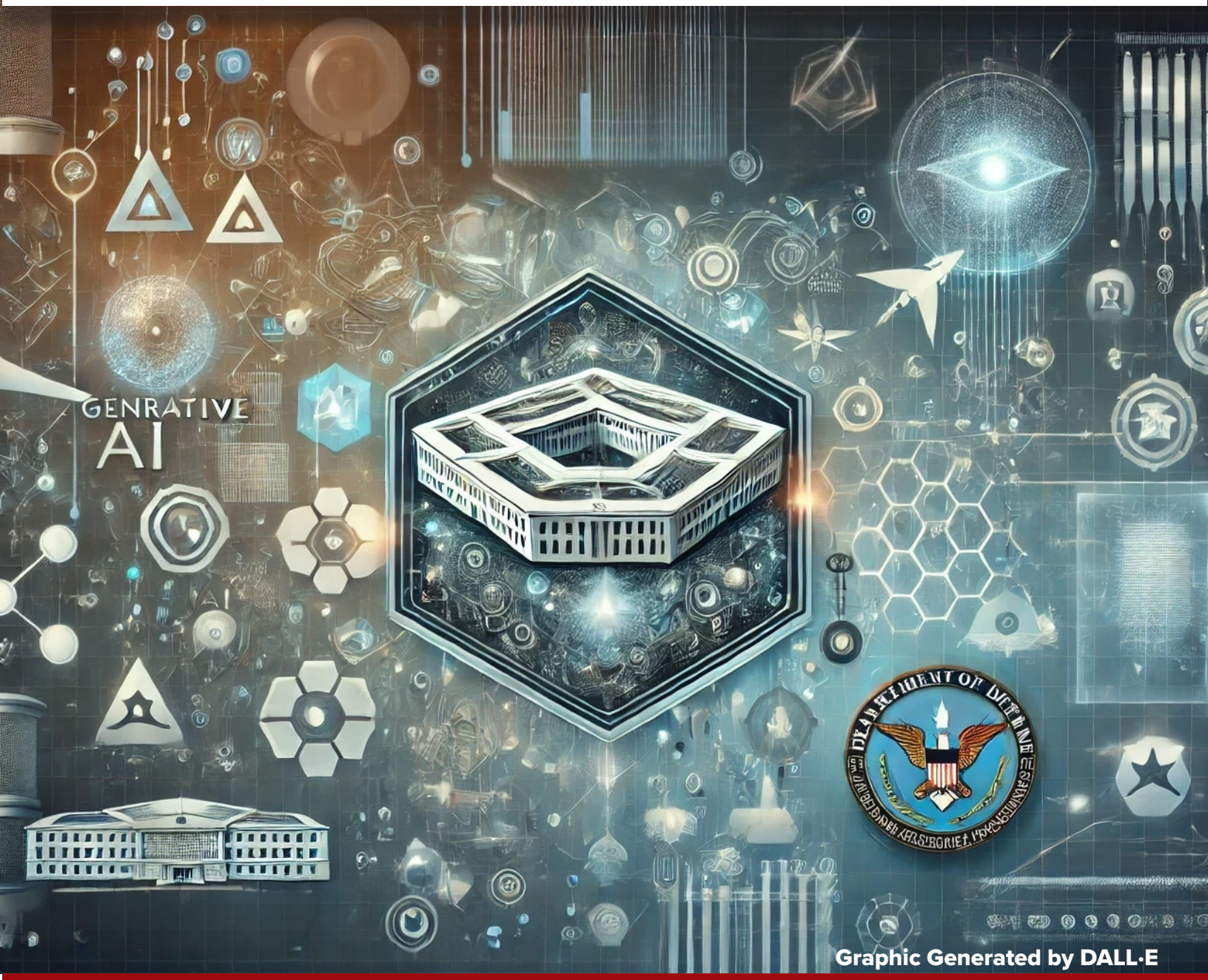
BENS INSIGHTS:
Elevating Thought Leadership in National Security

BUSINESS EXECUTIVES FOR NATIONAL SECURITY
SECURING AMERICA'S FUTURE



Graphic Generated by DALL·E

Adopting Generative AI: Pathways for Defense

Adopting Generative AI: Pathways for Defense

Like the private sector and other organizations striving to improve efficiency and effectiveness, the Department of Defense (DoD) has increasingly embraced generative AI. These advanced systems, capable of creating new content by learning from existing data, are being applied to a diverse array of military applications, from data synthesis to autonomous system design. Yet, for all its promise, the adoption of generative AI by the DoD has not been without hurdles. Complications with existing legacy systems and the need for training and reskilling personnel present significant challenges. To gain insights into these adoption challenges – and possible solutions – BENS recently interviewed industry leaders Benjamin Cheatham, Vice President of Data and Artificial Intelligence at Microsoft; Shaun Modi, founder and CEO of Capitol AI; John Zangardi, CEO of Redhorse Corporation; and Dennis Chornenky, CEO of Domelabs AI.

