INSPECTr Project

Intelligence Network & Secure Platform for Evidence Correlation and Transfer

Quarterly Newsletter: Fourth Edition

Edition: December 2021

Dear Colleagues,

Welcome to the INSPECTr project newsletter, a guide to our latest work and news.

INSPECTr Principal Objectives Brief Summary

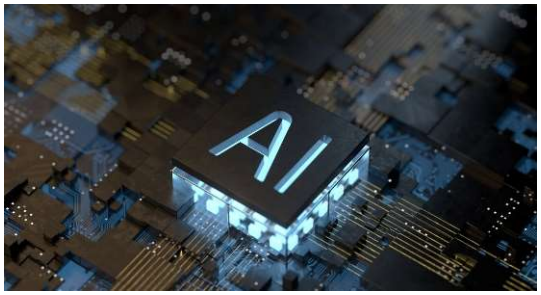
Intelligence
Network &
Secure
Platform for
Evidence
Correlation and
Transfe**r**

To develop a shared intelligence platform and a novel process for gathering, analysing, prioritising, and presenting key data to help in the prediction, detection, and management of crime in support of multiple agencies at local, national, and international level. This data will originate from the outputs of free and commercial digital forensic tools complemented by online resource gathering. The final developed platform will be freely available to all Law Enforcement Agencies (LEAs).

INSPECTr Newsletter Fourth Edition

In this, our fourth edition, we will provide updates on our last quarter activities, information about meetings and events attended, our upcoming events, recent dissemination activities and blogs on the subject of **AI as an Assistive Technology for LEAs** provided by CCI, (NUID) and our Consortium Partners, Inlecom (ILS), Trilateral (TRI), and Gendarmerie (GN).

AI as an Assistive Technology for LEAs



- **Artificial Intelligence (AI) Research Methodology and Application**
- **Natural Language Processing (NLP)**
- **Image processing**
- **Ethical, Legal, and Societal Impact Assessment Overview**

Artificial Intelligence (AI) Research Methodology and Application
Blog provided by INSPECTr partner Daniel Camara of the French Gendarmerie (GN)

Artificial Intelligence (AI) is an interesting research field and, depending on whom you ask, it either has no formal definition or has hundreds. Even specialists do not fully agree, with different researchers having different definitions and interpretations of AI, which comes to the fact that intelligence itself has different interpretations. What is intelligence? Some believe it must be linked to rationality, critical thinking, high order brain functions, others seek intelligence from things that may not even have a brain at all. However, a relatively well-accepted intuition is that a computer program belongs to the AI class if capable of behaving/giving answers, close to the ones a human would if presented to the same kind of situation/question. Another concept that is intrinsically linked to AI is the notion of error. If you have a way to solve a problem, that gives a perfect solution every time…. That, most probably, does not fit the general concept of AI. In some sense, it does not contradict the previous intuition, as humans solve problems in an instinctive way, and yes from time to time we make mistakes, *"to err is human"*. AI methods are heuristics, intuitions, that are not necessarily full-proof methods. The word heuristic originates from the Greek word "heuriskein", meaning "to find" or "discover". A heuristic is a practical method, which comes from, for example, an intuition, an abstraction, or even a pattern generalisation. It is an idea that is not sure to work every time but that, in reality, behaves pretty well.

As the definition of the AI field is "flexible", to say the least. Many different methods/heuristics are used to try to reach this "similar to human" kind of answer. This implies that the AI field is constantly evolving, and new methods are created to explore ways to solve general problems or work well in a specific domain. One of the most iconic artificial intelligence methods is the neural network. It is based on abstractions of how the human neurons work and are organised. Rosenblatt proposed the formalisation of a perceptron (the abstraction of a neuron) in the 50's, and even the modern deep learning approaches rely on his concepts. A perceptron is a simple sum function that add up the different received stimuli and have an activation function to propagate a signal. A neural network is an organisation of these basic units in series and layers. The modern deep networks are called deep because they have many layers of perceptrons. When considered individually, perceptrons may not be that impressive, but when organised in a network and trained, the provided answers may be quite impressive, even though there is no guarantee that the answers will always be correct. Deep learning methods are currently used in INSPECTr for example to perform facial recognition, for detection of child pornography and cars in images, text translation, among others.

