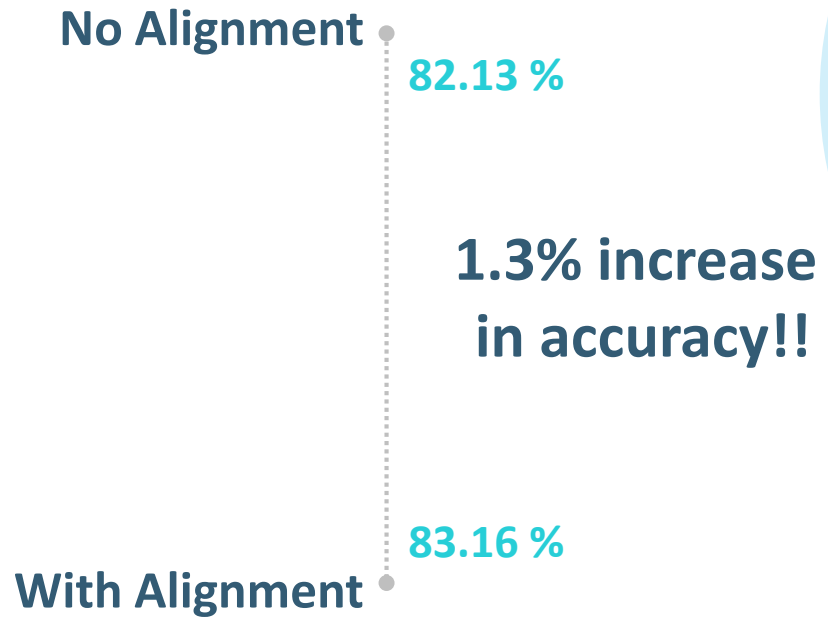


# Implementation

Experimental Setting

Model Architecture

Results





THANK YOU

For Your Time!