



## ***New Emergency Management in a Resilience Era Facing Health, Climate and Energy Challenges***

***6<sup>th</sup> to 10<sup>th</sup> December 2021***

Date and slot of presentation to be filled in shortly

Authors names and affiliations

## IMPETUS Partner – INSIKT



### AI RESEARCHERS AND DEVELOPERS

**COLLABORATING WITH GOVERNMENTS,  
LAW ENFORCEMENT AGENCIES AND OTHER  
INSTITUTIONS **TO FIND ANSWERS AND HELP  
THEM SAVE LIVES****



## IMPETUS Partner – INSIKT – Who we are

2016

Insikt Intelligence was founded **5 years ago**



Based in **Barcelona**, working globally



**Research-performing Technological SME**



**3 Seals of Excellence by European Commission**



## IMPETUS Partner – INSIKT – Problems to solve



**RADICALIZATION  
AND TERROR**



**DISINFORMATION AND  
MISINFORMATION**



**HATE SPEECH  
AND THREATS**



**FINANCING OF  
TERRORISM**

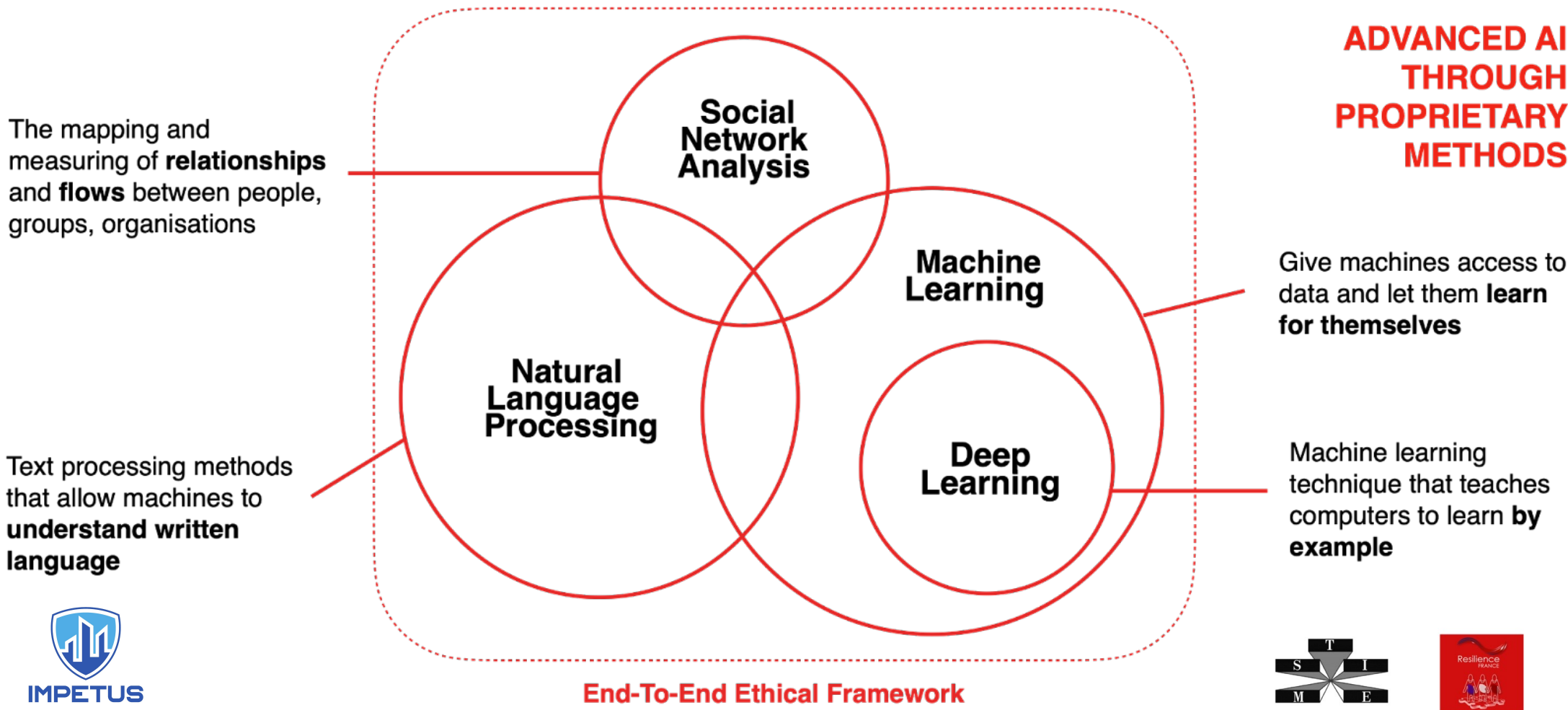


**DETECTION OF  
CRIMINAL ACTIVITY**

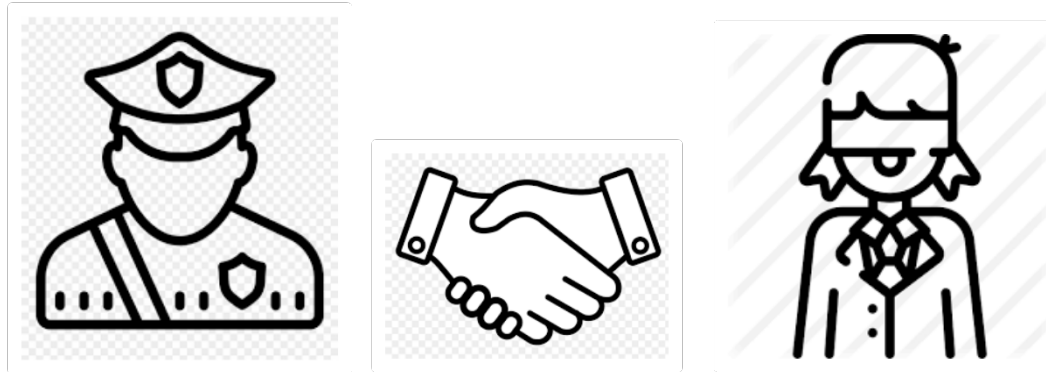


**PUBLIC OPINION AND  
BRAND REPUTATION**

