



**GOVTECH CHALLENGE SERIES 2.0**

# HOW TO DETECT AND MONITOR HATE SPEECH ON THE INTERNET?



**The Office of Inspector of Journalist Ethics**



**Ministry of Social Security and Labour**

**GOVTECH** LAB





# CONTEXT

## CURRENT SITUATION

- Currently, hate speech and hate crime in the public internet space is monitored through a keyword-based monitoring system.

## SCOPE

- This challenge is important, because the current keyword-based monitoring system does not match organisation's needs;
- A high number of hate speech cases go unnoticed, offenders are undetected, and a sense of impunity prevails, which in turn leads to an increase in such speech in public space.

## PROBLEM

- The number of cases provided by the keyword-based monitoring system is unreasonably high;
- The system is inefficient as a lot of the time is spent on reviewing the cases, even though most of them are “empty”, i.e. no actual hate speech is detected;
- In other words, this requires an unsustainable amount of human resources that the organisation cannot allocate.

The Office of Inspector of  
Journalist Ethics



Ministry of Social Security  
and Labour





# CHALLENGE

We are looking for an innovative tech solution that can identify and monitor hate speech on the internet as accurately as possible.

The Office of Inspector of  
Journalist Ethics



Ministry of Social Security  
and Labour





# SOLUTION

## KEY FUNCTIONS

- The solution would be able to semantically identify hate speech based on given keywords;
- Simple and user-friendly user environment;
- Each comment must be related to the primary source of information;
- Possibility to process and manage data provided by the system, e.g. statistics;
- Possibility for continuous system improvements (developing system versions).

## METRICS FOR SUCCESS

- The number of detected real hate speech instances;
- An accurate representation of semantic information.

## IMPORTANT ASPECTS

- Often, additional characters (dots, spaces) are inserted into the written expressions of hate speech. This makes it difficult for the monitoring system to detect it;
- Hate speech can be presented in the form of pictures or icons (emojis);
- Certain platforms have protection against these kind of solutions, which poses problems for detecting hate speech.

The Office of Inspector of  
Journalist Ethics



Ministry of Social Security  
and Labour





# OPPORTUNITIES

## INSTITUTION

- We are ready to provide all available information related to the development of the solution - keywords, examples of hate speech;
- In our monitoring process, we have an opportunity to propose the introduction of such monitoring system to the media;
- The organization would purchase the services of a properly functioning monitoring tool.

## MARKET

- According to our research, a similar tool of the public space monitoring operates in France;
- The developed tool has the potential to be used for monitoring activities in other areas.

**The Office of Inspector of  
Journalist Ethics**



**Ministry of Social Security  
and Labour**

