Written Testimony

of

Devaki Raj

Former Chief Executive Officer & Co-Founder, CrowdAI

Before the

U.S. Senate Committee on Homeland Security & Government Affairs

Introduction

Chairman Peters, Ranking Member Paul, and distinguished members of the committee, my name is Devaki Raj. I am here representing CrowdAI, a Silicon Valley based start-up leading the development of no-code artificial intelligence tools since 2016. Until a recent acquisition by Saab Inc., I was CrowdAI's CEO and co-founder.

Thank you for this opportunity to testify on "Governing AI Through Acquisition and Procurement" before the Committee on Homeland Security and Government Affairs. I would like to thank Chairman Peters for his leadership on AI initiatives and Ranking Member Paul for his leadership on improving the Small Business Innovation Research (SBIR) / Small Business Technology Transfer (STTR) program for small businesses and startups like mine.

I am a proud American—born in Ohio and raised in Massachusetts, Connecticut and California. So, it is an honor to be here and to present my testimony from the perspective of a small business and startup, the lifeblood of American ingenuity. Before getting started, I want to thank my team for their tireless efforts, as well as my family for their relentless support.

CrowdAI proudly serves the U.S. federal government across multiple mission areas, notably for disaster response and fighting wildfires in California, as well as countering narcotics trafficking in South America. The government recognizes the clear need and value of AI to these critical homeland security missions.

There have been multiple notable efforts towards using AI, especially during the early days of the AI pathfinders, such as Project Maven and the Joint AI Center (JAIC), where "failfast and adapt" was a shared rallying cry.

Indeed, we live in remarkable times with novel challenges: we are witnessing the dawn of generative AI.[1] The pace of technological advancement is such that everyday Americans cannot keep up[2], let alone consider regulations[3] and procurement practices[4]. But, change is needed. To remain stationary at this moment will only result in the U.S. being left behind by our allies and our competitors.

Should members of the committee take away anything from my testimony today, it is that AI must be thought of as a journey, not a destination. In this sense, I prefer the term "machine learning" to AI, as it better aligns with how we understand human development and how we should be thinking about this technology. It is in this difference between a discrete concept, AI, and a continuous process, learning, that today's procurement generally breaks down.

Today, I will share four main observations about AI procurement:

1. First, commercial off-the-shelf AI solutions need government curated data to be mission ready.
2. Second, AI procurement needs to include ongoing AI model training and the infrastructure to support that training.
3. Third, the rapid growth in open-source AI technologies necessitates rigorous testing and evaluation before and after procurement.
4. Finally, it is important to establish paths to programs of record for small businesses through project transition milestones.

To be clear, my testimony is from the perspective of an AI practitioner, based on the work CrowdAI has done with the U.S. government. But AI research is evolving, and rapidly so. The pace of this change is such that even I feel it moving past me at times. Just like AI, I too must continuously learn; which stands to highlight the importance of this and other hearings on artificial intelligence being held this week. I appreciate the invitation to speak and thank you for your interest.

[1] Gmyrek, P., Berg, J., Bescond, D. 2023. Generative AI and Jobs: A global analysis of potential effects on job quantity and quality, ILO Working Paper 96 (Geneva, ILO). https://doi.org/10.54394/FHEM8239

[2] https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/

[3] https://www.bloomberg.com/news/articles/2023-03-17/chatgpt-leaves-governments-scrambling-for-ai-regulations

[4] https://www.linkedin.com/posts/michaeljkanaan_pentagons-bridge-to-techs-private-sector-activity-6966421664184025088-P2LS

Commercial Off-the-Shelf AI Solutions need Government Curated Data to be Mission Ready

Often, today, we see solicitations ask explicitly or implicitly for ready-to-deploy automated solutions; however, government missions are inherently unique–their sensors, needs, and environments. Commercial off-the-shelf capabilities require government-furnished information to train to those unique missions, regardless if it is for homeland security, intelligence, or public health. Most commercially available AI systems must learn *a priori* about their operational use or they break down. We call this brittleness.

Now, this is a somewhat nuanced point, and some people may bristle at the notion, as certain government programs and companies have built themselves upon the fallacy of "ready-to-use." To be clear, the tools and architectures to create, modify, and operate AI do exist commercially, but the models are built on publicly and commercially available data and marketed aggressively with a "one-size-fits-all" mindset.

The reason for these assertions is that the knowledge and training data needed for many government missions is not in the public domain.

AI needs exposure not only to the data and domain it is to learn from, but that data must match the intended operations environment. In some cases of algorithm development, we can estimate or otherwise get closer to mission data; but, domain shifts from the training data to operations data, will still make models unreliable or brittle. In short, mission specific data are required for each algorithm because each problem posed by homeland security, intelligence, or public health is inherently unique. Accordingly, AI must faithfully represent the situations not only specified by the government, but validated by it too.