## IMPETUS Partner – INSIKT – Our methodology

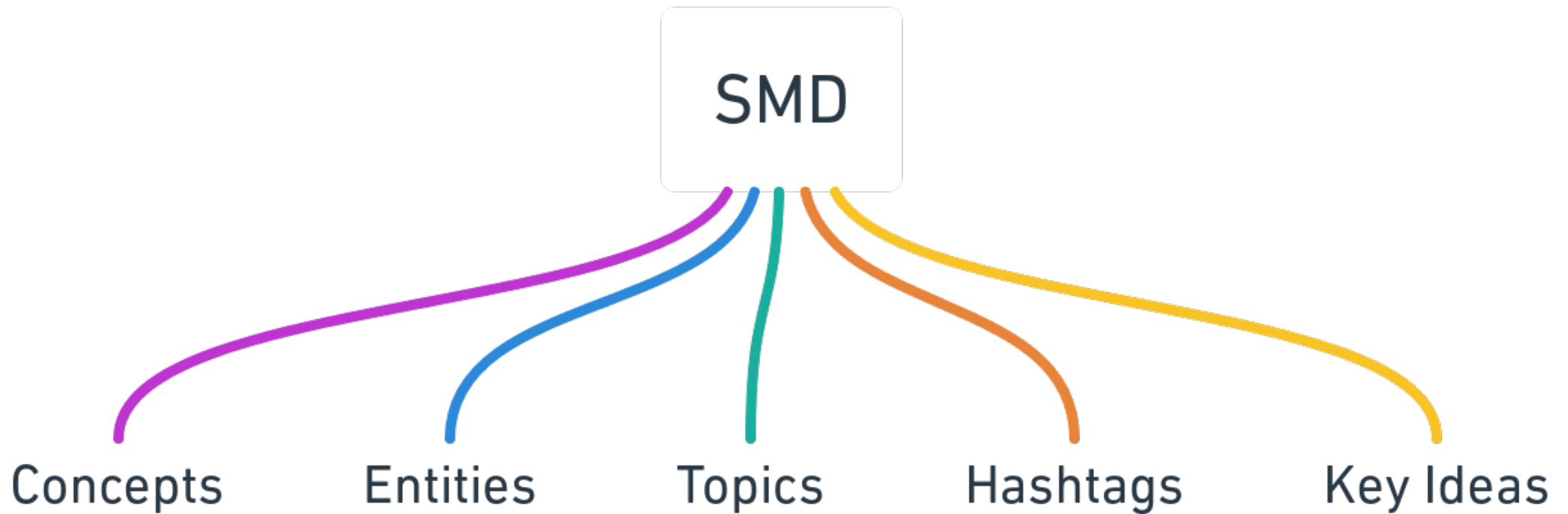


## Social Media Detection tool



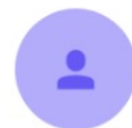
- Ethical-based monitoring of online text content for security in smart cities.
- Big challenge: develop an AI tool with machine learning models without knowing the domains within it will be applied.
- Product as flexible as possible.

## Social Media Detection tool



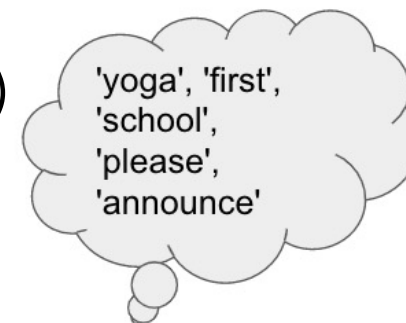
## Concepts

- Process text to be analysed (tokenization, remove stopwords, clean symbols, etc)
- Part-of-speech tagging: labelling all words with the function in sentence
- [NOUN, PROPN, VERB, ADJ, ADV, DET, etc]
- Select VERB, NOUN and ADJ
- Lemmatizer: extracts the lemma of the word (i.e. running->run)



ba56019f

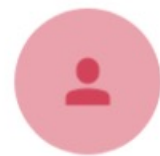
We are so pleased to announce that we are the first Mini Me Yoga school in Wales! 🚩





## Entities

- Process text to be analysed
- It recognizes three categories of entities: LOC-PER-ORG



