Coalition Situational Understanding Via Adaptive, Trusted and Resilient Distributed Artificial Intelligence Analytics

Alun Preece
Cardiff University
UNITED KINGDOM

preecead@cardiff.ac.uk

Dave Braines
IBM Research Europe
UNITED KINGDOM

dave_braines@uk.ibm.com

Federico Cerutti
University of Brescia
ITALY

federico.cerutti@unibs.it

Gavin Pearson
Dstl - Porton Down
UNITED KINGDOM

agpearson@dstl.gov.uk

Lance Kaplan
DEVCOM ARL
UNITED STATES OF AMERICA

lance.m.kaplan.civ@mail.mil

ABSTRACT

Artificial intelligence (AI) and machine learning (ML) promises transformative effects on coalition multi-domain and hybrid operations. AI/ML approaches that support situational understanding in the context of ad-hoc coalition operations at the tactical edge are of considerable current research interest. Coalition operations need distributed AI/ML that is robust to contested and complex multi-actor situations. Information with a high degree of complexity needs to be collected across a range of sensed modalities and processed at high tempo, aligned with human needs and capabilities. Research carried out in the joint US/UK Distributed Analytics and Information Science (DAIS) programme since 2016 is addressing coalition needs for adaptable, trusted and resilient AI/ML: adaptable AI refers to AI systems which can rapidly adapt in dynamic situations; trusted AI means that human users are able to rapidly calibrate their trust in AI systems; and resilient AI concerns AI systems which are resilient to adversary attacks and deception.

This paper focuses on DAIS research centred on the rapid exploitation and integration of coalition AI/ML assets including both symbolic (logic-based) and subsymbolic (deep neural network-based) approaches. To provide a focus for the paper, we consider settings involving detecting patterns of interrelated events that form situations of interest where only sparse training data (for ML) is available. Rapid trust calibration is addressed via a combination of explainable AI - involving both symbolic and subsymbolic approaches to explainability - and effective management of uncertainty - considering both aleatoric and epistemic types of uncertainty. While not the primary focus of this paper, resilience is considered by showing that the integrated neuro-symbolic system performs robustly against targeted model-poisoning adversarial attacks, and also that the processing of multimodal sensed data by explainable AI/ML services makes the integrated system much harder to attack. For easier assimilation of the programme of work, we use a single integrated case study of a coordinated attack in an urban setting based on the NATO Anglova exercise.

1.0 INTRODUCTION

Military operations typically involve working with coalition partners to resolve rapidly evolving situations where adversaries are adapting their tactics, techniques and procedures, and the behaviour of the civilian population is changing. Achieving coalition situational understanding (CSU) involves both insight, i.e., recognising existing situations, and foresight, i.e., learning and reasoning to draw inferences about those situations, exploiting assets from across a coalition, including sensor feeds of various modalities, and analytic services. Thus, military information processing systems need to be able to recognise significant patterns of activity which are distributed in time and space, in near real-time, without generating too many false alarms. In the language of information processing, this requires the ability to recognise the relationship between a set of individual events. Recent years have seen significant advances in artificial intelligence (AI) and machine learning (ML) technologies applicable to CSU. Deep learning-based AI systems are constantly improving their ability to recognise such individual events. However, such state-of-the-art ML techniques based on deep neural networks require large volumes of training data; unfortunately, representative training examples of situations of interest in CSU are usually sparse. Moreover, to be useful, ML-based analytic services cannot be 'black boxes'; they must be capable of explaining their outputs and quantifying their uncertainty in their decision-making to allow users to calibrate their trust in AI-based assets, often rapidly.

We describe an integrated CSU approach that combines deep neural networks with symbolic learning and reasoning, as well as techniques for explainability and uncertainty-awareness, integrated within an environment that facilitates human-AI collaboration. The approach supports processing of multi-modal sensed data, and emphasises two key features:

- Humans are able to inject new knowledge, or hypotheses, about patterns of activity rapidly through addition of new rules – which means patterns can be recognised in situations where there is insufficient time or data to train a deep learning model;
- The need for training data is significantly-reduced (especially important when examples of the patterns-of-interest are relatively rare), AI model training is faster, and detection accuracy is improved over 'pure' deep learning approaches.

Our approach is loosely-coupled, based on open source software in an open architecture, and the AI-based assets can run on edge-of-network devices, thus being suitable for tactical sensor systems. Our integrated approach is intended to result in AI-based CSU systems for Defence that are (1) *adaptive*, able to learn and adapt at the 'pace of the fight'; (2) *trusted*, meaning that human users are able to rapidly calibrate their trust in the AI-based assets; and (3) *resilient* to adversary attacks and deception. The paper reports research carried out in the joint US/UK Distributed Analytics and Information Science (DAIS) programme since 2016. For easier assimilation of the programme of work, we present the key scientific and technological components in the context of an integrated scenario: a coordinated attack in an urban setting.

