The latest by AIDA Project

Second Edition of AIDA's Bi-Annual Newsletter

July 2021

Welcome to the second edition of AIDA’s newsletter!

The AIDA project is now in its 10th month of implementation and is getting prepared for the first review. In these months, the consortium has achieved several milestones, as it has:

- Analysed and further defined the use cases (cybercrime and terrorism);
- Collected and defined the user requirements;
- Started the analysis of social, ethical, legal and privacy issues;
- Defined and specified the overall AIDA system;
- Defined the architecture of the AIDA system.

AIDA will equip Law Enforcement Authorities in Europe with AI-enabled tools to prevent, detect, analyse and combat criminal activities.

In the upcoming months, partners will focus on:

- Completing the analysis of citizens' security perceptions and impact of AI capabilities by LEAs on societal resilience
- The development of tools for data acquisition and for AI actionable intelligence.

New technologies to fight crime: a collaboration with PREVISION

The Horizon 2020 PREVISION project is developing advanced, almost-real-time, analytical support for multiple big data streams, subsequently allowing their semantic integration into dynamic and self-learning knowledge graphs that capture the structure, interrelations and trends of terrorist groups and individuals, cybercriminal organisations and organised crime groups.

Given the numerous commonalities in the goals of AIDA and PREVISION, as well as the similar nature of the challenges faced by the two projects, a collaboration was launched, which we are happy to present to our communities.

Such collaboration will focus on the following activities:

1. Organisation of joint dissemination activities and public events, including demonstrations;
2. Organisation of joint activities for societal impact assessment, to foster engagement of networks and work with focus groups of citizens who will provide their assessment of the societal impact of AI technologies used by LEAs, as well as share information and valuable suggestions;
3. LEAs of the two projects will be invited to participate (by invitation) to the public demonstration events organised by both projects for testing and evaluation;
4. Information sharing regarding available datasets that have been secured by the technical partners of the two projects;

5. Creation, improvement and sharing of domain ontologies, both for knowledge modelling and for exporting of data for prosecution. Joint activities and a thematic workshop will be organised to share experience on Semantic Engines and Knowledge Representation;
6. Specifications of selected tools that could potentially be used by both projects will be prepared by technical partners. Technical discussions will include the approach used for the wrapping of technical components and GUIs of the platforms.

A deep dive into the use of AI for public security: contributions and challenges of automatic speech recognition

Every day, terabytes of data are created and shared worldwide. People post their experiences and opinions over the Internet through different channels, like mails, blogs, and social networks, generating a huge flow of information. This is an appealing scenario for crime activities, which can exploit private and public communication channels to advertise propaganda factions, recruit new elements and strengthen their networks.

In this context, Artificial Intelligence (AI) and Machine Learning (ML) techniques become essential to analyse the vast, and rapidly increasing, amount of data available. Speech and language technologies, in particular, are critical to allow a fast and automatic inspection of dynamic contents, like audio and video files, which otherwise would require years of manual analysis. Indeed, audio processing technologies can actively contribute to fight the spread of organized crime and terrorism by providing Law Enforcement Agencies (LEAs) with valuable tools enabling the management and filtering of distinct information types.

Automatic speech recognition (ASR) plays a very important role on analysis of speech inventories, as it generates metadata (transcriptions and time stamps) either to allow for the indexation of spoken terms or to feed Natural Language Processing (NLP) tools in order to produce higher-level knowledge from the transcriptions. If ASR technology is fast enough, like the one used in AIDA, an ASR can also be used to monitor communications of target individuals in real time (e.g. lawful interception of phone calls) causing suspicious information to be promptly identified.

ASR technology, however, is language-dependent, that is, before a language can be supported, three main models must be built:

- acoustic models (statistical)
- pronunciation models (statistical or rule-based)
- language models (statistical)

Acoustic models (AM) are meant to produce phonetic probability values for each acoustic feature vector present at its input, every few milliseconds. They are trained through ML techniques over large amounts of spoken utterances (manually) transcribed at the orthographic level. Pronunciation models (PM) are used to bridge the gap between the phonetic and orthographic levels, since, by generating word pronunciations, they provide a means to convert phonetic sequences into known words (i.e. words belonging to the ASR vocabulary). Such models play a pivotal role both in run time (when the ASR is transcribing speech) and in acoustic model training, as they make it possible to convert the utterances' orthographic transcriptions into phonetic sequences that are later aligned with the speech signal in order to model the acoustics of all phonetic segment productions. Last, but not least, because multiple word sequences can give rise to the same phonetic sequence ("ice cream" vs "I scream"), language models (LM) are used to disambiguate such cases by penalizing odd word sequences or, at least, not so frequently observed ones. In addition to being language-specific, such models are, to some extent, task-specific (e.g. word sequences produced in a live sports report vs in a City Council meeting). Therefore, such models are trained over a large number of task-specific (or in-domain) texts. Moreover, since they are aimed at transcribing spoken utterances, on the one hand, and oral language differs from written language in many aspects (e.g. more interjections, more word repetitions and contractions, shorter sentences, more self-reference words), on the other hand, such texts should, whenever possible, consist of spoken utterance transcriptions. The ASR vocabulary (typically, around 100,000-sized) is mostly composed of the most relevant (frequent) word types observed within the collection of domain-specific texts used for training the LMs.