QAZWS0192

We went to Barcelona and had a very good time! ❤️❤️

{'type': 'location',  
'entity': 'barcelona'}

## Topics

- Process text to be analysed (tokenization, remove stopwords, clean symbols, etc)
- It vectorizes the text in a 300 dimensions embedding
- It finds the closest topics given a predefined list of topics by calculating semantic similarity
- It classifies the sentences in these topics
- List of topics: ['politics', 'family', 'fashion', 'education', 'accident', 'construction', 'financial and business', 'childhood', 'technology', 'civil unrest', 'language', 'racism', 'mass media', 'police', 'system of justice', 'jihadism', 'energy and resource', 'natural hazard', 'environment', 'defence', 'woman', 'weapon', 'health', 'immigration', 'employment', 'tourism and leisure', 'transport', 'fascism', 'government', 'war', 'science', 'music', 'crime', 'property', 'literature', 'entertainment', 'award and prize', 'religion', 'art', 'drogues']



CNYH8888

MUSIC

Talented artists who are at the pinnacle of their music generation are usually innovators, not duplicates.

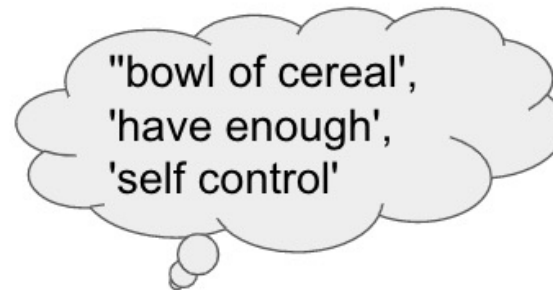


## Topics

- Process text to be analysed (tokenization, remove stopwords, clean symbols, etc)
- Part-of-speech tagging: labelling all words with the function in sentence [NOUN, PROPN, VERB, ADJ, ADV, etc]
- Select VERB, NOUN, ADJ, PREP and DET
- Looks for the defined patterns (i.e. NOUN+VERB, DET+NOUN+VERB, etc)



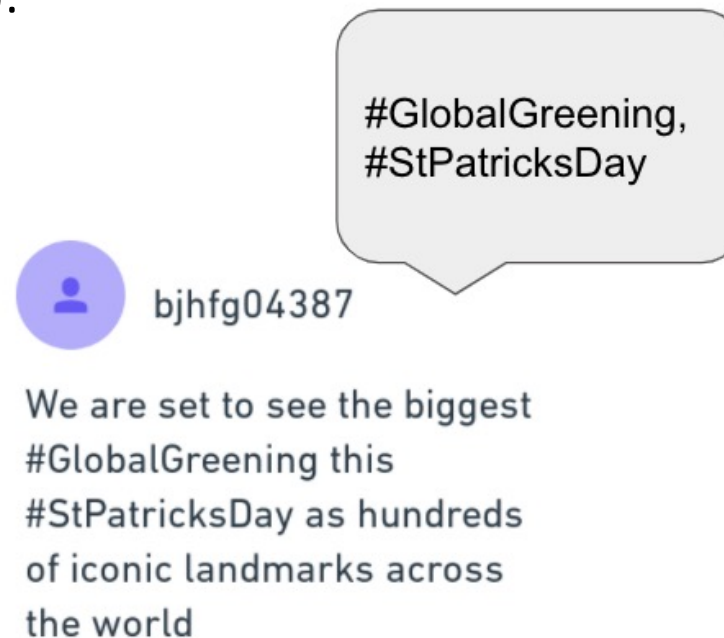
ba56019f



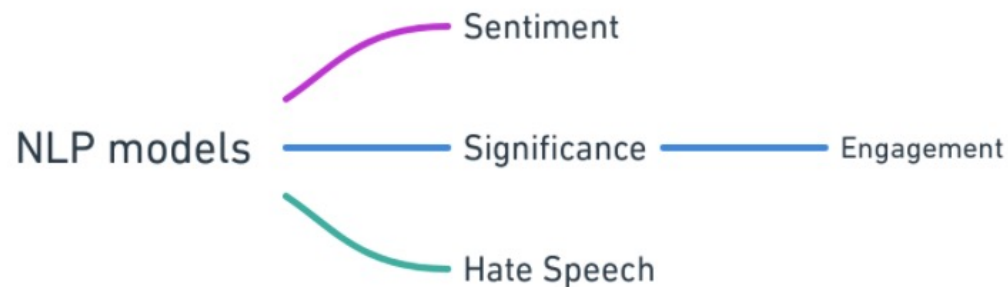
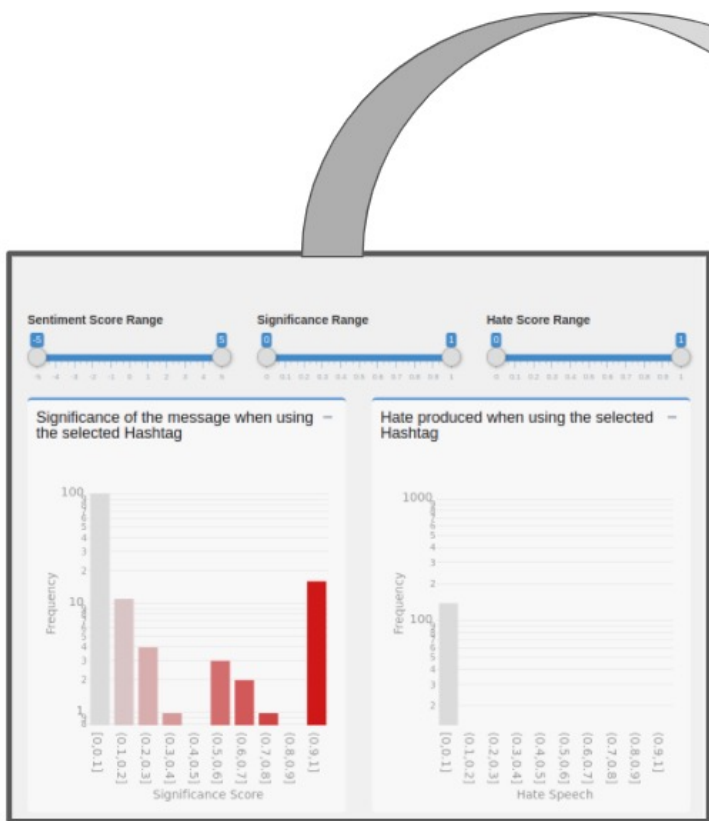
People who say they have enough self control to just eat one bowl of cereal are wrong