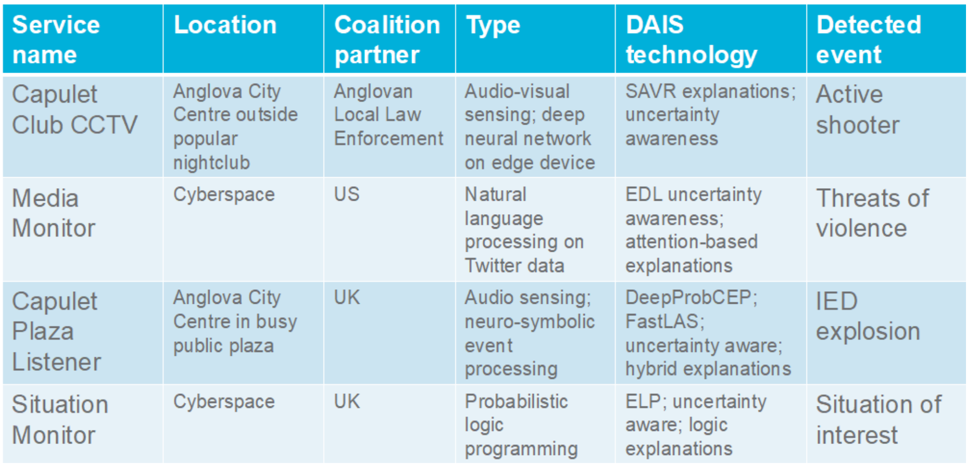
2.0 OVERVIEW OF THE SCENARIO & AI-BASED SERVICES

We demonstrate the research in the context of monitoring a rapidly-evolving situation in the NATO Anglova urban setting [1] via diverse coalition AI assets processing multimodal sonsor data with management of situational uncertainty. We envision a situation where events indicate growing threats to, and attacks on, a section of the Anglova civilian population, the 'Capulet' community. We focus on four AI services, each chosen to showcase particular DAIS research technologies, summarised below and in Table 1:

- **Capulet Club CCTV:** a CCTV processing service owned by Anglovan local law enforcement located outside a popular nightclub frequented by members of Anglova's Capulet community detects an *active shooter event* via deep neural network (DNN) processing of audio-visual data running on an edge device. To build trust and assure robustness, the service is able to provide multi-modal explanations for its outputs with associated uncertainty.
- **Media Monitor:** a service owned by the US coalition partner and located in cyberspace detects *threats of violence* directed against the Capulet community via DNN natural language processing (NLP) of Twitter data. Again, the service can offer appropriate explanations and is equipped to provide uncertainty awareness.
- **Capulet Plaza Listener:** an acoustic sensor-based service owned by the UK coalition partner and located in a busy public plaza in the Capulet part of town processes audio data streams via a neuro-symbolic AI architecture. This service is able to detect events such as an *IED explosion* when trained on a relatively small set of training data.

- **Situation Monitor**: this service, owned by the UK coalition partner and located in cyberspace, performs probabilistic logic programming via the Evidential Logic Programming engine [2] with rules injected by subject matter experts to detect that the three preceding events (*active shooter*, *threats of violence*, *IED explosion*) constitute a situation of interest, i.e., escalating violence directed against the Capulet population of Anglova.

Table 1: Overview of scenario AI services.



