PCLOB 7/11 AI Forum

Dean Souleles Opening Remarks

Framing question: "What are the best ways for the IC to use AI in its counterterrorism efforts? What are the risks of using AI in this context and how can we best reduce/minimize those risks?"

AI and its constituent technologies are very important tools for the US IC to apply to all its mission areas and nowhere is it as critical as in the counterterrorism arena. But it is essential that AI be applied in a legal and ethical way that consistent with our shared values.

I will talk a little bit about the ways is which AI is important to intelligence and counterterrorism in particular but before we do that it is important define a few terms. This may seem obvious but it's important to be clear about what we mean. In its broadest sense the IC's counterterrorism consists of the following two broad responsibilities:

- To collect, analyze and share actionable intelligence
- To detect and disrupt threats

In addition, NCTC, the National Counterterrorism Center at the Office of the Director of National intelligence is responsible for maintaining the authoritative database of known and suspected terrorists.

So that is the IC and the CT mission. Now what do we mean by AI?

That's not quite as easy to answer.

In the current media environment, one could be excused for thinking that AI is synonymous with Large Language Models and Chatbots. For those who weren't deeply involved in technology ChatGPT appeared to come out of nowhere just about two years ago and now is seemingly everywhere. But LLM's are just the latest in a long line of machine intelligence tools that have become increasingly more useful over the last decade.

In 2012 Yann LeCun and Geoff Hinton demonstrated that neural network based supervised machine learning image recognition systems could perform better than competing systems and sometimes better than humans. That demonstration triggered a tidal wave of innovation that continues to this day. Its applications have proven to be broad and wide –machine translation, speech recognition, image recognition in every field ranging from radiology to jet engine design.

A simplified way to thing about AI is to divide into a couple of sub-classes of technology:

- Supervised machine learning is characterized by requiring large amounts of "ground truth" training data – usually provided and curated by human experts. In that way machine classification systems.

- Unsupervised machine learning is used to automatically represent data in useful ways that simplify or compress it or help to find hidden patterns and connections in the data.

- Reinforcement learning is yet another set of AI technologies where the system learns how to behave in a way that increases reward by interacting with the environment.

- Deep learning is a set of technologies that work across all three of the areas I just described. Deep learning uses large quantities of pf data to figure out how to do complex things by searching for the combination of weights that best describes the data.

There are many nuances and other specific techniques and tools but those broad categories account for 90% of what we call AI today.

Turning back to the IC.

It's difficult to believe now but in 2017 when we began to develop AIM – Augmenting Intelligence Using Machines for the IC no national level federal agency had published an AI strategy. There were no executive orders, no one in the Office of Science and Technology Policy was talking about AI. The PCLOB certainly wasn't. And yet the commercial world was on a breakneck pace with new innovations happening almost weekly. At that time very few people in government or in the public even understood what the topic was. AI was the magic box of the tech giants – Google, Amazon, Microsoft and Facebook.

So we set about to create an AI strategy for the IC. We focused on four big ideas:

- Creating a secure digital foundation and focus on data
- Invest in the gaps where the commercial sector was not doing enough
    - Different risk modes mean the need for different security models
    - AI security and vulnerability are paramount in ways they may not be in other enterprises
- And in the long term we need better sense making and understanding

Our initial use cases focused on automation of routinized tasks. Answering questions that start with "who, what, where and when" but not "how" or "why" or "what if".

It turns out that AI is good at some things –

- Pixel peeping and counting – cars in parking lots and airplanes on runways, recognizing faces or matching fingerprints, translating text, even summarizing large quantities of text. ChatGPT turns out to be very good at that.

And bad at other problems and highly dependent on quality training data

- Which specific airplane is that – which needle in a pile of needles?
- Why are there fewer airplanes today than yesterday?
- What is going to happen next?

These are the questions for the human analyst – that no AI system is currently capable of doing reliably.

Today nearly every way the IC uses computers can be reduced to a question of search

- Show me a this, find me a that, what does this say, how does this relate to that, what is the network of relationships

In the counterterrorism domain one can readily see the use cases:

- Keeping track of the known terrorists – who are trying to obfuscate their personas
    - Facial recognition and other biometrics are becoming increasingly useful
- Network analysis to find patterns in communications collections to discover new individuals and groups of individuals who may be working together
- Combining open-source information with classified collections to provide indications and warning of potential terrorism activities.

All these mission areas benefit from appropriate uses of AI, but we understood then and now that if we are to apply these tools to national security we must better characterize and understand the risks.

- We know that **AI can learn the wrong thing** if it is not given appropriate training data or if the training data is somehow poisoned or deliberately biased.
- We know that **AI can do the wrong thing** and worse it can do it with confidence. The makers of the chatbots are so anxious for us to trust them that they have anthropomorphized them in an effort to make them seem more credible – and yet we know they hallucinate.
- Of particular interest to the IC is the fact that **AI can reveal the wrong thing** – it's been shown that to given access to a machine learning model it is possible to reverse engineer the model to learn about the data the model was trained with just potentially jeopardizing our sources and methods.

And like all other applications of technology, it is important that we hold these technologies to the same high standards that we hold our people to. We must insist that AI generated results be accurate, fair, defensible and transparent. That is why at the same time we developed AIM we developed the "Principles of AI Ethics for the Intelligence Community and the "Artificial Intelligence Ethics Framework for the Intelligence Community" – hear are just a few of things the ethics framework tells about AI.

- Be used when it is an appropriate means to achieve a defined purpose after evaluating the potential risks;
- Be used in a manner consistent with respect for individual rights and liberties of affected individuals, and use data obtained lawfully and consistent with legal obligations and policy requirements;
- Incorporate human judgment and accountability at appropriate stages to address risks across the lifecycle of the AI and inform decisions appropriately;
- Identify, account for, and mitigate potential undesired bias, to the greatest extent practicable without undermining its efficacy and utility;
- Be tested at a level commensurate with foreseeable risks associated with the use of the AI;
- Maintain accountability for iterations, versions, and changes made to the model;
- Document and communicate the purpose, limitation(s), and design outcomes;

- Use explainable and understandable methods, to the extent practicable, so that users, overseers, and the public, as appropriate, understand how and why the AI generated its outputs;
- Be periodically reviewed to ensure the AI continues to further its purpose and identify issues for resolution; and,
- Identify who will be accountable for the AI and its effects at each stage and across its lifecycle, including responsibility for maintaining records created.

In summary – AI represents tremendous opportunities to improve the counterintelligence mission to enable the IC to deliver better and more timely intelligence to national leaders and policymakers. But it can only do so if the underlying systems are safe, reliable, secure and implemented in an ethical manner consistent with our national values.