The ROXANNE platform for supporting Law Enforcement practitioners in criminal investigations by analysing multi-modal data

Costas Kalogiros
(AEGIS IT Research)
contact: ckalog@aegisresearch.eu

Joint work with:
Srikanth Madikeri (IDIAP) & Andreas Alexopoulos (AEGIS IT Research)

June 9th 2022

This project has received funding from the European Union’s Horizon 2020 Work Programme for research and innovation 2018-2020, under grant agreement nº833635.
LEAs’ pain points

- ROXANNE run a survey on LEAs’ requirements
- 121 responses were collected from 40 countries highlighting that amount of data to be processed and analysed is the main pain point
Knowledge level related to ROXANNE learning topics

- ROXANNE also ran a survey on LEAs’ training requirements
- 31 responses were collected from 8 European countries

Only ~10% of the participants have received some training on all 4 key ROXANNE technologies!
ROXANNE Objectives

LEA

Speed up investigation of large complex criminal cases (mixing SIM, cross-border, multiple-languages, use of nicknames, detection of leader, innocents, …) and train LEAs on using state-of-the-art technologies

DEV

Platform to combine evidence extracted from multimodal sources with network analysis

RESEARCH

Bi-modal interaction between processing technologies and network analysis

ETHICAL

Implementing an ethics-by-design and privacy-by-design approach
ROXANNE Overview
Speaker Verification & Network Analysis

A wiretapped call

Voiceprint repository

New wiretapped call

Speaker verification process

Updated Voiceprint repository

Network Analysis

This project has received funding from the European Union’s Horizon 2020 Work Programme for research and innovation 2018-2020, under grant agreement n° 833635. © ROXANNE Consortium
The ROXANNE synthetic dataset

- ROXANNE prepared a synthetic dataset of 200+ simulated wiretapped phone calls and related metadata (i.e., Call Details Records (CDR) info)
- The screenplay involves three drug dealing cases
  - A drug distribution case in Prague universities
    - The police has wiretapped the two mobile phones of Krystof, as well as some of his partners.
  - A drug distribution case in Prague city centre
    - The police has wiretapped the mobile phones of Alexo
  - A drug lab
    - The police has wiretapped the mobile phones of Tuan and Hoang
- The investigator uses the Autocrime platform in order to identify whether and how these suspects are connected.
Training Platform

https://roxanne.kemea-research.gr/

Registration is via invitation only

https://docs.google.com/forms/d/e/1FAIpQLSc9kiBOvqmWV2YAqKnkAIHQLGG49xQhMOFozKmcbR5D-eMaQ/viewform
List of courses

- Autocrime Platform
- Voice Activity Detection (VAD) and Speaker Diarization (DIAR)
- Speaker Identification (SID)
- Voiceprint Extraction
- Gender Identification
- Automatic Speech Recognition
- Topic Detection
- Named Entity Recognition
- Mention Network
- Network Analysis
Next Steps

- The Autocrime platform is currently being extended to support additional modalities (e.g., video) and export results to other known formats

- On Oct 6th 2022, the ROXANNE project will run the final field test in Lyon, France (INTERPOL premises)
  - Let us know if you are interested to participate on site

- On Jan 2023, the ROXANNE project shall release:
  - the ROXANNE platform free of charge to interested LEAs
  - the synthetic dataset to other researchers (expanded to also include videos and chat messages)
Thank you!

Any questions?
Backup slides
Autocrime components and interactions

1. GUI
2. Back-end
   - Core Back-end
   - Speech
   - Video
   - NLP
   - Network Analysis
3. Core Back-end
   - JSON Output
4. Middleware
5. Export
6. Local Filesystem
   - temp JSON Output
7. Local Filesystem
   - i2_Output
8. GUI
9. autocrime
10. GUI

Autocrime components and interactions.
# Evaluation of Autocrime Technologies

<table>
<thead>
<tr>
<th>Technology</th>
<th>Partner(s)</th>
<th>Method</th>
<th>Status</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker Verification</td>
<td>BUT &amp; Phoena</td>
<td>Energy-based VAD + VBy</td>
<td>[green]</td>
<td>DER 5.91%</td>
</tr>
<tr>
<td>Speaker ID</td>
<td>BUT &amp; Phoena</td>
<td>ResNet architecture</td>
<td>[green]</td>
<td>99.95% speaker accuracy</td>
</tr>
<tr>
<td>Open set Speaker ID</td>
<td>BUT</td>
<td>Same as SID</td>
<td>[green]</td>
<td>90%</td>
</tr>
<tr>
<td>Gender ID</td>
<td>Idiap</td>
<td>GMM-based</td>
<td>[green]</td>
<td>66% accuracy</td>
</tr>
<tr>
<td>LID</td>
<td>Idiap</td>
<td>ECAPA-TDNN on Speechbrain</td>
<td>[green]</td>
<td>76.5% accuracy</td>
</tr>
<tr>
<td>ASR</td>
<td>Idiap</td>
<td>Wav2Vec 2.0 + LM, English</td>
<td>[green]</td>
<td>46.3% Word Error Rate</td>
</tr>
<tr>
<td>ASR</td>
<td>HENS</td>
<td>Hybrid TDNN AM + 3-gram LM, English</td>
<td>[green]</td>
<td>48% Word Error Rate</td>
</tr>
<tr>
<td>Topic</td>
<td>Idiap</td>
<td>Zero-shot</td>
<td>[green]</td>
<td>Not measured yet</td>
</tr>
<tr>
<td>NER</td>
<td>USAAR</td>
<td>BERT-based</td>
<td>[green]</td>
<td>Not measured yet</td>
</tr>
<tr>
<td>Mention network</td>
<td>USAAR</td>
<td>Custom co-reference analysis module</td>
<td>[green]</td>
<td>89% accuracy</td>
</tr>
<tr>
<td>Network analysis</td>
<td>LUM</td>
<td>Community Detection, Social Influence Analysis, Link Prediction and Node Embedding</td>
<td>[green]</td>
<td>Community Detection: F1-score 75%, Link Prediction: 67.22% accuracy (Top-5)</td>
</tr>
</tbody>
</table>