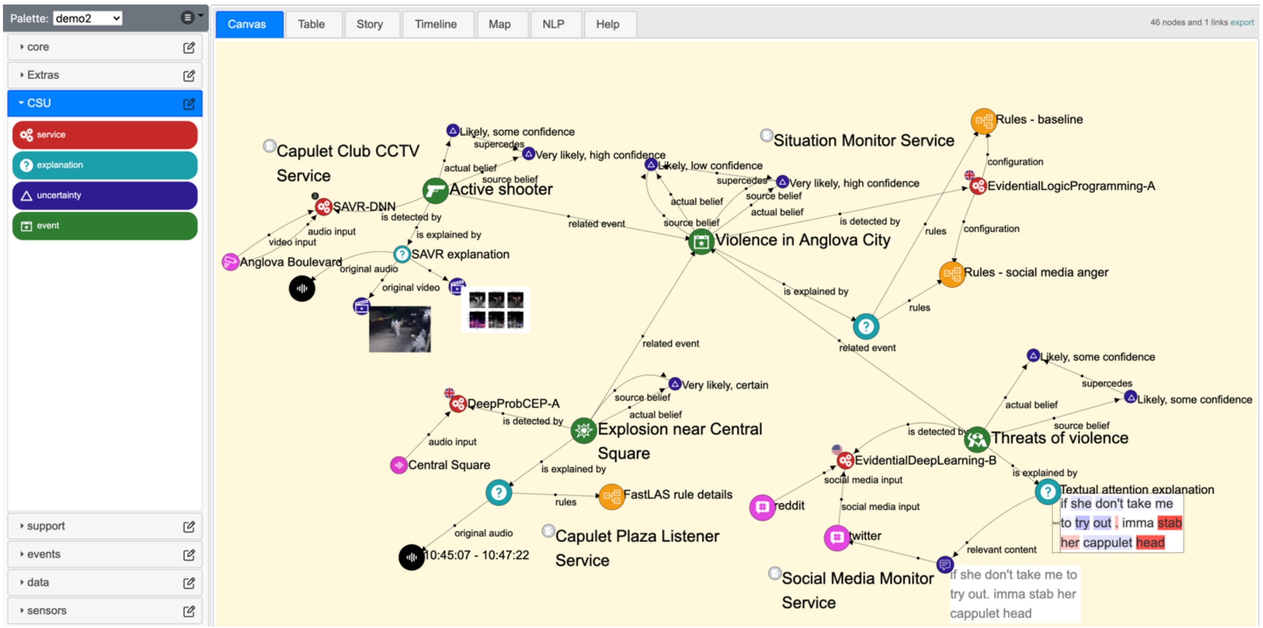
Service name	Location	Coalition partner	Type	DAIS technology	Detected event
Capulet Club CCTV	Anglova City Centre outside popular nightclub	Anglovan Local Law Enforcement	Audio-visual sensing; deep neural network on edge device	SAVR explanations; uncertainty awareness	Active shooter
Media Monitor	Cyberspace	US	Natural language processing on Twitter data	EDL uncertainty awareness; attention-based explanations	Threats of violence
Capulet Plaza Listener	Anglova City Centre in busy public plaza	UK	Audio sensing; neuro-symbolic event processing	DeepProbCEP; FastLAS; uncertainty aware; hybrid explanations	IED explosion
Situation Monitor	Cyberspace	UK	Probabilistic logic programming	ELP; uncertainty aware; logic explanations	Situation of interest

3.0 THE COGNI-SKETCH HUMAN-AGENT SENSEMAKING ENVIRONMENT

Our scenario features a situational understanding example based on the collaboration between human and machine agents in a sensemaking software environment. The purpose of this environment is a virtual workspace where human and machine agents can rapidly build and exchange knowledge through two key user interface affordances: *tellability*, injecting new knowledge and information (e.g., rules and facts), or *explainability*, going deeper into rationales of why information exists or has been added. Figure 1 shows an overview of the Cogni-Sketch environment [3]. The palette on the left-hand side provides a set of capabilities and services that can be used, customized or extended, and shared as needed for each use-case. For example, we have the standard core entity types, as well as specific additional extensions to support CSU:

- **Service** (red ‘cog’ icon): the AI machine services that have been deployed into this environment, usually in a distributed setting onto the sensors themselves, but they could also be run remotely in cloud infrastructure.
- **Explanation** (light blue question mark icon): additional information from the service which provides some form of explanation for any events that are detected.
- **Uncertainty** (dark blue triangle icon): all observations and inferences come with uncertainty information which is essential in terms of any downstream inferences or human assessment.
- **Event** (green icon customised for different event sub-types): the events that can be detected by services in this environment, these are specialized into a number of concrete types, based on the capabilities of the services, e.g.: *explosion*, *violence*, *shooting* and *situations*.

The Cogni-Sketch canvas on the right of Figure 1 shows the services that were introduced in the previous section: the Capulet Club CCTV service, the Situation Monitor service, the Capulet Plaza Listener service and the Media Monitor service. Two of these (Capulet Club CCTV and Capulet Plaza Listener) are specifically located in Anglova City based on the sensors they are connected to, and the AI processing is located directly on these sensors, i.e., at the edge of the network. The other services are not located in specific geo-locations but operate in cyberspace, with the AI processing running in appropriate locations within the coalition cloud. The figure shows the 'end state' of the sequence of events outlined in Table 1; this state will be built-up over the next three subsections.



if she don't take me to try out . imma stab her cappulet head

if she don't take me to try out. imma stab her cappulet head

Figure 1: Cogni-Sketch human-agent sensemaking environment.