Generative AI is here. If Defense leaders want to stay at the forefront of innovation and ensure America's military is able to continually meet today's increasingly complex security environment, they need to incorporate its capability quickly and effectively.

Benjamin Cheatham
Vice President of Data and Artificial Intelligence
Microsoft

Shaun Modi
Founder & CEO
Capitol AI

John Zangardi
CEO
Redhorse Corporation

Dennis Chornenky
CEO
Domelabs AI

Interviews have been edited for length and clarity. The views expressed are their own and may not reflect those of their employers or BENS.

What integration strategies should the Department of Defense prioritize to ensure a seamless and ethical deployment of generative AI? Considering the potential risks associated with AI in national security, how can the DoD establish frameworks for safe use, transparency, and accountability throughout the development and deployment lifecycle?

Benjamin Cheatham: I think before we talk about integration strategies—which to me is a reasonably specific question about bringing new technologies into an existing tech stack—I might step back and just say, how should the government, broadly speaking, think about Generative AI? Its potential opportunities for impact, its limitations, etc.

The rate of change in foundational models is exponential, with significant leaps such as from GPT-3 to GPT-4, demonstrating rapid evolution. This underscores the importance for organizations, including the government, to embrace open architecture to avoid dependency on specific models or vendors. The challenge lies in harnessing this technology to enable model swapping and using multiple models for various tasks. Since ChatGPT's release, there has been a proliferation of both proprietary and open-source large language models, as well as smaller, lightweight models that are easily deployed at the edge and trained for specific tasks.

As government departments consider integrating these models, two governing principles should be kept in mind: avoiding vendor lock-in to ensure access to leading-edge technology and the ability to switch quickly and determining which problems to solve with specific models. Large foundational models like GPT-4 and the latest Claude excel at both general and specific tasks, but require substantial storage and computer resources, necessitating cloud hosting.

Many defense applications lack the cloud infrastructure needed for large model deployments, making edge deployment in high-security environments essential. Companies like Microsoft are developing secret and top-secret clouds to securely host multimodal models, but adoption is slow and the tooling on high-secure clouds is less advanced than in commercial clouds. This creates an opportunity for the government to strategically consider where models are deployed. The commercial cloud offers flexibility and a range of technology options, but as security clearance levels increase, the environment becomes more constrained, making it harder to leverage large models fully. Additionally, preventing data leakage into models is crucial. Platforms like Azure ensure proprietary data remains secure and does not leak back into the model.

Shaun Modi: Generative AI is such a fast-moving domain. It is absolutely critical that startups and agile organizations are brought-in to work with the Department. Research comes out every week, if not every day, in this domain. Whether you're trying to streamline logistics and supply chains for DoD, which is ostensibly one of the largest employers in the world, or you're trying to understand open-source information from our adversaries, there are different models for different use cases. DoD needs to bring in small teams to understand the data, and can compare with the models that are available from the outside – creating a fast, iterative loop. The large defense contractors and the largest tech companies are not going to be the right partners for this fight.

John Zangardi: To be fair, DoD has already done a lot. If you think about things like the ethical AI principles that came out in February 2020, or the DoD-responsible AI toolkit, the work that the National Institute of Standards and Technology has done in this area, the work that has been coming out of the Defense Innovation Unit, etc. all these point to a national security apparatus that is moving quickly and smartly to think about AI.

I liken this to years ago when we were looking at data center consolidation and adoption of the cloud. If you go back 10-15 years, it took years to get to a point where the government was moving very smartly to the cloud. And now they've been doing it for a long time.

At the moment, we do not fully understand generative AI, what it means, and its applications. There are many ways AI can be useful, analyzing imagery from ISR, proactively detecting cyber security threats, providing in-depth data analysis, etc. But on the flip-side, there are a lot of problems with it. It uses a lot of computational assets, and it can be used by bad actors to create attacks. The other side of that coin is you could use generative AI to simulate those attacks to train your security operations center personnel. The point is, there is risk that has to be balanced against those advantages just in the cyber security realm.