## Hashtags

- Process text to be analysed (tokenization, remove stopwords, clean symbols, etc)
- `def is_hashtag(self,text):`  
    `return text[0]=='#'`



# Big Data visualization



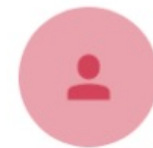
## Sentiment

- This analysis is lexicon based, which means it is based on the quantification of the sentiment for each word, the result is the average of the sentiment values of each word of the sentence.
- Results are from -5 (most negative sentiment) to 5 (most positive sentiment).

 CNYH8888

-2.6

These pictures aren't the best quality

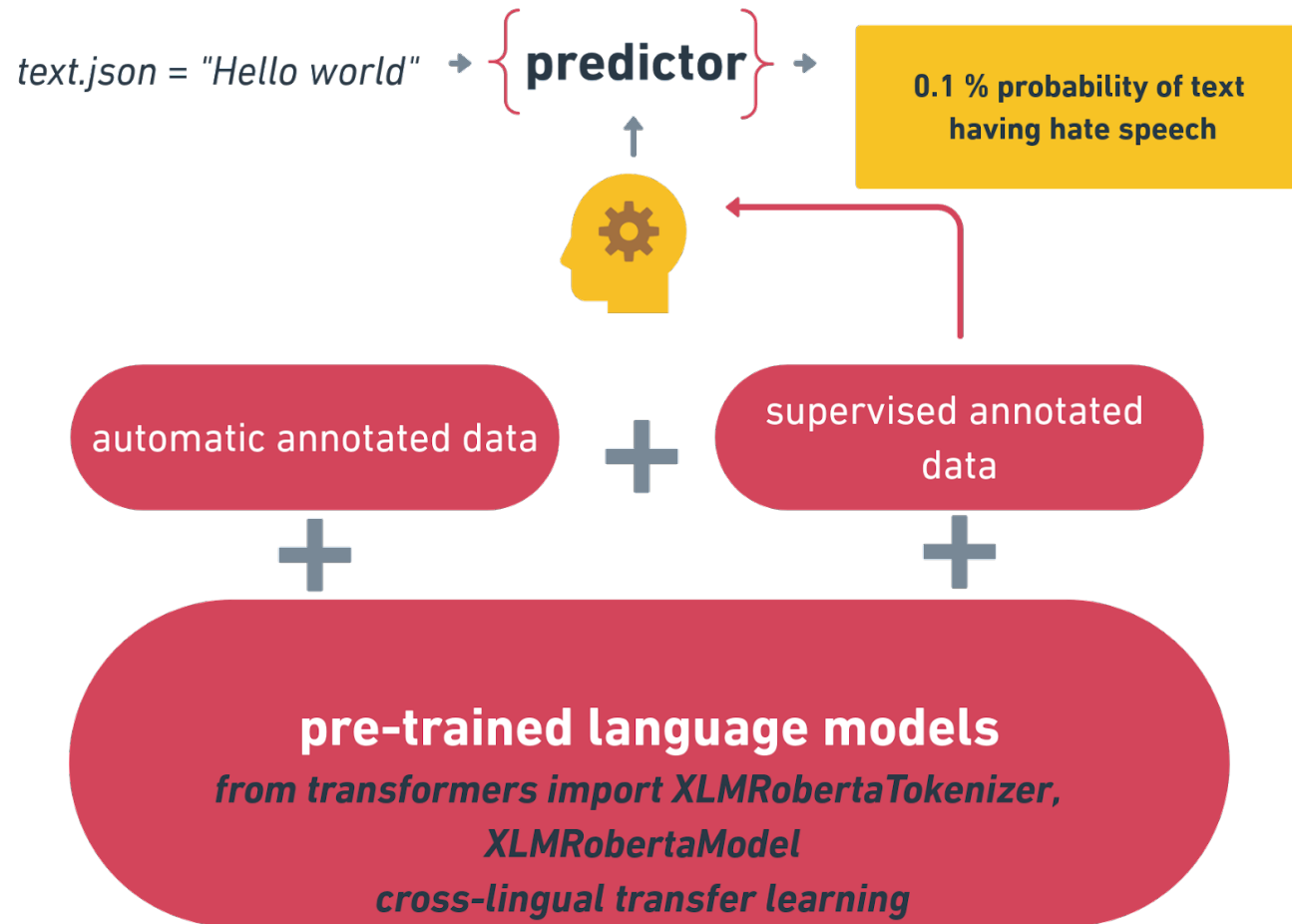


QAZWS0192

2.5

We went to Barcelona and had a very good time! ❤️❤️


# Hate Speech



## Hate Speech - Examples


**Hate speech** covers many forms of expressions which advocate, incite, promote or justify hatred, violence and discrimination against a person or group of persons for a variety of reasons.

0.7

 QAZWS0192

Black children are getting killed by domestic terrorism and the news is barely covering it.

0.9

 ba56019f

Assholes believe that fucking bank of england rules the world 😏..its fucking ruling boundaries are till the fucking UK



## Significance score = engagement



nshares  
ncomments  
nlikes  
nfriends  
nfollowers