Commonly, software environments to support sensemaking involve graphical interfaces that allow users to 'connect the dots', affording flexibility in adding increasing context and meaning as the sensemaking process progresses, for example, following the Pirolli & Card model [4]. Commercial tools such as I2[1] and research prototypes such as CISpaces[2] support the process with varying strengths and weaknesses. Generally, there is a tendency to either favour higher-level sensemaking (i.e., schematization and case-building, often via formal representations) or lower-level (i.e., pre-formalisation via 'shoeboxing' and exploratory assembly of evidence). Many analysts fall back on generic mind-mapping or concept-mapping tools for the latter because of their ease of use and lack of formality. However, none of these tools currently allow rapid integration of AI services and their necessary explanatory affordances.

4.0 EXPLAINABLE MULTI-MODAL EVENT DETECTION

Figure 2 shows the result on the Cogni-Sketch canvas of the Capulet Club CCTV service detecting an active shooter situation in front of a nightclub that is frequented by members of Anglova's Capulet civilian population. The explanation for the event contains links to the relevant raw input from the sensor, which in this case is the CCTV video and audio feed (shown bottom-left in the figure, and in enlarged form on the right). This can be seen on the canvas as the SAVR (Selective Audio-Visual Relevance) explanation which takes the form of an attention-highlight video showing the audio and visual relevance of the scene.

SAVR employs *selective relevance* [5], a postprocessing step that can be applied to an explanation such as the example produced by the layerwise relevance propagation (LRP) technique [6]. Here we review an

[1] https://www.ibm.com/uk-en/security/intelligence-analysis/i2

[2] https://cispaces.org

example of the SAVR application of the method (i.e., selective relevance applied to both the audio and video streams of a multimodal model). The goal of the method is to decompose explanations for heterogeneous data into more interpretable components. For instance, for the video, a spatio-temporal input, selective relevance decomposes the explanation into its spatial and temporal counterparts. The temporal element is shown on the far right in Figure 2. This functionality, not present in the base LRP explanation, reveals to us specifically whether regions in the video are relevant because of their appearance or because of the motion taking place within them. Within the context of this example, selective relevance is seen to highlight the shooter's arm (upper right quadrant of the fat top-right image in the figure). Similarly, temporal elements of the audio track are highlighted on the spectrogram in the lower right image.

The chief benefit to coalition operations from SAVR is increased robustness of the human-AI system. This includes robustness against adversarial attacks involving input spoofing because the focus on temporal and multimodal features makes it significantly more challenging for an attacker to provide spoof data: spoof data needs not only to fool a classifier in each modality, but also to generate plausible modal and temporal/spatial distributions of relevance.

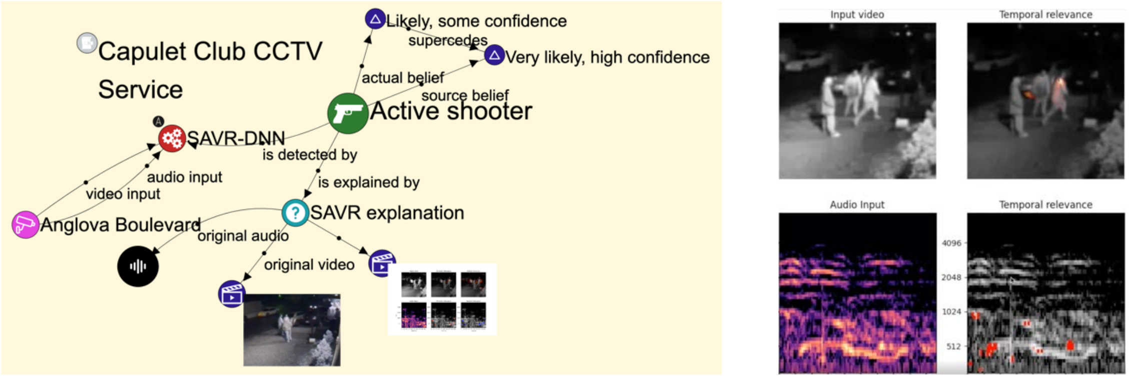
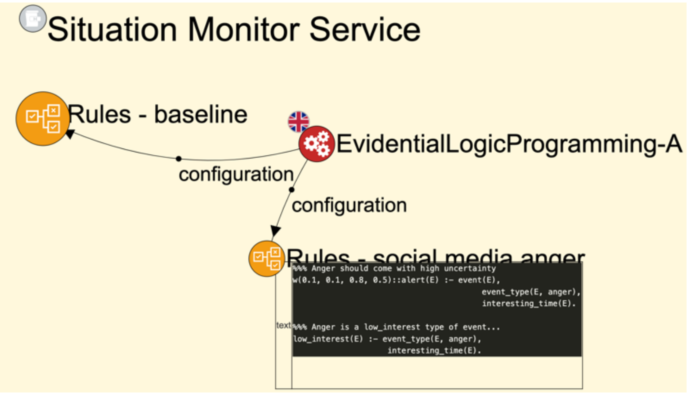


Figure 2: Multi-modal event detection with SAVR explanation.