Earlier this year, for example, CrowdAI released a toolset that automated remote monitoring of military airfields using "few shot learning" techniques. The tool scales globally, meaning every air force worldwide, while using only a tiny fraction of the training data, in some cases as few as a single image or even a line drawing. This revolutionary advancement in computer vision still had to be vetted by our government partner because we sourced information on aircraft specifications from the internet. We are experts in AI, not in the complexities of Russian, North Korean, and Chinese militaries.

The problem of training robust models to government missions gains complexity as we move from sensor to sensor. The domain shift within a satellite constellation can cause brittleness. Even more so is the shift from commercial sensors to sensitive ones, such as unmanned systems, like aerial drones, or overhead collection platforms operated by

the National Reconnaissance Office. There is no way for industry to know, in advance, what those systems and their data look like unless they have been or are on contract.

When we start delving into other areas of AI, such as large language models and generative AI, which draw their education from a massive corpus of publicly available information, but scant government information, similar domain shift issues arise as well. The greater the domain shift, the more training that is needed, transforming the initial model to its government purpose and no longer “commercial off the shelf”.

To accelerate this process, departments and agencies pursuing AI modernization should compile and then furnish datasets so that during contract execution it enables faster transition from commercial off-the-shelf to government off-the-shelf. The National Geospatial-Intelligence Agency, to its credit, for years has been working to curate and redefine how its structures GEOINT data to be machine readable. A herculean task for sure, but a necessary one. Similar efforts, across the government, could be started, if not already done, to identify which missions across the enterprise can be machine-augmented. Force ranked by priority, start curating mission-specific datasets to allow faster and more accurate validation of commercial models and accelerated model tuning. This will help to promote capabilities from Research & Development and transition them to Operations & Maintenance. It is important that I acknowledge that there are constraints that come with use of government data (statutory limitations, privacy, security, etc). But these, for us, have been largely overcome through strong relationships and the “mission-first” attitude we share with our government and industry partners.

However, once this government information is provided, AI isn’t a “set it and forget it” solution.

Ongoing AI Retraining: A Continuous Learning Process

As mentioned earlier, AI is a journey. AI procurement needs to include ongoing AI model retraining and a retraining infrastructure.

Currently, procurement processes often buy AI as a one-off software solution, like a copy of Microsoft Word for your desktop. However, due to the nature of machine learning, it is critical to procure AI technologies with the ability to continuously incorporate new data as sensors and missions change, as they invariably do.

Current AI software that is commercially-produced, needs to integrate with current government systems, and continually update on new data to account for new locations, sensors, ecosystems, and missions.

For example, in 2018, CrowdAI collaborated with the California Air National Guard to automate wildfire mapping using MQ-9 drones, a collaboration we proudly continue.

Our initial predictive models performed extremely well in northern California's forested regions where wildfires were common and we had access to government furnished data.

Unbeknownst to us at the time, when the Air Force drone arrived on station, a different sensor was flown for overwatch. Furthermore, the wildfire's epicenter was in a suburban area. As evolving wildfire epicenters shifted and sensors were updated, our models required retraining to maintain operational relevance. These changes represented a twofold domain shift from the previous year's model, which needed to be retrained. Anticipating these operational inevitabilities, we built model retraining into our contract to avoid unnecessary contract options or extensions that, during disaster missions, could cause fatal delays.

While AI models are flexible, they still require contracting officers to include model retraining in contracts, ensuring alignment with evolving mission data. It's a dynamic process, akin to software updates, not just a one-time procurement.

For that AI development project, we increased wildfire map production from just one per day to updates pushed electronically to firefighters every 30 minutes. This was a momentous improvement for our first responders. And our goal in the next six months is to achieve real-time situational awareness, fully automated, from the overhead sensor to users' mobile devices. With our partners at the 163rd Operations Support Squadron, the Hap Arnold Innovation Center, and California Department of Forestry and Fire Protection, this will be reality before next fire season, further saving lives and providing an exemplar for others to follow.

Model training and retraining is not only inevitable, it is the rule. As the world constantly evolves, it is important for a model to continue learning and not be bound by its original training. Contracts for AI must include continuous learning. This ensures that the government is not using a stagnant model that declines in time, but one that continuously improves and adapts as it is exposed to more data.

For the purposes of operationally-ready AI, there is no functional difference between development and operations, as they are continuously interlinked. Therefore, we must move past AI being funded as one off software purchase, but build procurement vehicles that bake in ongoing updates or service level agreements. This is not only because AI needs re-training but also to provide the procuring officer with technology that is cutting edge. It is important that commercial tools, available today, can be implemented at the enterprise level to operate AI at the pace of mission as both the AI and sensor technology continues to evolve.

Rigorous Testing and Evaluation of AI Solutions

Today, anyone can go online and download a trained computer vision model. For example, it is easy to find detection models using state-of-the-art architecture trained on millions of images collected all over the Internet. For many use cases, it is a great tool for experimentation and use.

Cyber threats aside, the risk of using publically available AI architectures creates room for moral hazard. Companies today are incentivized (if not required!) to claim AI expertise, even when it is limited. There often is little-to-no ability by a procurement officer to verify vendor claims of the robustness of their models: Are these models purpose-built or open sourced? How will it perform on my data? These questions are near impossible to answer *a priori*.

