**Opportunity Title:** Threat Model Considerations in Systems that Use Neural Networks at the Edge
**Opportunity Reference Code:** ICPD-2022-37

**Organization** Office of the Director of National Intelligence (ODNI)

**Reference Code** ICPD-2022-37

**How to Apply**

**Create and release your Profile on Zintellect** – Postdoctoral applicants must create an account and complete a profile in the on-line application system. **Please note: your resume/CV may not exceed 2 pages.**

**Complete your application** – Enter the rest of the information required for the IC Postdoc Program Research Opportunity. The application itself contains detailed instructions for each one of these components: availability, citizenship, transcripts, dissertation abstract, publication and presentation plan, and information about your Research Advisor co-applicant.

Additional information about the IC Postdoctoral Research Fellowship Program is available on the program website located at: https://orise.orau.gov/icpostdoc/index.html.

If you have questions, send an email to ICPostdoc@orau.org. Please include the reference code for this opportunity in your email.

**Application Deadline** 2/28/2022 6:00:00 PM Eastern Time Zone

**Description**

**Research Topic Description, including Problem Statement:**

With machine learning advances in hardware and software being powered by strong commercial interest and consumer awareness of data security/privacy it is becoming more practical to deploy trained models to the edge of a system on low power devices to filter information, make decisions and provide context to data before it is aggregated on back-end systems

Neural networks are benefiting from advances in core principles, software libraries and optimized hardware to run them making an eco-system that is moving quickly but still relatively immature in terms of its success deploying solutions to “production”.

What threats and opportunities does using neural networks at the edge of a system present? I.e., how could somebody affect a system that has a neural network as a component, either to defeat, deceive or coerce the system?

Other thoughts for steer are:

- What new attack vectors are there in a system that uses a neural network on an edge device?
- We consider “edge” devices to be low power sensors to small servers with a preference for the sensor end, potentially with specific hardware optimized for running neural networks (TPUs, GPUs, FPGAs, Neuromorphic chips).
- Consider what information may change an adversary’s approach, e.g., if you knew the hardware, firmware, software, model deployed, training dataset, validation dataset etc.
- Consider the cost of each attack option vs the reward/impact (disruption, disable, data breach etc.)

**Example Approaches:**

One approach could be to consider the neural network as an isolated “black box” component and affect the inputs in a way that is imperceptible to human eyes but can change the output in some way, there is previous work in this field of deceiving neural networks and in some ways the approach can be considered similar to a GAN based approach trying to create one network that can fool another.

Related to the previous approach would be the “white box” approach where the internals of the neural network are known and the attacks focuses on replacing/changing the network to alter its outputs while maintaining enough of its behavior to limit detection.

**Opportunity Title:** Threat Model Considerations in Systems that Use Neural Networks at the Edge
**Opportunity Reference Code:** ICPD-2022-37

Another approach would be to consider specialized hardware that may be used to run inference on and how that could be interfered with to affect the performance of the system. CPU security has received some attention and improvement with issues like Spectre and Meltdown receiving media attention. Attacks focused on information leaking during transfer (see: https://ieeexplore.ieee.org/document/8715004) in CPUs are common but are TPUs different? How about neural networks deployed to FPGAs? Will neuromorphic hardware be free from those issues given it doesn't separate memory and storage or will it have a new set of unique issues?

When systems contain sensors without processing at the edge security often focuses on validating the input signals, establishing trust in the hardware/software and looking for "normal" data being received based on historic data. Are there opportunities to build consensus between sensors on events when they can all apply context and describe the event e.g., a person is there because I can see them and hear them? – Can this be used to increase trust in the individual devices and spot disruption?

**Relevance to the Intelligence Community:**

The IC needs to understand the threats and opportunities when creating distributed sensor networks that have intelligence and business logic spread across the system. Specifically, the security considerations of using neural networks as components but more broadly their use on the system in terms of trust and threats. For example a CCTV network or a building management system.

**Key Words:** Cybersecurity, Neural Networks, Edge, IoT, Machine Learning

## Qualifications

**Postdoc Eligibility**

- U.S. citizens only
- Ph.D. in a relevant field must be completed before beginning the appointment and within five years of the application deadline
- Proposal must be associated with an accredited U.S. university, college, or U.S. government laboratory
- Eligible candidates may only receive one award from the IC Postdoctoral Research Fellowship Program

**Research Advisor Eligibility**

- Must be an employee of an accredited U.S. university, college or U.S. government laboratory
- Are not required to be U.S. citizens

## Eligibility Requirements

- **Citizenship:** U.S. Citizen Only
- **Degree:** Doctoral Degree.
- **Discipline(s):**
  - **Communications and Graphics Design** (2 👁)
  - **Computer, Information, and Data Sciences** (16 👁)
  - **Earth and Geosciences** (21 👁)
  - **Engineering** (27 👁)
  - **Environmental and Marine Sciences** (14 👁)
  - **Life Health and Medical Sciences** (46 👁)
  - **Mathematics and Statistics** (10 👁)
  - **Other Non-S&E** (2 👁)
  - **Other Physical Sciences** (12 👁)
  - **Other S&E-Related** (1 👁)
  - **Physics** (16 👁)
  - **Social and Behavioral Sciences** (27 👁)