Smart Investigation

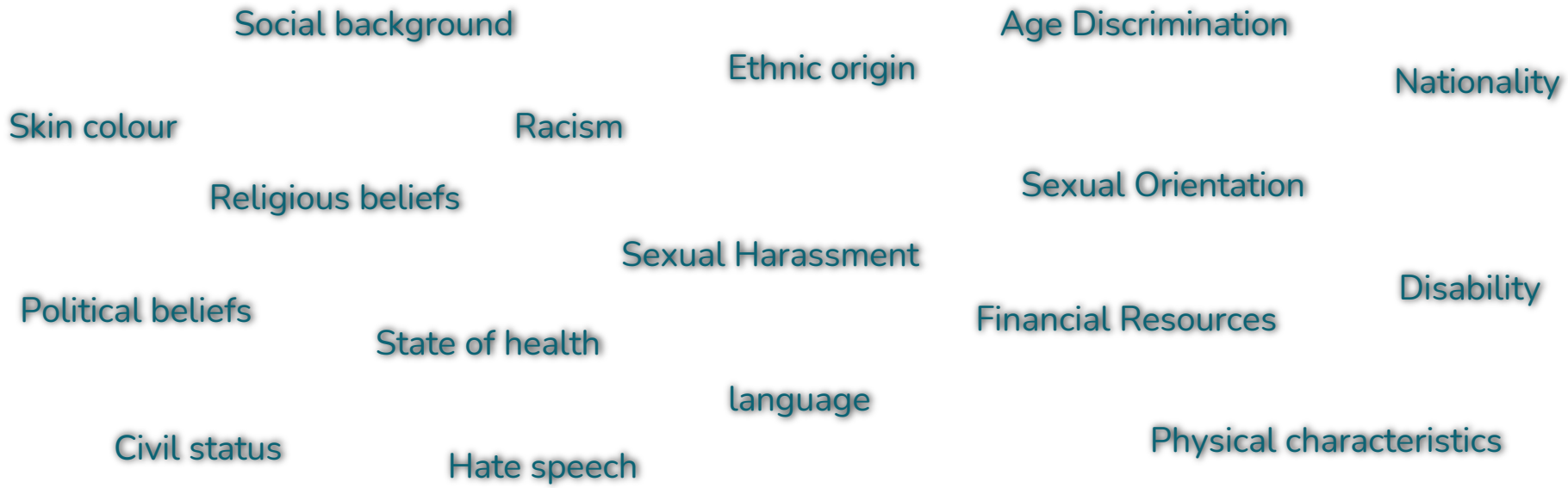
Social Media Scanner

By BeCode SMS

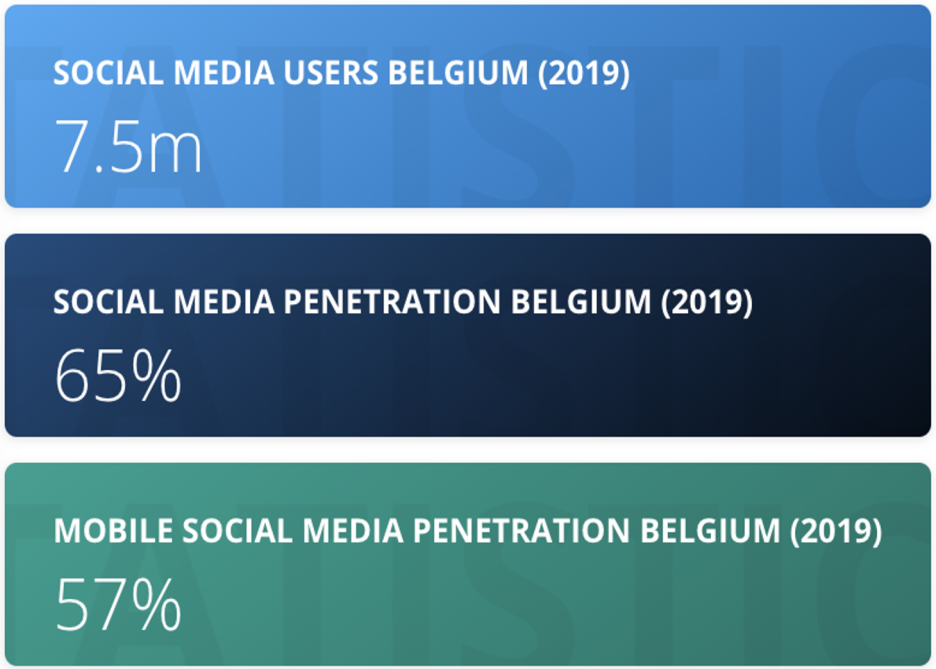
How to extract intelligence from social medias or other internet channels?