Source selection panels for contracting have the difficult task of interpreting changing rules about AI and inferring proposing companies' credentials and ability to deliver. Because AI moves so fast, industry has been particularly susceptible to research that only exists in white papers and technical proposals. Companies recognize that to be competitive they must sell the art of the possible versus the science of today. The result can cut either of two ways: the company delivers on its promises or it does not.

On account of how procurement works today, companies often project future potential capabilities, regardless if they are possible, to win federal awards. The risk of development failure comes later and falls on the government. The only risk to these companies is failure to secure contracts. This phenomenon can be mitigated through more rigorous market analysis and, more importantly, sound model testing and evaluation—during solicitation of capabilities, such as with AI competitions that grant favorable terms to participants for commercialization, and after delivery on an ongoing basis.

Codifying government-wide standards for AI testing and evaluation would help mitigate unverifiable corporate claims. As we move from mission to mission, department to agency, everyone's thinking today about AI is different. Those differences not only leave room for risk, but drive confusion both in industry and government.

Both quantitative and qualitative evaluations of AI are important: quantitative evaluations give information on how well the model is doing across large amounts of data, but can sometimes obscure important information by only reporting aggregate metrics. We've found that it is equally critical to evaluate data samples to get a sense of how and why a model is performing, and more importantly, qualitatively if the model is accomplishing the task assigned to it.

An example CrowdAI encountered was finding airstrips used by narco-traffickers. We built a tool that found areas meeting our description: long, straight dirt paths cleared of vegetation. However, in addition to finding airstrips in the jungle, it often also found dirt roads, which were prolific throughout the region. The point being that evaluating AI isn't as simple as setting an arbitrary performance score. Yes, those metrics matter, but they represent only part of the story. Performance must also consider if the capability solved the problem, in this case finding runways.

This is why I recommend both quantitative and qualitative metrics. Knowing these upfront can alleviate confusion later about why or how a system performed in an operational environment. For this to happen, we must educate our federal workforce on AI fundamentals and, then, provide specific training on how it relates to their role. I commend Chairman Peters for leading AI legislation to address some of these challenges through the AI Training for the Acquisition Workforce Act.

Establishing Paths to Programs of Record for Small Businesses through Project Transition Milestones

With no-code tooling, such as ours, and the advent of generative AI, it will not be long before we see AI being developed in every government agency, business, university, and home across the country, whether people know it or not. Like when the appstore for the iPhone opened itself to 3rd party creators, I anticipate an explosion of AI tools and models beyond what we can imagine today. So, rather than force innovative solutions through one or a few large prime contractors or government innovation units, we should normalize standards and federate acquisition to benefit all businesses.

Startups, at first glance, may appear agile and resilient; but, on government timelines, they are fragile and exposed. The "procurement Valley of Death" continues to take its toll not only on small businesses, but the government's efforts to source and deploy AI, as well. Even after selection, with funding obligated and sole-source authorization under SBIR, we have seen contracts take the better part of a year to award—a lifetime for a small business. So, we look for other means.

One of the fastest ways to get on contract is through a subcontract with a systems integrator. But, working through a prime contractor is a double-edged sword. Startups, so eager for work, can reveal their innovations and thinking, giving primes new ideas. It is a risk we take in order to get the work—to live to fight another day.

The preferred alternative is direct award from the government. The Small Business Innovation Research program is one such vehicle, and I want to acknowledge the members of this committee for SBIR's reauthorization last year. Thank you for continuing this valuable program.

It is difficult to overstate the transformative capabilities the SBIR program has for small businesses like mine. SBIR's sole-source award policy allows small businesses to compete with large prime contractors. This rule helps level the playing field, increasing the types of new technologies that the government can procure, as well as the speed with which they can test and evaluate them.

SBIR's phased approach spans requirement validation, initial development, and transition, which provides a more navigable path than perhaps any other contracting mechanism in government. The SBIR program has the added benefit of protecting intellectual property and promoting small businesses.

However, SBIR is struggling.

Over the last few years, we've watched the quantity of SBIR awardees balloon, the size of awards shrink, and the quantity of transitioned projects slip. This suggests to me that the current system is incentivized by the wrong metric: recognizing departments and agencies for the quantity of awardees versus the number of transitions to programs of record.

I have found it crucial as a small business owner that any government procurement have clear transition milestones for a path to a program of record. A great example of a transition partner is Naval Air Warfare Center, which includes project transition milestones in its contract milestones.

Conclusion

In conclusion, the needs and resources of government missions are unique, requiring tailored AI solutions. Therefore, procurement vehicles must reflect the iterative nature of AI, and the introduction of standards for testing and evaluation will promote more effective AI adoption. Most importantly, AI education and relevant procurement training are both imperatives for increasing AI adoption across the federal enterprise.

Finally, in all phases of an AI project lifecycle, remember that machine learning, just like human learning, is a journey and not a destination.

Thank you for your time.