The *active shooter* event has been assessed to be "Likely, some confidence". This is a natural language representation of a subjective logic opinion [7], that distinguishes the amount of belief, disbelief, and epistemic uncertainty in the truth of a given proposition. Each event comes with a source uncertainty assessment that can be modified upon ingestion into the platform. The source uncertainty, here, has been discounted at 80% because the sensor and service are run by the local Anglovan authorities whereas the coalition user maintaining situation understanding is from the UK. This information is all available in the graph but is hidden by default to ensure that the users are not overwhelmed, but can navigate to any relevant information as needed.

5.0 SITUATION MONITORING WITH KNOWLEDGE INJECTION

The analyst has seen the shooting event and has formed a hypothesis of escalating violence against the Capulet civilian population. There is no time or data to re-train the Situation Monitor service to learn this hypothesis, but we can readily modify the rules used by the Evidential Logic Programming inference engine to inject the new hypothesis in near real-time (an instance of the Cogi-Sketch *tellability* affordance noted in Section 3). The analyst therefore would like to direct the Situation Monitor service to be particularly focused on any discussion on social media that may indicate increasing unrest, specifically by looking for early indications of violence towards Capulets. This is a technical update and would likely be made by an operator who has responsibility for maintaining the services, but the Cogni-Sketch environment can also support the input of higher-level information from which code such as this can be generated. After reviewing the current configuration the user decides to extend the rules and link this into the configuration, as shown in Figure 3. This could either be done by extending the existing configuration rule, or by creating a new additional node as shown here. The rule is now live and has updated the running Situation Monitor service accordingly.


Situation Monitor Service
Rules - baseline
EvidentialLogicProgramming-A
configuration
configuration
Rules - social media anger
text

```
%%% Anger should come with high uncertainty
w(0.1, 0.1, 0.8, 0.5)::alert(E) :- event(E),
                                   event_type(E, anger),
                                   interesting_time(E).

%%% Anger is a low_interest type of event...
low_interest(E) :- event_type(E, anger),
                   interesting_time(E).
```

Figure 3: Injection of a new situational rule to take account of threats of violence on social media.

As time passes, the Media Monitor service, processing social media data on Twitter, signals a credible, “Likely, some confidence,” threat to Capulets, shown in the bottom right of Figure 4. This service employs a DNN model based on BERT embeddings [8] trained with the evidential deep learning (EDL) approach allowing the model to express epistemic uncertainty when facing unseen data [9]. At the neural layer, an EDL model is specially trained to characterize the amount of relevant evidence for the various alternatives in light of the input (sensor) data and the data used to train the AI network. In this case, the input is Twitter text messages and the training data consists of samples of threatening/angry messages previously obtained from Twitter and labelled as such by human subject matter experts. In many different applications, it has been demonstrated that EDL can detect out-of-distribution test samples. Furthermore, accuracy increases when deferring decision-making on highly uncertain test data. The EDL framework can be applied to numerous target classification systems. It allows such systems to alert decision makers when it no longer is able to provide reliable recommendations, which is possibly due to changes in the operational environment relative to how the system was trained. This enables decision makers to rely on the system only when it is reliable.

Returning to the scenario, the ‘triggering’ Twitter message processed by the Media Monitor service is shown at the bottom left of Figure 4, along with an attention-based explanation that highlights (in red) the parts of the tweet that caused it to be classified as threatening. The EDL uncertainty is shown here mapped into in natural language as “Likely, some confidence”. Due to the earlier rule change, the Situation Monitor service is in turn triggered by the detection of this new *threats of violence* event, creating a situation based on both the new event and the previous *active shooter* event, both of which are targeting the Capulet community (see the links between the three green event nodes in Figure 4). Initially the uncertainty is “Likely, with low confidence” with the rationale noting that this is computed from component events, which are indicated via the “related event” links from the *active shooter* and *threats of violence* events. The rules continue to run and the situation will be updated as any new related events are detected, with the certainty being revised accordingly.

