The Alan Turing Institute

Call for reviews of literature

The Alan Turing Institute is seeking to commission literature reviews to inform future research directions.

Each review should be up to 5,000 words in length and should summarise relevant academic and policy literature, identify knowledge gaps and highlight opportunities for future work. They should be written for an interested non-specialist audience and should set out the key questions in the field, the empirical challenges to answering these questions, the approaches existing work has taken, and the results found. The reviews should be international in scope.

To apply, please send an email to Mark Briers (london-arc@turing.ac.uk) briefly describing:

- Which review(s) you are interested in;
- The researchers who would be involved and their suitability for conducting the review;
- The focus and scope your review would take and the literatures you would draw on;
- Permission from your academic lead (where relevant).

Selection will be based on technical suitability. Payment will be negotiated based on experience and suitability. Reviews would be expected to commence as soon as practically possible.

The deadline for submitting interest is 15th December 2019.

Any questions should be addressed to Mark at the above email address.

The Alan Turing Institute
The British Library
96 Euston Road
London NW1 2DB

+44(0)30 0770 1912
info@turing.ac.uk
turing.ac.uk

1) Audio at the Edge

There is a requirement to understand the limitations of AI at the edge in terms of Size, Weight, and Power combined with model accuracy. We require an understanding of the tradeoff between model size and performance at the edge, and to demonstrate through tangible examples the benefit of moving data enrichment to the edge of the network. There is a specific interest in the deployment of speech to text and speech recognition models on edge devices.
Over the last couple of years there has been significant development in the application of Deep Learning pipelines to audio processing with particularly relevant work from the Google AI team ("Streaming End to End Speech Recognition for Mobile Devices" using an RNN Transducer network), and Nvidia's GPU acceleration of the Open source Kaldi library (https://devblogs.nvidia.com/nvidia-accelerates-speech-text-transcription-3500x-kaldi/).

2) Arabic OCR

There is a requirement to understand and demonstrate state-of-the-art methods in academia for OCR applied against both printed and handwritten Arabic script (Arabic is a priority but there is also interest in other non-Latin-based texts such as Cyrillic). We are specifically interested in the use of innovative methods for augmenting training data in order to improve the performance of classifiers. This may include the use of adversarial/generative approaches in order to synthesis training data, building on approaches tested and demonstrated at previous ICDAR conferences.

3) Model Security and Inversion

Increasingly the UK Government are looking to deploy machine learning based models onto internet based. Often these models are trained on sensitive data which has raised questions with regard to UK policy on the classification of such models. Given the rise in the number of adversarial attacks being demonstrated against trained models, especially model inversion attacks, research is required in order to inform the handling of trained models using an evidence-based approach.

Use cases for the types of models of interest include: object detection, face recognition, speaker ID, Speech to text, sentiment analysis, document classification, and entity extraction. There is a requirement to:

- Track state of the art research in model inversion approaches;
- Exploration of methods which are available and in-development to protect ourselves against such adversarial attacks.

4) Real-world model poisoning

Researchers have demonstrated the ability to undertake targeted backdoor attacks against Face Recognition algorithms, which cause the system to behave erratically under real-world conditions. This falls under the category of adversarial attack referred to as Model Poisoning. Model Poisoning is of particular concern to the UK Government as they are extremely difficult to detect once training has been completed, and many of the underpinning models (e.g. ImageNet, BERT) are open source, with little knowledge of the training process and data.

5) Approaches to Low-shot learning

The UK Government would like to embrace the potential capability enhancements which Deep Learning based systems offer across a number of different data modalities. These include but are not limited to: image classification, object recognition, activity detection in video, text classification, entity extraction, sentiment analysis, face recognition, NLP, and audio processing. However, most commercial classifiers do not transfer well into the Government domains due to slight differences in the data to which they are applied, or the nature of the classifiers. There is a requirement to train systems within Government on small volumes of data, however in most cases there is not enough training data for deep learning systems to give reasonable performance. Low-shot learning may offer a solution to this challenge.

7) Online Deception

Across government there are significant concerns regarding criminal use of generative capabilities in support of face and voice synthesis. Wider applications of such capabilities will significantly degrade and undermine trust in online information and could lead to widespread deception of individuals online. This is a rapidly emerging field which requires agile research in order to stay on top of the threat, and to help develop potential counter-measures and detection capabilities.

8) Masquerading

Use of video-conferencing online is becoming ubiquitous in the modern age. With the increased maturity of deep fake capabilities, how can we assure individuals of the identity on the other end of the video-conference? How long before GAN and similar based technologies reach a point where real-time masquerading is a mature threat? How is the technology developing? What are the risks now and in the near future? What methods are effective for detecting such deception?

Prior research of relevance to this requirement include: pix2pix, vid2vid, Face-It, and Lyrebird.

9) Automated Image/Video Forensics

There is a requirement to develop automated approaches for detection of fake and/or manipulated imagery and video. The use of imagery and video manipulation is widespread for both benign and malicious purposes and takes many forms. These include:

- Imagery alteration using software tools such as Photoshop
- Image splicing
- Image synthesis using GANs
- Video splicing (e.g. Deep-fakes)

There is a need to understand what capabilities already exist, benchmark these capabilities against a collated dataset, understand the limitations of detection capabilities, and develop

new innovations to address these shortfalls. Desired outputs include: literature reviews, demonstrators, code, and research papers.

Researchers should take the following considerations as part of the research:

- There is a greater interest in generic capabilities which are able to detect more than one type of manipulation/synthesis. Use of classifiers for specific approaches (e.g. Style-GAN) whilst interesting are of less operational utility;
- Researchers should assume that no prior knowledge of the approach is known (e.g. what model is being used);
- The use-case is to find a manipulated image or video in a large volume of un-manipulated data, therefore false positives are a concern;
- Modern day cameras often introduce some post processing at the sensor end (e.g. contrast stretching) which can manifest as a manipulation feature for automated detectors - we would like to minimise and account for these type of distractors;
- The capabilities should be completely automated and should minimise the need for a human forensics expert.

10) ML/Transformations on low side encrypted data

Within UK Government there is an increased push to conduct data processing and analysis on internet connected systems, to reduce storage and processing costs, and facilitate access to state-of-the-art developments in the commercial world. However, adoption of connected systems is often hindered by the classification of datasets processed within government. Making use of developments in Privacy Enhancing Technology (PET), users are able to store and query data on such systems in a secure manner. We need to understand whether this is possible, and what practical technologies exist which enable UK Government to conduct more data analysis on encrypted data held on internet connected systems.

The Alan Turing Institute
The British Library
96 Euston Road
London NW1 2DB

+44(0)30 0770 1912
info@turing.ac.uk
turing.ac.uk