Social Media Scanner

The Spread of Discriminating Ideas



Overview of Social Media



1.7 million Facebook users in Belgium that were between 25 to 35 years old.

https://www.statista.com/topics/5525/social-media-in-belgium/#dossierSummary__chapter1

Social Media Intelligence

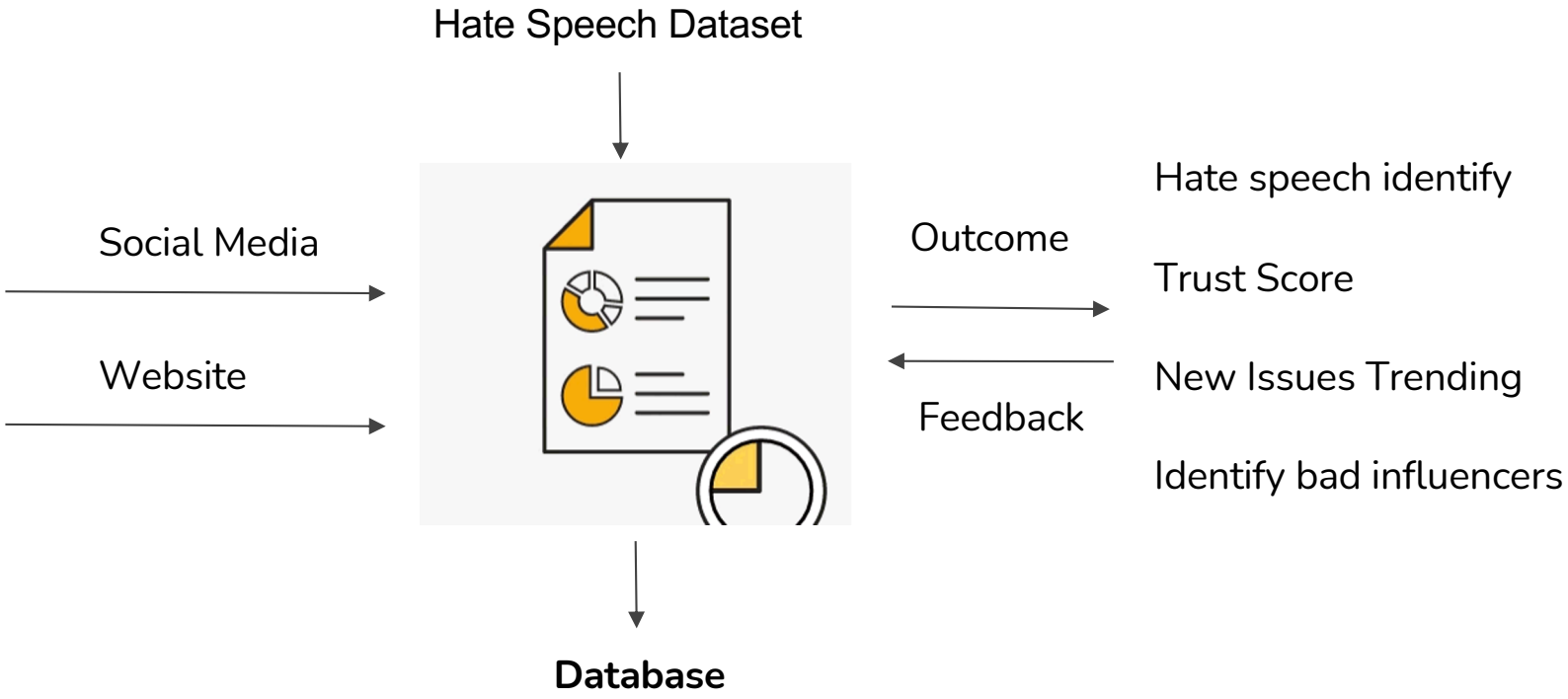
obtain text from social medias / website