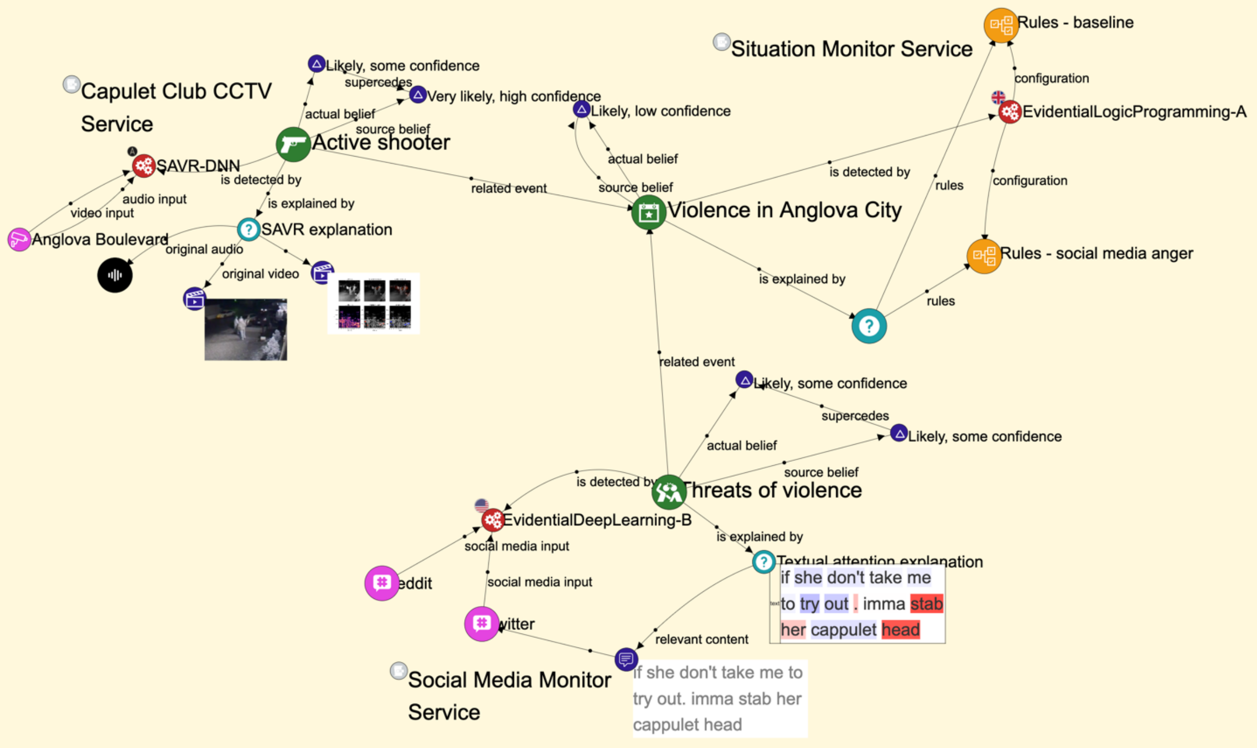


Figure 4: The unfolding situation with linked *shooting* and *threats of violence* events.

6.0 NEURO-SYMBOLIC COMPLEX EVENT DETECTION

Finally, an explosion event is detected by the Capulet Plaza Listener service. As with all detected events, a link to the explanation is shown along with the original audio should the analyst wish to assess the event for themselves – shown in Figure 5. This AI service is very confident, generating a “Very likely, certain” classification for this event, and since this listener service is run by the UK coalition partner there is no modification to this certainty. The new explosion event is also linked to the unfolding situation, due to the spatio-temporal proximity, and the certainty for the situation is updated accordingly.

