JRC141468

European Commission

SCIENCE FOR POLICY BRIEF

Data and Tools to Counter Terrorism

SecData Platform – Transforming media information overload into data and insight

HIGHLIGHTS

- → Collection and analysis of open-source media information on security-related events and incidents to generate data and insights, facilitating and supporting threat assessment processes.
- → Automated process using Large Language Model (LLM) filtering, translation, duplication detection, clustering, categorisation, summarisation and assessment (enhanced by the use of Retrieval-Augmented Generation (RAG)).
- → Data visualisation supporting analysis through several functions: mapping of geolocated data, filtering (focus on user-selected aspects), creation of personalised charts to explore further and gain insights into trends.
- → Newsletters inform users about events at chosen frequencies (daily, weekly, monthly) providing a general overview or focusing on a particular subject, category or country.

'What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. ' – Herbert A. Simon

Analysis of security-related event data

In an interconnected world, ideas and experience of emerging threats are quickly shared through propaganda, publications and social media. Event-based assessment of security incidents can help decision-making in quantifying threat levels. The objective is the collection of open-source media information to generate data and insights in order to facilitate and support threat assessment processes.

Joint Research Centre

Europe Media Monitor

The European Commission's '*Europe Media Monitor (EMM)*', developed by the Joint Research Centre's Text and data mining unit, allows to explore and understand current news reported by the world's online media. Monitoring thousands of news sources, the system uses advanced information classification and extraction techniques.

EMM monitors the live web through curated pages from around 10000 key news portals world-wide, plus 20 commercial news feeds. It retrieves over 450.000 links to articles per day in 80 languages around the clock.

JRC research of security-related event data focuses on the application of secured Large Language Models to media news metadata gathered on a daily basis through EMM. The results of the analysis are made accessible in the *JRC Security Data (SecData) Platform* in the form of topic-specific 'data universes'.

The universes cover areas of terrorism- and extremism. Furthermore, in line with the *Counter-terrorism agenda for the EU (2020)* and the *EU's policy on countering the potential threats from non-cooperative unmanned aircraft systems (UAS)*, the SecData Platform supports the Commission's counter-drone expert group with a specific UAS Incident Database.

The automated universe development environment allows for a quick extension to other topics in line with EU policy.

A multi-stage process to transform news into data and insights

A retrieved news article goes through a multi-stage assessment process before integrating a database as an event or incident. Initially, articles are retrieved from EMM based on universe-specific keywords, elaborated in collaboration with subject matter experts. The articles then go through an automated process using Large Language Model (LLM) filtering, translation (into English, if the original article in another language), duplication detection, clustering, categorisation, summarisation and assessment. This uses Retrieval-Augmented Generation (RAG), optimizing LLM output by referencing authoritative knowledge bases external to the model's training data sources, before generating a response. Going through this process, the media information is transformed into data and pre-analysed, allowing analysts and interested stakeholders to gain insights and reduce information overload.

Data visualisation, analysis and notification

Data visualisation supports analysis through several functions: mapping of geolocated data, filtering (focus on user-selected aspects), creation of personalised charts to explore further and gain insights into trends.

Fast-paced evolution of underlying LLM's

Artificial Intelligence and Large Language Models are fast evolving fields. The JRC uses an internal platform to securely explore the use of pre-trained Large Language Models and offers access to a wide variety of Generative AI models. The underlying LLM models in the SecData Platform are constantly adapted as new, more comprehensive models become available. There are a number of JRC reports that explore how the Sec Data Platform has evolved and adapted over time. Fore more detailed information, please consult:

JRC140529, *JRC Terrorism/Extremism Database - Large Language Models and prompt engineering for news article summaries*, 2025

JRC135160, *JRC Terrorism and Extremism Database - Data Quality Optimisation for Machine Learning*, 2024

JRC134961, *Use of Large Language Models for location detection on the example of the terrorism and extremism event database*, 2023

In addition, find out more on the JRC hub for general scientific research on Artificial Intelligence *here*.

Both preconfigured and personalised Newsletters can inform users about events at chosen frequencies (daily, weekly, monthly) providing a general overview or focusing on a particular subject, category or country.

https://counterterrorism.ec.europa.eu/