Even if deep learning is quite in vogue in the last few years, and with good reason (it is a general method that can be used to solve an important range of problems), it is not the only AI method. A broad range of other methods exist. For example, the crime forecasting implemented in INSPECTr is based on the analysis of time series and on the intuition that crimes have an opportunity factor, which is basically random, but also a facilitator factor, which means not all the places and times are fit for the commission of a crime. If we can map these places, we can have a good intuition of where crimes may happen in the future. Three things are required for a crime, a victim, a perpetrator, and a place where these two meet. Not necessarily at the same place at the same time, but somehow the paths of these have crossed at some point. However, some places are more fit for a crime commission than others. For example, a dark alley or a tourist attraction with many distracted people. Moreover the crimes have a seasonal component, e.g., not much happens in a winter sky station during the summer. However, it is also a tendency factor, e.g., a team that is committing residential burglary tends to spread their activities in a region over a period of time. So, if we analyse the criminality of different regions, considering criminal seasonality and tendencies, to forecast the next days/months criminal levels of activities it is not possible to guarantee that the

criminal pattern in a given region will be the same tomorrow as it was today, but most probably, tomorrow's criminal profile for that region will be close to today than any day two months behind. Moreover, in general, it can be assumed that this week will present more similarities with the same week of the last year than with the same week six months ago.

However, a series of other possible methods could be used for criminal analysis and LEAs could profit from them. We could highlight genetic algorithms as a popular and reasonably general method. Genetic algorithms are abstractions of the natural evolutionary process. Through crossover and mutations only the fittest individuals survive and so produce better chances for their offspring, and through generations, the population will evolve to get the best of their environment. AI engineers' interpretation of it is to encode the available parameters linked to the solution of a given problem, are organised in genomes, and randomly assign values to these genomes. Each one of these genomes is compared to the objective function, which gives a score to that genome, and gauges how fit that specimen would be, regarding the objective function. Through crossover and mutations, we change the genetic pool and as generations pass by, we expect that each generation will be more fit than the previous one. At some point, evolution would provide a good enough answer to the problem we want to solve.

Hundreds of other heuristics exist, some are more suited to solve LEAs' problems, some less. Some people could even argue that we would need to apply LEA based heuristics to solve some of the specific LEA-related problems, and they are probably right. However, AI methods should be used on "hard to solve problems", the ones we cannot distinguish a clear pattern linked to the problem or solution to that problem. If one can distinguish a pattern that can be transformed into a set of defined rules, AI methods may not be required to solve it. Considering that AI-based methods imply an error probability, if you solve the problem using standard and precise algorithms, it will always be better.

Natural Language Processing (NLP)
Blog provided by INSPECTr partner Panos Protopapas of Inlecom (ILS)

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) and linguistics concerned with the task of making machines "understand" text found in documents and automatically extract information contained in these to automate various processes by using it. Some common NLP tasks are described below, ordered from those considered relatively easier for a machine to perform to those considered more difficult:

- *Part of speech tagging (POST):* to determine the part of speech of words in a document, given the context. For example, in the phrases "click on this link" and "some languages include click sounds", the word click should be recognised as a verb and adjective respectively.
- *Lemmatisation:* to remove the inflectional endings of words appearing in a document (due to different tenses, cases, voices, genders, etc.) and retrieve a "base" word (lemma).
- *Named entity recognition (NER):* to classify named entities found in a document into pre-defined categories (locations, persons, organizations, dates, etc.), with the addition of more categories possible via training.
- *Sentiment analysis:* to classify the polarity of a given document, that is, whether it is of a positive, negative, or neutral sentiment.
- *Topic modelling:* to infer "abstract" topics occurring within a document. The main idea is that when a document is about a particular topic, some words will appear more often than others. For example, a document about *cars* is more likely to contain the words *engine*, and *transmission* and

less likely to contain the words *bone* and *bark* which are more likely to be present in a document about *dogs.*

- *Summarisation:* to sum up the main points of information contained in a document and create a short, coherent, and fluent summary outlining the text's major points.

Although the output of these tasks might not seem very useful on a per-task basis also considering that the machine performance in these tasks is not yet at par with human performance, one should not overlook the speed increase that the usage of NLP can provide, in situations where the number of documents to check is too large and checking them manually is not realistically possible or economically viable. To this effect, the synergies between these NLP outputs and the positive effects they can have when post-processing searching or filtering through these documents is of interest.

When searching these documents, the *POST* and *Lemmatization* techniques can greatly "generalise" the results of a search query. For example, in the search query "he is holding a gun" the pronoun (he) can be overlooked, while the fact that the verb form is provided in the present continuous tense (is holding) and the article-noun pair in singular number (a gun) can be non-important. This results in the phrase "they were holding guns" to appear in the results. Moreover, if the NER model is trained to classify firearms, the term gun can also be "generalised", and the search query can return "she held an AK47".

Furthermore, the outputs of the NER, sentiment analysis and topic modelling techniques can greatly enhance the filtering techniques of users going through NLP-analysed documents. If said documents have been tagged with the outputs of the NLP models, a user could filter for example for documents mentioning a particular location and/or date, having a negative sentiment, and some firearm appearing in the list of words defining the topic.

Finally, the summarisation results can also aid the human processing of documents by enabling users to scan through a document for an accurate and brief summary of its contents prior to deciding on investing the time to read it.