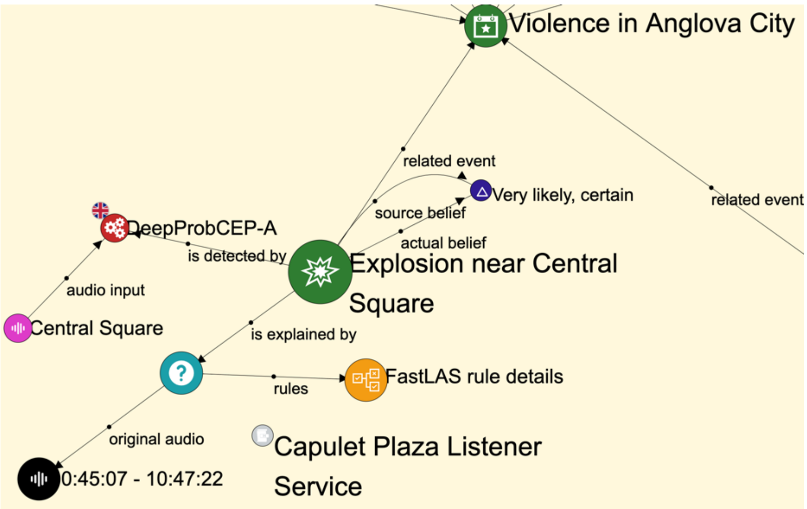
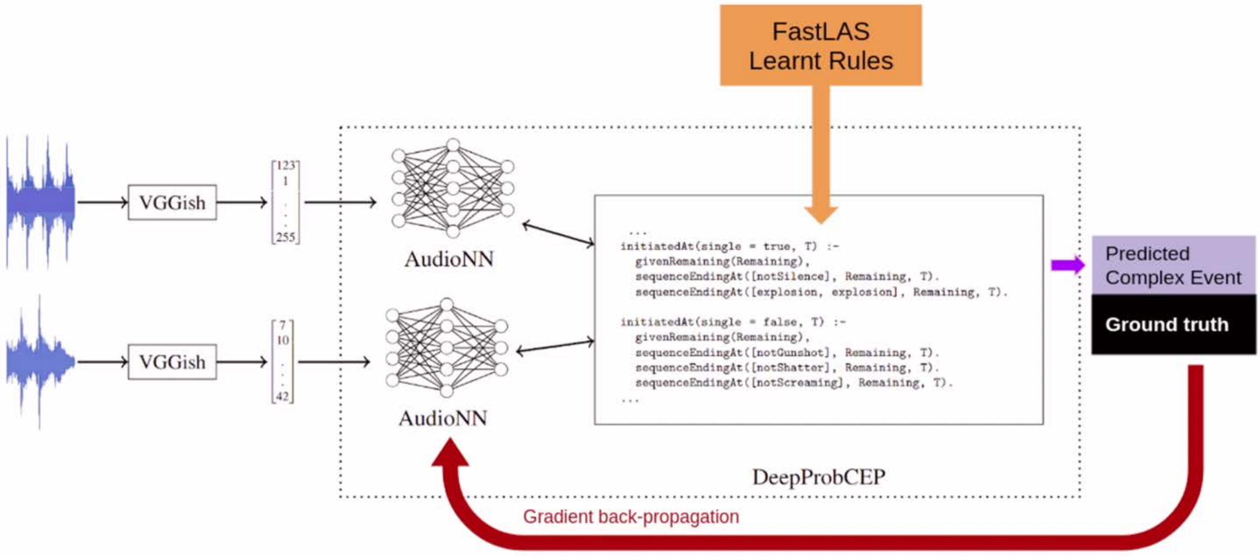


Figure 5: Explosion event detected by the DeepProbCEP Capulet Plaza Listener service.

We will now examine implementation details of the Capulet Plaza Listener service, shown in Figure 6. This service is implemented using DeepProbCEP, a hybrid neuro-symbolic architecture designed to perform complex event processing (CEP) [10]. DeepProbCEP combines a neural network and a logic (rules) layer which allows us to train the system using far less data than neural only approaches. The first step is to split the input audio into one second segments. Each of these segments is then passed through VGGish, a state-of-the-art feature extractor [11]. VGGish outputs a feature vector which is then fed into a neural network, AudioNN in the figure. AudioNN will output a classification of which sounds are present in the given second of audio. Based on these classifications, the logic layer will detect the situations of interest defined in the rules. Compared to simple neural architectures [12] and state-of-the-art neuro-symbolic approaches to CEP [13], DeepProbCEP (i) needs less labelled training data thanks to its end-to-end learning capability, (ii) is robust against noise and adversarial attacks in the form of training data poisoning, and (iii) can classify individual events as a by-product of the end-to-end training.

DeepProbCEP allows experts to manually define the conditions for a situation of interest similar to the rule injection example shown in Section 5. However, in this case we have used FastLAS [14] to learn the rules through inductive logic programming. This has allowed us to train DeepProbCEP in and end-to-end manner without the necessity of an expert to define the rules. In the centre of Figure 6 we can see some of the rules that have been written by FastLAS. Despite being automatically generated, these rules are still understandable by humans with a logic programming background. For instance, the first rule defines that a single IED attack will occur if an explosion sound is detected for at least 2 seconds without being interrupted by complete silence. This is the rule that detected the explosion for this demo. In order to train the system we first obtain the rules using FastLAS. This was done using a very small dataset with different situations of interest we wanted to detect. Then the whole system can be trained in an end-to-end manner. In order to do this, the training audio must be provided to the system which will then use the neural network and rules to predict if a situation of interest is occurring. This will then be compared to the ground truth and used to generate a gradient for back-propagation. Using DeepProblog [15] the logic layer is made differentiable which allows us to train the neural network to detect the sounds that appear in the audio.



```
 ...
initiatedAt(single = true, T) :-
  givenRemaining(Remaining),
  sequenceEndingAt([notSilence], Remaining, T).
  sequenceEndingAt([explosion, explosion], Remaining, T).

initiatedAt(single = false, T) :-
  givenRemaining(Remaining),
  sequenceEndingAt([notGunshot], Remaining, T).
  sequenceEndingAt([notShatter], Remaining, T).
  sequenceEndingAt([notScreaming], Remaining, T).
...
```

Figure 6: Overall architecture of DeepProbCEP for the audio setting used in the Capulet Plaza Listener service.