If you just look at AI in and of itself, there is the major risk of intellectual property being leaked. There's the risk of data training via the introduction of malware. We have to think about third party compliance for Personally Identifiable Information, General Data Protection Regulation, or the California Privacy Rights Act. DoD has done a lot of this. They have organized around it. They are moving out smartly. But anytime you bring new technology in, there is a learning curve. It takes time.

Dennis Chornenky: Every enterprise that's adopting AI – whether it's a federal agency, DoD component, or a private sector organization – they should start with an AI strategy. It should answer key questions and create a compelling narrative for internal and external stakeholders. It helps answer why that organization should adopt AI in the first place, how it will ensure safety through AI governance, and how it will execute efficiently with an AI adoption roadmap.

The AI adoption roadmap helps lay out the use cases that you consider to be the highest potential for your organization, and then match them to the different ways in which you can operationalize those use cases. This is where we get into the buy-vs-build decisions. You have three buckets broadly. One is building it in-house, if an organization has significant IT capabilities. Another one is doing an individual acquisition with an individual vendor. The third bucket is using a platform like Amazon Web Services or Microsoft Azure, because they've got a pipeline of capabilities, features, and tools that may align well with your goals.

If we step back a little bit, we can think of AI strategy at a high level. That helps to answer some of these key questions, including around creating an AI-ready workforce, ensuring that we have AI-ready data, data governance, what partnerships we want to get into, and upskilling with analytics curricula and data literacy. These kinds of things that may be necessary to ultimately get to a point where the enterprise and the workforce is AI-ready and able to take advantage of new AI technologies.

Once we have the strategy level addressed, we can step down to the AI governance level. This is where we can look at four key components or risk lenses:

- **Fairness and Regulatory Compliance:** this includes risks of AI producing unfair outcomes, amplifying existing disparities and biases, and not being in compliance with rapidly emerging policies and regulations.
- **Technical Risks:** this incudes risks that models may not always perform as intended, that they can be manipulated by threat actors, and a broad range of other technical risks. The more we rely on AI to power our mission-critical systems, the more new AI-specific threat vectors this introduces. Beyond guarding against data drift or model drift in terms of performance, systems must be protected from cyber attacks, ensuring data security, lifecycle monitoring, and overall technical reliability and resilience.
- **Financial Risks:** scattershot investments and sprinkling money around in the hope that AI capabilities will mature on their own rarely pays off. Organizations don't have visibility into their AI portfolios and don't manage these technologies in an organized way. Most AI projects tend to be buried within larger software and IT budgets that makes it difficult to see the big picture. And the regulatory environment is forcing us to start carving out AI and managing it separately from other IT. This is a good thing because it gives us the opportunity to manage AI more efficiently.
- **Domain-Specific Validation:** AI needs domain-specific validation for mission-specific environments, related datasets, surrounding technologies, and integrations. There can be a tremendous amount of variability in how different AI models perform in different circumstances and we do not yet have robust performance evaluation frameworks for most of the key use-cases in their unique environments.

For each of these broad risk areas we can apply frameworks that helps us to mitigate those types of risks:

- For regulatory compliance, as well as fairness, bias, ethics, all those questions, we can apply a **responsible AI framework**.
- For technical risks, we can use **model operations (or "modelops")** as a **framework**. This is where we can establish lifecycle monitoring, real-time assurance, reliability and robustness, resilience of systems, defense mechanisms against data poisoning and other forms of AI model manipulation in this rapidly emerging threat environment.
- For the financial component, we can use a **portfolio management framework** that helps us to organize inventories and better understand costs and value. There is a relevant concept of "total cost of ownership," which means not just the cost of building a model, but all the costs related to its deployment, management, and impact on operational workflows that may also require integrations with other systems. This also includes the training necessary for end-users, change management, and cultural impact.
- For the last element of domain-specific validation, sectors and institutions need to develop **domain-specific frameworks** to evaluate performance of applicable models in unique mission environments. There is too much variability here to try to assign any single framework.

Once there is an AI governance policy in place, an organization can proceed to develop and execute on an AI adoption roadmap. Part of the approach here is to balance a bottom-up process with a top-down process for use-case discovery and prioritization, and then apply decision frameworks for buy-vs-build approaches to operationalize the use-cases you want to prioritize.

Given the specialized nature of generative AI and the skills required to effectively implement it, what steps should the Department of Defense take to address the need for training and skill development among military personnel and technical staff?