Image processing in INSPECTr
Blog provided by Ben Roques of the CCI project coordination team (CCI/UCD)

The INSPECTr platform implements an image processing system aiming at reducing the workload of investigators when dealing with a very large amount of media content. After media extraction through the use of INSPECTr platform's specific tools, or after ingesting media directly into the platform, law enforcement officers can automatically annotate pictures using different image recognition models. Those annotations can then be filtered and searched depending on the needs of the investigation.

The system needs to be able to ingest a large amount of data. Oftentimes, storage supports linked to an investigation contain thousands of media files. Law enforcement officers should be able to send all the files to be processed in the background and then search for/prioritise investigative actions. A model is a machine learning classifier that was trained on specific data. For the purpose of INSPECTr, the image processing pipeline only uses convolutional neural networks as those were proven very effective for image processing tasks. There are many different architectures, with new ones frequently published in scientific journals. The system possesses many different neural networks trained on for a specific task. It was designed to be very modular. Adding and removing models needs to be easily done. This is due to the fact that the field of machine learning is always evolving and moves really fast. That is why the system tries to be model agnostic. Model specific pipelines are pushed at the very end of the data flow.

The system is organised around micro-services. Pictures to be analysed are sent to a main endpoint that forwards tasks to a queuing system. A worker collects the tasks from the queue and forwards them to a model micro-service. Models can be swapped, added, or removed without bringing the system down. Once the model is finished, a webhook is sent to the main endpoint. The result is then cached into a database.

The system is not designed to make automated decisions. The idea is to triage a large number of pictures to lighten the investigator's workflow. For instance, cases related to child sexual exploitation oftentimes contain a lot of media files. The system is able to detect the presence of children and/or nudity in a set of pictures. The investigator can then, for example, search for the face of a suspect inside the set of pictures containing both nudity and children using face recognition technologies. This example is only a small use case. It is important to note that because many more models can be added, the possibilities will vastly increase over time.

Ethical, Legal, and Societal Impact Assessment Overview
Blog provided by INSPECTr partner Joshua Hughes of Trilateral Research (TRI)

As part of the ethical approach to research in INSPECTr, Trilateral conducted an impact assessment of ethical, legal, and societal issues that could be raised by the INSPECTr project and technologies. As has been discussed in other blogs, privacy and data protection issues are also key considerations. It is important that these issues are considered so that the final version of the INSPECTr platform developed in the project is the most ethical, legally optimal, socially acceptable, and privacy-respecting as possible. It is crucial that this work is done because the intended recipients of the INSPECTr platform are law enforcement agencies whose officers come from, and contribute to, the society that is being policed.

As noted in other blogs, Trilateral analysed both the project and technologies being researched to develop a list of requirements, and some of these were discussed in more detail during workshops (the list of requirements is provided in D8.5 Ethical, Legal and Societal requirements for the INSPECTr platform and tools). The workshops helped to specify some of the details needed to fulfil some of the requirements and, for others, provided space where options for fulfilling the requirement could be discussed. In addition, through the Ethics and Privacy-by-Design work carried out by Trilateral, some additional requirements have been developed.

How all of these requirements can be fulfilled is an evolving conversation that will continue until the end of the project. As the technologies are still being researched, it is not possible to give a final overview as to how the requirements are being fulfilled. However, generally, it can be said that the INSPECTr platform is being researched in a way that enables data to be collected, analysed, and shared in appropriate ways that respect the applicable standards; end-users will be fully informed about what tools can and cannot do, what their limitations are, and how to use them through a detailed training provision. Overall, the ethical, legal, and societal requirements developed in the first part of the project are likely to be fulfilled in the second half, leading to a platform that will enable law enforcement to engage in analysing large amounts of data from complex investigations in an ethical, legal, and socially acceptable way.

- INSPECTr Monthly Project Meetings
- INSPECTr Weekly Technical Meetings
- INSPECTr LSG Monthly Meetings
- Ethics Work Package Monthly Meetings

INSPECTr Monthly Project Meetings

These continue to be held monthly where an overview of the activities undertaken in each work package are reported on by work package leaders to the Consortium.

INSPECTr Weekly Technical Meetings

Weekly technical meetings are held in order to support the finer detail of the project development, give close attention to particular issues, and bring about resolution so that the project continues to develop on track between the monthly meetings.

INSPECTr Law Enforcement Steering Group (LSG) Monthly Meetings

These meetings provide a forum for collaboration between our law enforcement, technical, and ethics partners. It is of primary importance that that technical features under development in the INSPECTr platform are developed in a way that is relevant to and mirror the LEA investigative workflow. An ethical oversight at these meetings ensures that privacy and data protection requirements from applicable legal frameworks are observed, developed, and embedded.