$$\text{engagement} = 0.2 * \text{scoreShares} + 0.8 * \max(\text{scoreComments}, \text{scoreLikes})$$

$$\text{impact} = \max(\text{scoreFollowers}, \text{scoreFriends})$$

## Social Network Analysis

Several metrics are computed:

- **Activity:** Total number of times the user has been **mentioned** or has **mentioned others**.
- **Influencer:** Quantifies the influence of a person. An influencer is someone whose opinion is deemed as important by their peers.
- **Spreader:** Determines how much the user acts as a spreader of information. That is, **the ability of the user to spread information quickly over the network.**  
**Role:** Determines the role of the user.
  - An opinionated, active user that mentions other people a lot, but is not mentioned as much.
  - A conversationalist. The user interacts bidirectionally with other people (replies and is being replied to).
  - An external influencer. Is mentioned a lot but does not contribute to the conversations.

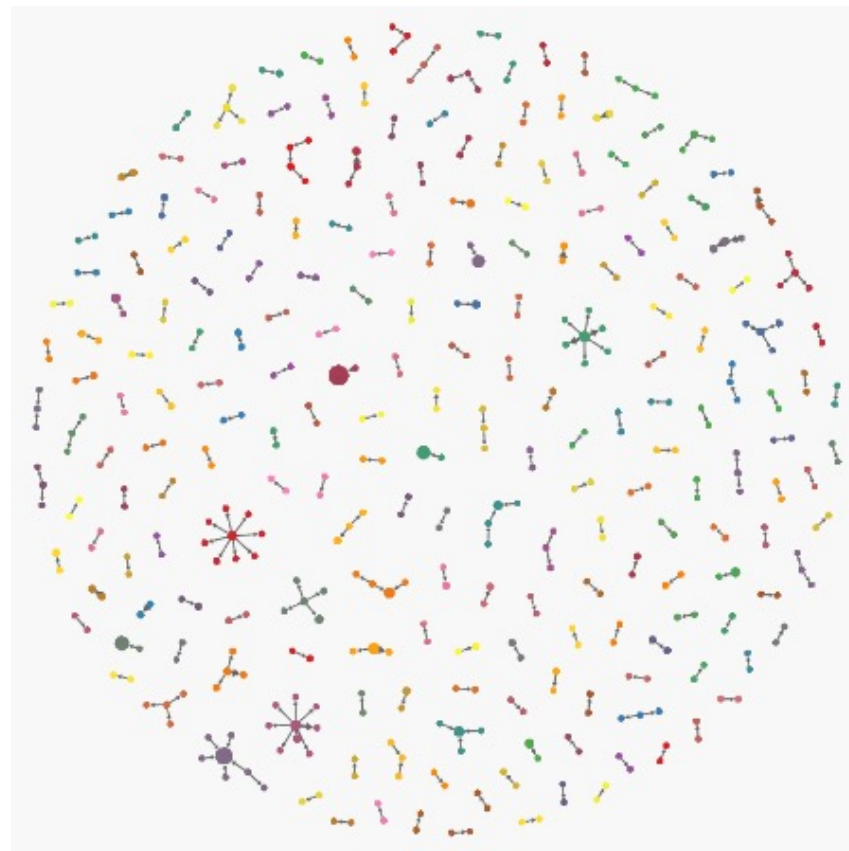


## Social Network Analysis

*Process of investigating social structures through the use of networks and graph theory.* (Wikipedia, 2020, 5th April)

Network is a Graph. This is composed by:

- **Nodes/Vertices.** These represent the Users.
- **Edges.** These represent the relationships between the users (mentioning and being mentioned).