Currently, state-of-the-art speech recognition systems are able to achieve performance comparable to human transcribers when dealing with data presenting clear acoustic conditions (e.g., read or planned speech like the one found in broadcast news). Nevertheless, speech recognition accuracy easily gets lower when dealing with a different scenario containing adverse conditions, such as in the case of spontaneous speech or environmental disturbances. The performance degradation is usually due to mismatches between the type of data used for training the models and the actual data they have to work with. Data mismatch, however, represents only one of the many challenges that the identification of terrorism-related content will pose.

It is expected that the multimedia content to analyse will contain a variety of background noises, which will hamper the comprehensibility of the foreground speech also for a human listener. Non-native accents and the reference to foreign-origin person and place names (pronunciation models are less effective in predicting the acoustic realization of such words) will challenge the speech recognizer with unexpected or atypical sounds (entropy of AM probabilities will be increase). In the same line, an intense emotional state is also foreseen to contribute with a detrimental impact on voice analysis. This is typically observed in hate speeches produced by charismatic leaders either for recruiting new followers, or for convincing them to take action against their "opponents". Finally, another probable source of disturbance could be chants and prayers, expected to be present as a constant background in the audio.

In this context, adaptation approaches usually allow to partially overcome this issue by reducing the discrepancy between data. However, for security reasons, it is very difficult, if not impossible, to use typical audios gathered by LEAs in their investigations to improve the speech recognition models (AM, PM and LM). Therefore, such adaptation must be carried out either through the deployment of automatic model building tools that can run with no technical supervision on the LEAs premises, or through exploiting publicly-available data that is close to the target domain.

Both approaches have their pros and cons. On the one hand, if the pronunciation model is making mistakes, they must be identified and corrected manually, and LEAs officers are not aimed or expected to be capable of doing that, so even though the first approach enables the models to be trained over appropriate materials, pronunciation mistakes can limit the effectiveness of such models. On the other hand, following the second approach cannot assure improved results in all target speaking styles and conversations (language mismatch in phone call conversations). In spite of all cons of the second approach, it was, so far, the one possible.

The first adaptation procedure focused on the language models and, in the absence of a sufficiently large corpus of terrorist audio transcriptions, a large set of general-purpose news story transcriptions available in the internet was crawled. Then, a large list of terrorist-related keywords was compiled from news articles about the most famous terrorist attacks occurred in the last 25 years, Europol reports on Terrorism, and famous terrorist manifestos, in addition to a LEAs-provided keyword list. The full keyword list was then used to rank the news articles collected earlier in accordance with their keyword density. The first Terrorism-adapted LM built in AIDA was derived from the higher-ranked articles of the corpus (top 40%), and proved to give rise to observable transcription improvements. Further adaptation, both at AM and LM level, will be carried out when enough in-domain data is available.

News and opportunities:

Europe fit for the Digital Age: a EU proposal on AI

In April, the Commission published its legislative proposal for new rules on trustworthy Artificial Intelligence in the EU

Read more

Citizens' counter-strategies to the use of AI by the police

Our partner CENTRIC will present a paper during the 2021 Conference on Information and Knowledge Engineering, on 26-29 July. Stay tuned to read it!

Read more

European Union
Terrorism
Situation and
Trend report
2021

EUROPOL

Terrorists attempted to take advantage of the pandemic: TE-SAT 2021 is published

The 2021 EU Terrorism Situation and Trend Report outlines the features, facts, figures and trends concerning terrorist attacks and arrests in the EU in 2020.

Read more

Countering right-wing terrorist online propaganda

On 27 May 2021, the 1st Referral Action Day against right-wing terrorist online propaganda was coordinated by Europol, joined by 28 international partners. A total of 1,038 items were assessed for referral to the OSPs.

Read more

This project has received funding from the EU's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 883596. The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

AIDA is coordinated by

Mr. Ernesto La Mattina

Engineering Ingegneria Informatica S.p.A.

Contact us at:

info-aida-project@eng.it