INSPECTr Network of Living LEA Living Labs

Use Case mocked but realistic evidence has been prepared by our Law Enforcement project partners for experimentation and testing of the INSPECTr platform's functional and non-functional characteristics and have been developed to utilise the majority of the technological developments. Numerous forensic analysers and intelligence gathering tools are required to process both digital and non-digital items. These Use Case scenarios, and their related evidential material, will be used to test the platform during the iterative Living Lab exercises that have been scheduled throughout the project.

Ethics Work Package Monthly Meetings

These meetings continue and are held to reinforce the project's Ethics-by-Design approach and allow time for deeper consideration and exploration of ethical issues that arise. In tandem with these activities there are regular consultations held with the project Ethics Advisory and Review Group (EARG), a mandatory board established to offer relevant independent expertise on ethics issues generally within the project and more specifically on the areas of processing of sensitive data (behavioural/personality and biometric data), social media data collection, risk of mass surveillance and data security.

Further Opportunities for INSPECTr Dissemination and Cross-Project Learning and Collaboration

Octopus Conference Lightning Talks 16-18 November 2021 Octopus conference 2021 (coe.int)

The INSPECTr project participated in a lightning talk at the Octopus Conference 16th-18th November 2021. The first part of the talk was a presentation of the overall project. The second part explained the ethical consideration of the use of machine learning to conduct investigation and how INSPECTr tackles those issues.

Overview of the INSPECTr Project - Presented by Ray Genoe - CCI, UCD.

"Imagine a harmonised approach to cybercrime investigations, where the analysis of big data is made possible by standardisation of all investigative tools. AI assisted technologies, linked case discovery and evidential exchange across jurisdictions are just a click away. This is what the INSPECTr project aims to achieve. It is an LEA led H2020 project that seeks to disrupt existing LEA processes for the benefit of law enforcement and citizens. The final platform will be freely available to LEAs and will be complemented by SME technology and a robust capacity building program. This talk will describe the project and briefly outline the challenges faced."

An AI-powered investigation tool - Presented by Ben Roques - CCI, UCD.

Artificial intelligence is generally deployed to replace humans. For example, Facebook's automatic content moderation aims to replace human moderators or Amazon grocery stores do not require any clerks. While tools that can take decisions without human interaction can be beneficial, there are extensive ethical issues. From military applications (killer robots) to healthcare (IBM Waston), humans need to stay in control over the final decision. The INSPECTr project aims to deploy artificial intelligence capabilities to guide and support law enforcement officers while keeping humans closely involved in the decision making process. This lightning talk focuses on ethical considerations in deploying AI-assisted investigation pipelines and how the INSPECTr project deals with the issue for its image classification component.

Europol Innovation Lab

Europol is dedicated in supporting the European Law Enforcement community to innovate and make the most of opportunities offered by new technologies.

On 13th October 2021 the French Gendarmerie were invited to present their research work in the first edition of Europol's Innovation series and among the works presented were some that have been developed in the context of the INSPECTr project. The INSPECTr innovative tools presented were the **automatic text translation tool** and the **stylometrics tool**.

FREETOOL 3 Showcase - A Showcase of FREETOOL Investigation Tools for Law Enforcement

In January of 2022, the FREETOOL project, nearing the end of its third cycle, held a showcase to demonstrate the current suite of tools to interested LEAs. The showcase also provided an opportunity to discuss the tools' future and demonstrate how FREETOOL's evidence visualisation, digital forensic and intelligence gathering tools are being enhanced through integration with the INSPECTr platform. CEPOL kindly hosted the webinar to ensure maximum impact. It quickly became one of their largest ever, with 609 participants from 42 countries and organisations attending the day-long event. The feedback and increased stakeholder interest were very encouraging. We are very grateful to both FREETOOL and CEPOL for their fantastic support.

Conferences, Workshops, and Future Events

INSPECTr Consortium Attendance at Conferences and Workshops

- Innovation Demo Series - 13th October 2021, hosted by the Europol Innovation Lab.
- Octopus Lightning Talks - 17th-18th November 2021, hosted by Council of Europe.
- 2021 CEPOL Research and Science conference Vilnius - 1st- 3rd December 2021. The INSPECTr project had registered for this event which was postponed.
- Freetool 3 Showcase, hosted by CEPOL. Held via webinar on 26th January 2022.

INSPECTr Consortium Attendance at Forthcoming Events

- EAFS 2022 - Stockholm - May 30th - June 3rd 2022
  EAFS is a triannual conference on forensic science hosted by ENFSI (European Network of Forensic Science Institutes).
- SRE 2022 - Security Research Event
  1st-2nd March 2022

Closing

We look forward to updating you further in April 2022 with our fifth edition of the INSPECTr Newsletter. In the interim, communications from our readers are welcome and if you wish to contact us or subscribe to our Newsletter you can e-mail us directly at inspectr@ucd.ie. Further information and updates can also be found on our project website https://inspectr-project.eu/.