## Social Network Analysis

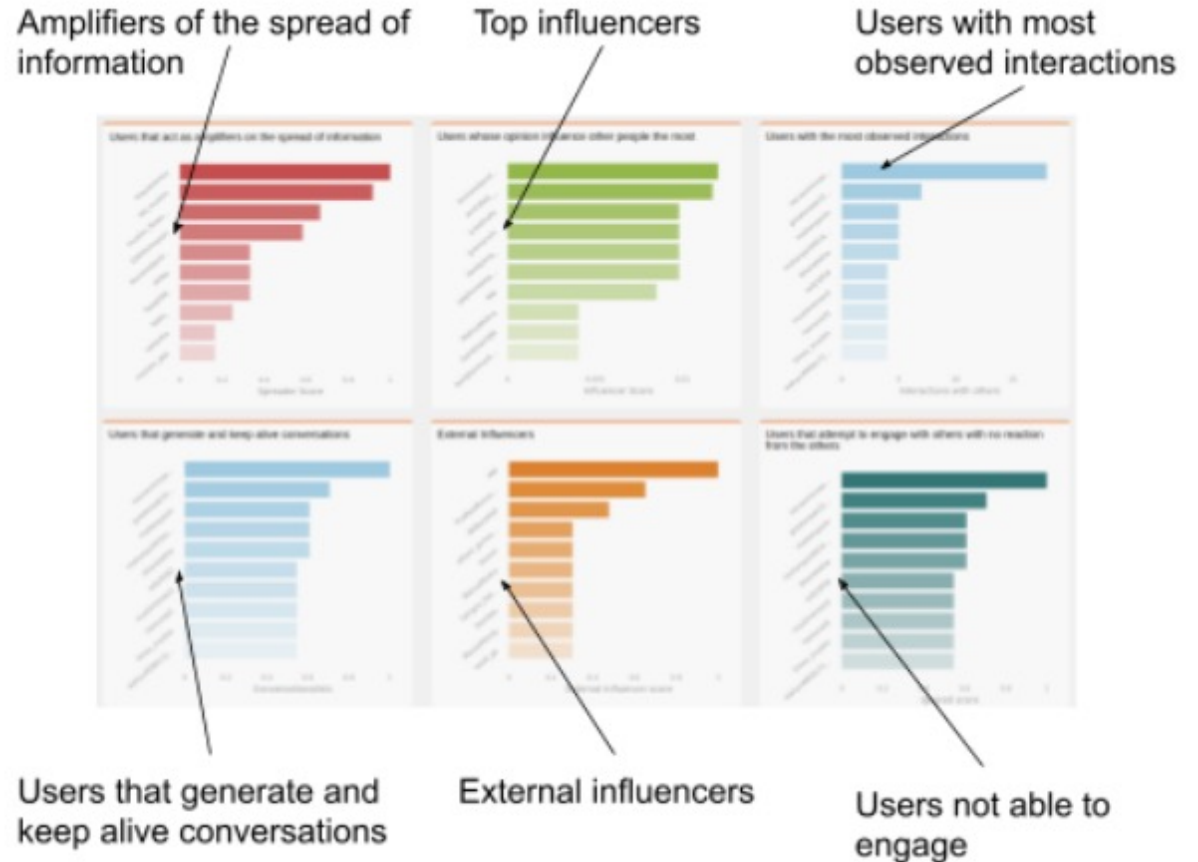
These metrics can be used as visualization parameters to easily see how each user plays its role in the networks.



## Social Network Analysis

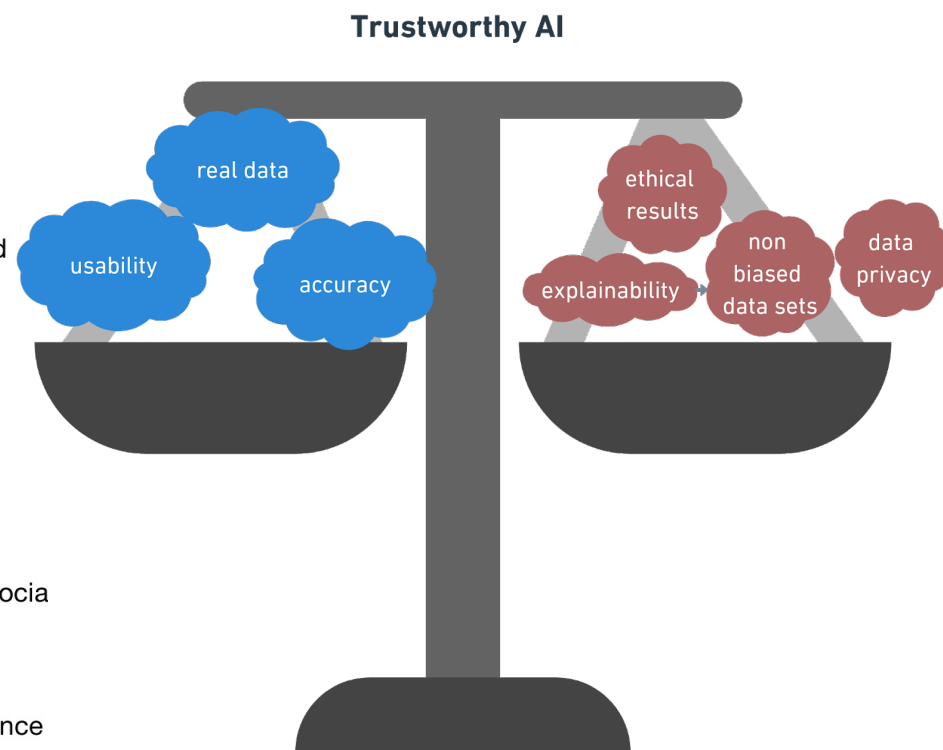
**Disruption.** Refers to the act of breaking ties in a network via multiple approaches so as to leave it disconnected. One can disrupt the network effectively by focusing on the most important users, according to different criteria and the goal to be achieved.