7.0 CONCLUSION

In this paper we have shown how multiple pieces of research carried out in the US/UK Distributed Analytics and Information Science (DAIS) programme enable patterns of situational events to be recognized where there is insufficient time or data to train deep learning AI models. We highlighted two key capabilities:

- Humans can inject new knowledge about patterns of activity through addition of new rules.
- AI models can be trained more rapidly, with a much-reduced need for training data and improved detection accuracy compared to ‘pure’ deep learning approaches.

The scenario highlights a loose coupling of hybrid types of AI systems in an open architecture, and how the resulting applications are able to run on edge-of-network devices, thus being suitable for use in tactical sensor systems. DeepProbCEP operates on commercial off-the-shelf hardware (a standard laptop) and SAVR has been tested on a Jetson Nano device, and is capable of generating audio-visual explanations in real-time. Moreover, we demonstrated AI services operating in multiple domains – physical and cyber – on a variety of data modalities, with awareness of situational uncertainty. The overall goal of our research is to offer a means to enable Defence to adopt AI systems able to learn and adapt at the ‘pace of the fight’, to ensure our allied coalition’s tempo of understanding and action is not overmatched by adversaries.

ACKNOWLEDGEMENT

This research was sponsored by the U.S. Army Research Laboratory and the UK Ministry of Defence under Agreement Number W911NF–16–3–0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the UK Ministry of Defence or the UK Government. The U.S. and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

REFERENCES

[1] N. Suri, K. M. Marcus, C. van den Broek, H. Bastiaansen, P. Lubkowski, and M. Hauge, “Extending the Anglova scenario for urban operations,” in *International Conference on Military Communications and Information Systems*, 2019.

[2] F. Cerutti, L. Kaplan, A. Kimmig, and M. Sensoy, “Probabilistic Logic Programming with Beta-Distributed Random Variables”, in *33rd AAAI Conference on Artificial Intelligence*, 2019.

[3] D. Braines, F. Cerutti, M. R. Vilamala, M. Srivastava, L. Kaplan, A. Preece, and G. Pearson, “Towards human-agent knowledge fusion (HAKF) in support of distributed coalition teams,” in *AAAI FSS-20: Artificial Intelligence in Government and Public Sector*, 2020.

[4] P. Pirolli and S. Card, “The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis,” in *International Conference on Intelligence Analysis*, 2005.

[5] H. Taylor, L. Hiley, J. Furby, A. Preece, and D. Braines, “VADR: Discriminative multimodal explanations for situational understanding,” in *23rd International Conference on Information Fusion*, 2020.

[6] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, “On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation,” *PloS One* 10, 2015.

[7] A. Jøsang, A, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*, Springer, 2016.

[8] J. Devlin, M-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding”, *NAACL-HLT*, 2019.

[9] M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” in *Advances in Neural Information Processing Systems*, 2018.

[10] M. R. Vilamala, H. Taylor, T. Xing, L. Garcia, M. Srivastava, L. Kaplan, A. Preece, A. Kimming, and F. Cerutti, “A hybrid neuro-symbolic approach for complex event processing in noisy and adversarial set- tings,” in *36th International Conference on Logic Programming*, 2020.

[11] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, et al., “CNN architectures for large-scale audio classification,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017.

[12] J. Roldán, J. Boubeta-Puig, J. Luis Martínez, and G. Ortiz, “Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks,” *Expert Systems with Applications* 149, 2020.

[13] T. Xing, L. Garcia, M. R. Vilamala, F. Cerutti, L. Kaplan, A. Preece, and M. Srivastava, “Neuroplex: Learning to detect complex events in sensor networks through knowledge injection,” in *18th Conference on Embedded Networked Sensor Systems*, 2020.

[14] M. Law, A. Russo, E. Bertino, K. Broda, and J. Lobo, “FastLAS: Scalable inductive logic programming incorporating domain-specific optimisation criteria,” in *34th AAAI Conference on Artificial Intelligence*, 2020.

[15] R R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. D. Raedt, “DeepProbLog: Neural probabilistic logic programming,” in *Advances in Neural Information Processing Systems*, 2018.