Take action

Make report

Identify the probability of the text to be discriminative

Concept : Social Media Scanner



Prototype

Testing

In [33]:
```
def predict(text,model):

    text = pd.Series(text)
    tfidf = vectorizer.transform(text).toarray()
    tweet_tags = []
    for t in text:
        tokens = basic_tokenize(preprocess(t))
        tags = nltk.pos_tag(tokens)
        tag_list = [x[1] for x in tags]
        tag_str = " ".join(tag_list)
        tweet_tags.append(tag_str)
    pos = pos_vectorizer.transform(pd.Series(tweet_tags)).toarray()
    feats = get_feature_array(text)

    #Now join them all up
    X = np.concatenate([tfidf,pos,feats],axis=1)
    my_prediction = model.predict(X)

    if my_prediction[0] == 2:

        print('This sentence is neutral.')

    elif my_prediction[0] == 1:

        print('This sentence is offensive.')

    else:

        print('This is a hate speech.')

    return
```

In [34]:
```
text = 'we are going to party tonight'

predict(text,lg)
```

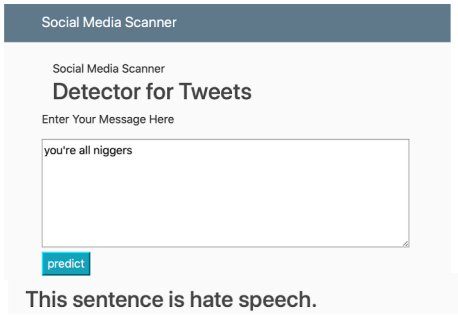
```
This sentence is neutral.
```

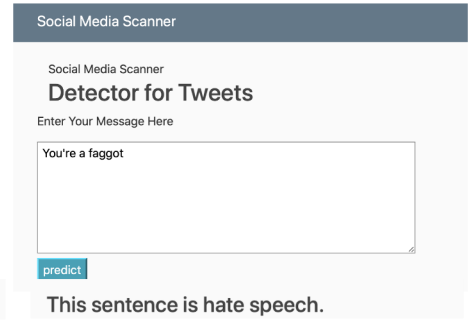
In [35]:
```
text = "you're all niggers"

predict(text,lg)
```

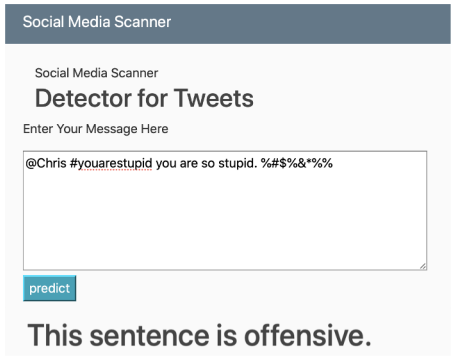
```
This is a hate speech.
```

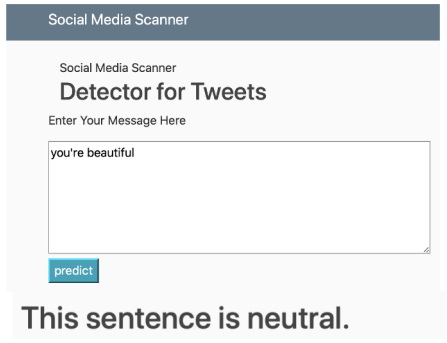
In [36]:
```
text = "You're going to die!"

predict(text,lg)
```

```
This sentence is offensive.
```