Using the metrics before, the disruption of the network can be understood from different perspectives.



# Social Network Analysis

- 1 Human agency, liberty and dignity**  
Positive liberty, negative liberty and human dignity
- 2 Technical robustness and safety**  
Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility
- 3 Privacy and data governance**  
Including respect for privacy, quality and integrity of data, access to data, data rights and ownership
- 4 Transparency**  
Including traceability, explainability and communication
- 5 Diversity, non-discrimination and fairness**  
Avoidance and reduction of bias, ensuring fairness and avoidance of discrimination, and inclusive stakeholder engagement
- 6 Individual, societal and environmental wellbeing**  
Sustainable and environmentally friendly AI and big data systems, individual wellbeing, social relationships and social cohesion, and democracy and strong institutions
- 7 Accountability**  
Auditability, minimisation and reporting of negative impact, internal and external governance frameworks, redress and human oversight



## AI Ethics

### Human Agency, Liberty and Dignity

**Challenge:** Enable the ability for humans to be autonomous and self-governing

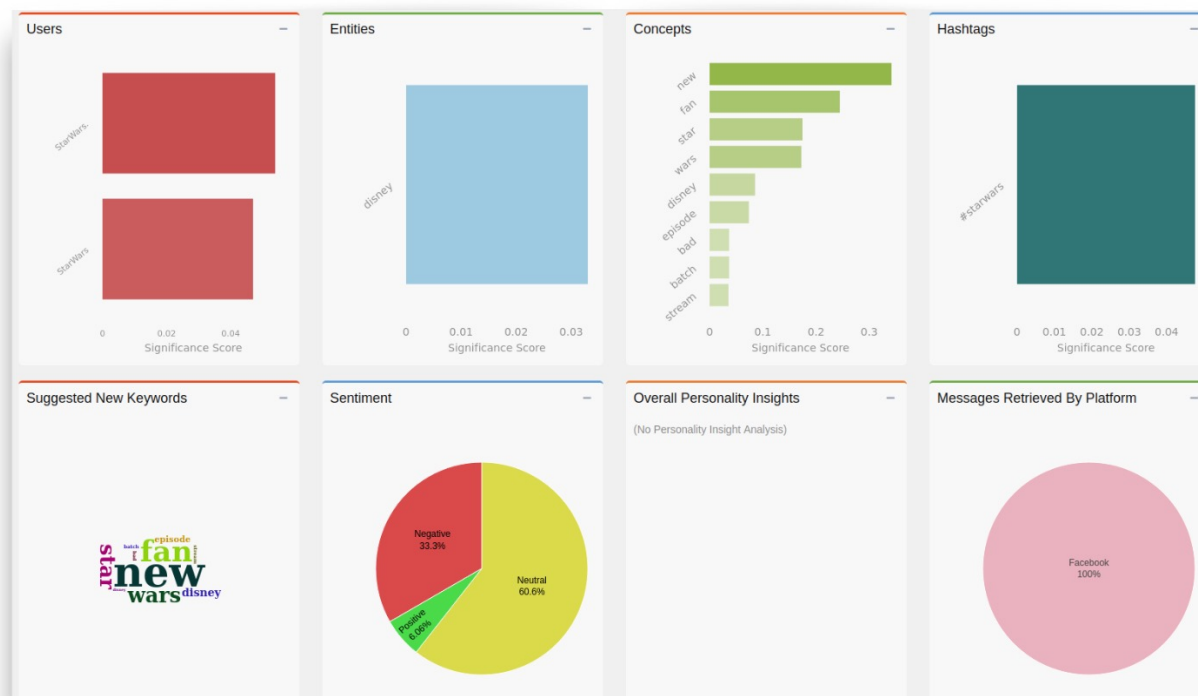


## AI Ethics

### Human Agency, Liberty and Dignity

We work with **human-in-the-loop** approach. Our ML models don't take autonomous decisions, and neither automatic actions, on the contrary, they provide information to the human, who decide the actions to be taken.

We must ensure that the **user is contextualised and trained** in the use of the tool. The users must be aware that they own decisions may end up including bias in the human-in-the-loop analysis.





## AI Ethics

### Technical Robustness and Safety

**Challenge:** Ensure that the tool that we are developing is safe and secure.



## AI Ethics

### Technical Robustness and Safety

Security and technical robustness helps **avoiding** the use of the tools by **malicious users**.

- Use of cloud services that are ISO certified
- User profile management as well as user profile policies
- Connections to Servers through HTTPS TLS and SSH
- Use of a VPN
- Use of a Firewall
- Continuous scanning for vulnerabilities



## AI Ethics

### Privacy and Data Governance

**Challenge:** Make sure that the system does not violate or infringe upon the right to privacy, and that private sensitive and/or personal data is well-protected.