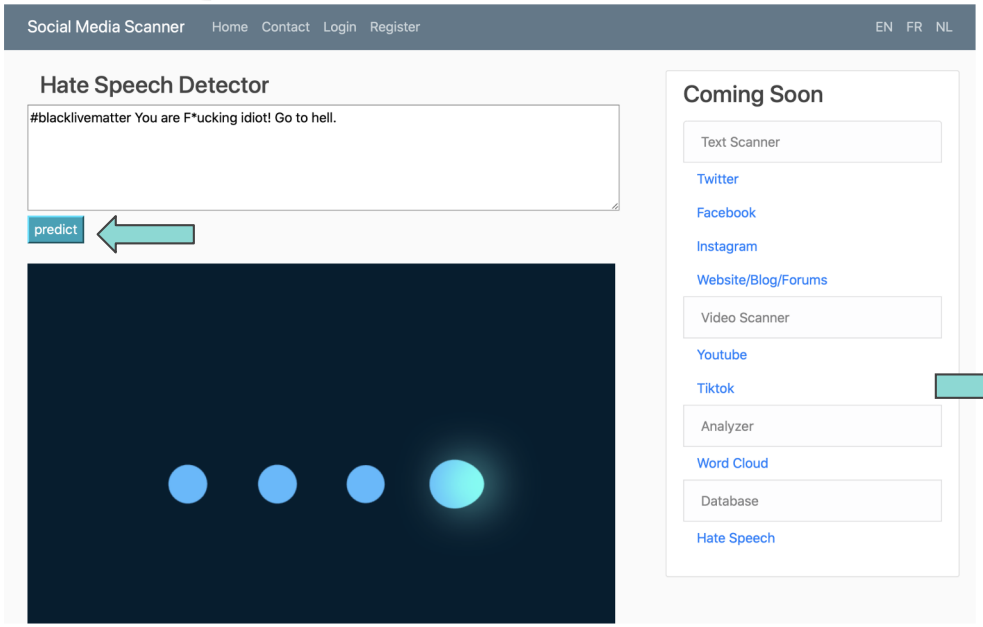




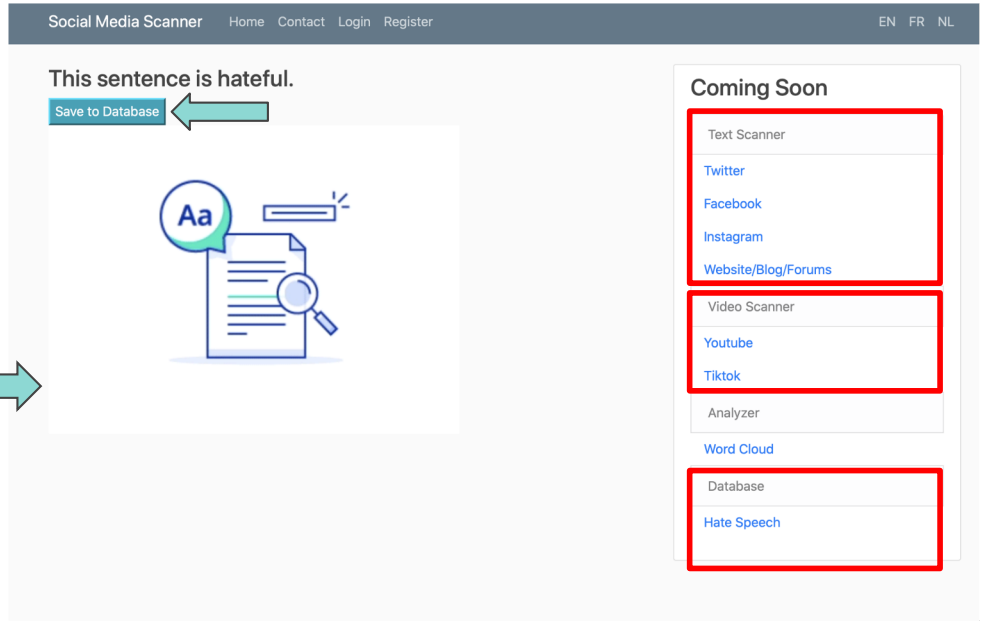




Showcase



Predict hate speech



Result and Save to Database

https://social-media-scanner.herokuapp.com

Conclusions

- To gather social media intelligence and detect hate speech content from social media or website.
- To provide valuable information for police for new issues trending.
- To create a database for hate speech model development.

Knowledge Leap

- Enormous case study done
- Concept formulation
- Model building for Hate Speech Detection
- Deployment on Hate Speech Detection online

LIVE DEMO

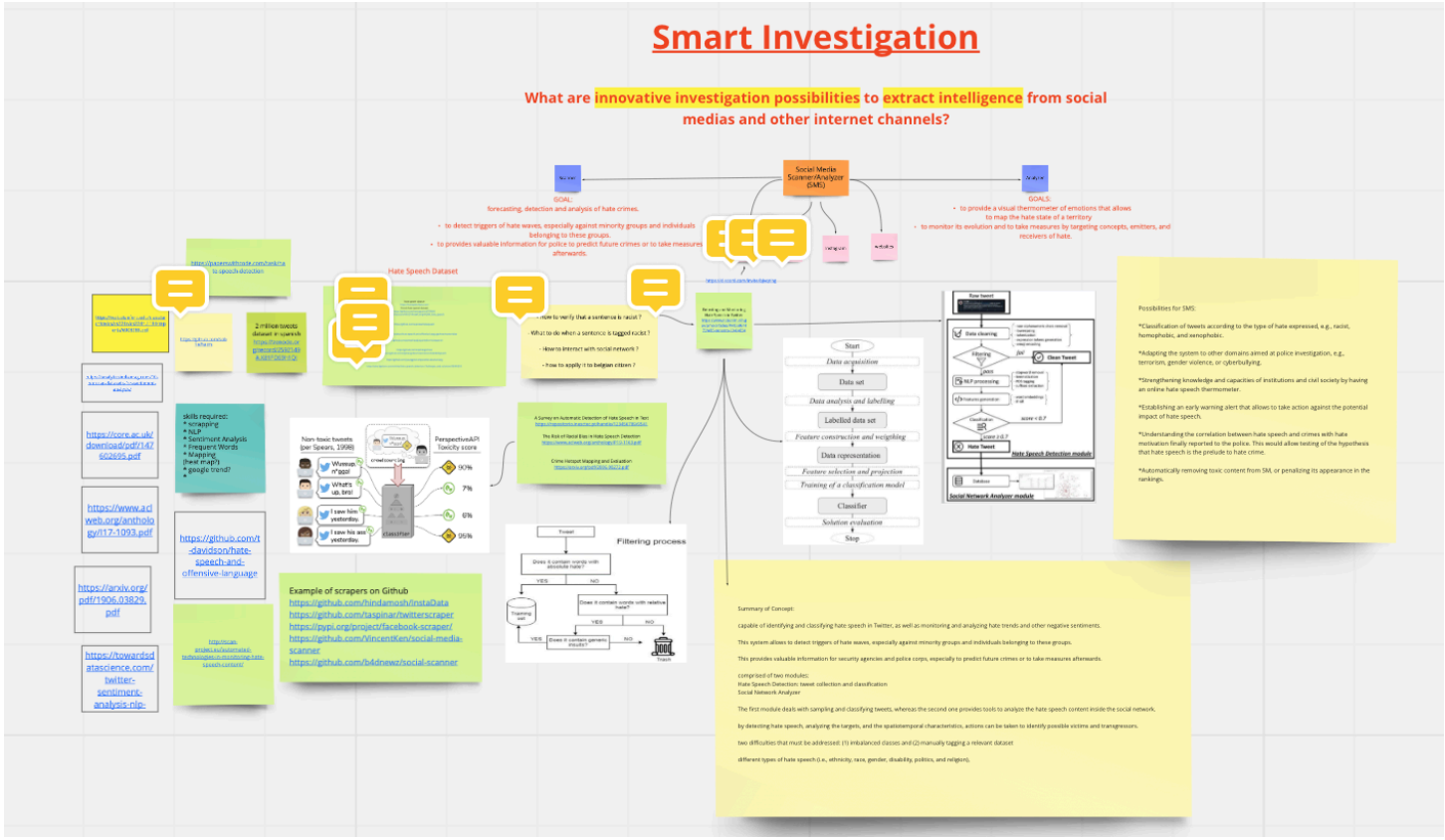
https://social-media-scanner.herokuapp.com