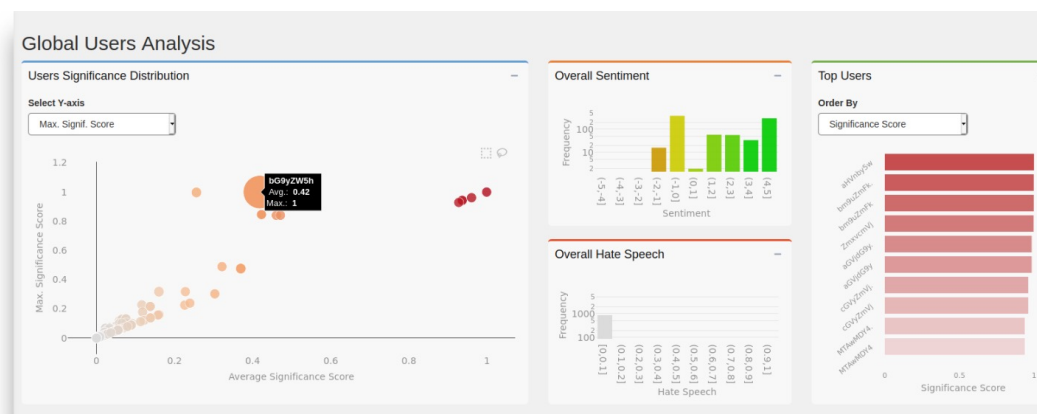


# AI Ethics

## Privacy and Data Governance

### Data management

- The **use of the data** is strictly limited to the corresponding project.
- The **access to the data** is strictly limited to a limited number of personnel participating in the corresponding project.
- **All data is deleted** after the end of each project or after a specific period.
- Full data **anonymization** [process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party]
- **Pseudonymization** - restricted access for de-anonymized to authorized authorities/LEAs.



## AI Ethics

### Privacy and Data Governance

#### De-anonymization process

Data will only be revealed after an authorized user asks for specific users information and introduces the authorization password.

Data de-anonymization will be done offering a “Show Information” button where individual user information is presented:

- Users --> General --> Individual User Analysis --> Profile Information, Messages with Filters Applied (username)
- Content --> Messages (Selected Message's User Info)
- Networks --> Interactions --> Observed Interactions Table → show de-anonymized info for certain users
- Location --> User Location --> Selected User Info

When clicking the button, the user will be asked to introduce a password for the decryption process. Only authorized user will be allowed to perform that process.



## AI Ethics

**Diversity, Non-discrimination and Fairness**

**Challenge:** Identify bias in AI models and remove it.



## AI Ethics

### Diversity, Non-discrimination and Fairness

Test the bias in ML models for protected groups.

- We input models with what should be considered neutral input and check the results



- Work towards the fairness of the models by:
  - Debiasing the training datasets
  - Optimising the algorithms with the right metrics

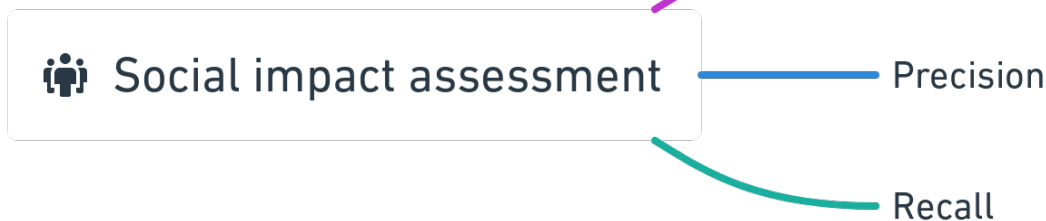
## AI Ethics

### Diversity, Non-discrimination and Fairness

**Debiasing the datasets** by replacing the sensitive terms for protected groups by generic ones in all our training data sets:

```
bias_topics = {'location':location, 'sexual orientation':sex_orient,  
              'religion':religion,'race':race,'politics':politics,  
              'gender':gender,'colour':colour, 'Name':name}
```

- Annotators should be gender, nationality and religious **diverse**.
- **Fairness on the models: the metrics criteria based on social impacts.**
- Hate speech detection --> **Precision** is optimized instead of accuracy or recall



“ All *lesbians* and *jews* have to go to church ”

Aligned 300-D word embeddings

“ All *sexual orientation* and *religion* have to go to church ”





## AI Ethics

### Transparency

**Challenge:** Make understandable how our tool achieves its decisions.



## AI Ethics

### Transparency

- The **original text** is always shown to the user to facilitate the understanding of the analysis
- Detailed **training** and user-manual is offered to the **end-users**, including the explanation of all the methods used in the analysis, to enhance transparency in the use of AI and Big Data.

## AI Ethics

**Individual, societal and environmental wellbeing**

**Challenge:** Reduce the harm possibly caused to individual, societal and environmental wellbeing.



## AI Ethics

### Individual, societal and environmental wellbeing

**Quantization of the models** → smaller models consume less energy in the production environments. Using quantization we have reduced 70% of the size of the models, from 1G to 300Mb.



# References

- ▶ Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach: Philip Brey, Björn Lundgren, Kevin Macnish, and Mark Ryan