Github Repository

https://github.com/kaiyungtan/Social-Media-Scanner

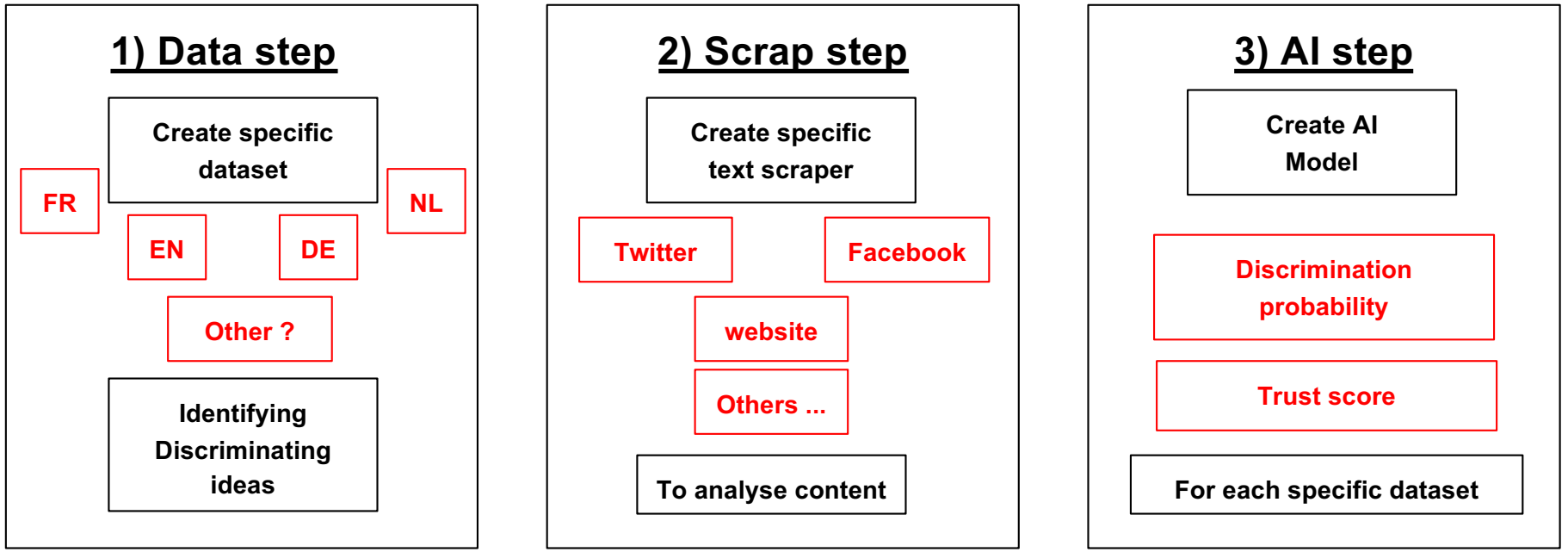
Thank You

Ideas Board



https://miro.com/app/board/o9J_lcCfeB8=/

The simplified big picture



The simplified big picture