Benjamin Cheatham: We need to separate what's happening in AI broadly from what's happening with generative AI or large language models. The operation of large language models is transformative because they change the user interface. While they are complex mathematical constructs with deep neural networks, their societal impact comes from the fact that you don't need much technical knowledge to use them. You can interact with these models using natural language and get good answers. In a world trained by Google to write short queries, we've found that providing detailed prompts with context helps the models perform better. Learning how to write effective prompts is a skill, but it doesn't require understanding the mathematics behind the models. Instead, it involves writing well and referencing relevant knowledge, historical events, literature, and media. Providing moral context also improves the accuracy and usefulness of the results. Essentially, we're learning to write well to unleash the power of these models.

When it comes to training military personnel and technical staff, rather than implementing a large training program, I suggest large-scale experimentation. The challenge for the Department of Defense is to create a safe environment that allows thousands of employees to experiment. Over the past couple of years, with access to AI tools embedded in Microsoft products, I've constantly improved my skills through experimentation in a protected environment. This means trying things out in a secure Azure tenant where confidential information isn't leaked but harnessing the power of these models to access both open-source and proprietary data.

Shaun Modi: This space is moving so fast. It's important for the workforce to read the latest research on what's happening in industry, but remember, these are just tools. The first and most important thing is understanding the problems that need to be solved. For those in government that are experts at procurement or bringing in new technology, it's creating that space to understand the needs of the end-users and then creating requirements that can go and let industry come in and help solve those problems. There's this tendency with the government to want to build it themselves, rather than bring it in from the outside. That fundamentally won't work. You don't need to become an expert at generative AI, you don't need to understand how all these models work. You need to understand the problems you want to solve and then work with industry to figure out how we can apply these generative models to create content. Think about all the communications and reporting that happen on a daily basis and how many human hours go into solving those things. What if generative AI can summarize these things in one tenth of the time? We're already seeing that in the private sector and

industry, whether it's white papers, business memos, or communications to customers – that's where the time savings aspect comes in. Generative AI will be a time saver for DoD and the best thing we can do is to identify the things that need to be cut down in terms of time and cost savings.

John Zangardi: There's an argument out there that everyone needs to be very technical if you're going to do this. I have been around program managers for a long time, and there's two kinds of program managers. There are the ones who are very technical, and there are the ones who are very business oriented. I think you need both perspectives as you go forward. You need that technical acumen, but you also need the business side. There have to be opportunities for people to use these generative AI tools in low-risk environments.

If you look at the military academies, bringing AI in is really important. Same with the War Colleges. Start bringing it in where people who have the operational and warfighting experience are, and are talking about how AI might be implemented. The real key here is talking about and understanding the strengths and weaknesses of generative AI in particular scenarios that are warfighter related, within the DoD. There are a lot of startups out there, big tech companies, and system integrators who are making generative AI training available. They should use it.

With the establishment of Task Force Lima, which is designed to assess, synchronize, and employ generative AI capabilities across the DoD infrastructure, how can it ensure the successful adoption of generative AI within the DoD and foster buy-in from a notoriously rigid institution? Are there particular initiatives, collaborations, or frameworks that Task Force Lima could try to implement to expedite the integration process and address unique challenges associated with generative AI in military operations?

Benjamin Cheatham: It comes back to security and providing the government with the assurance they need. This concern isn't limited to the government but also extends to defense contractors who face similar adoption challenges due to perceived and real risks. Investing in and enabling capabilities on the classified side is crucial to address these challenges. The commercial cloud environment has limited benefits for the Department of Defense if there isn't a scaled version operating in secret or top-secret environments. Much of the critical defense work cannot be done without these high-security environments. Therefore, my encouragement to the government is to be more aggressive about migrating applications and workloads to these secure cloud environments. Without a distributed environment to handle the massive storage and computing challenges that these models entail, it's very difficult to unleash their power.

Shaun Modi: The largest companies are part of the problem. That's why our defense budget is ballooning, and the warfighter is provided with lackluster tools. The agility that we have is because the government only knows how to do business with either entrenched interest or the largest of companies. Personally, I've been underwhelmed by the Chief Digital and Artificial Intelligence Office's (CDAO) effort. I'm in industry every day at the cutting edge of AI, and I really don't see much of their presence on the ground, working with startups proactively, and trying to bring them into government.

The usual answer you hear is that the DoD needs to fix their data problem first. To me, the whole point of generative AI is that you can take messy data, PDFs, and other files, feed it into a model, and get an output. Let's not let good be the enemy of perfect here. Let's modernize and start bringing in the next generation of companies. Anduril is a great example, but it can't be the only one. There's a lot of work to do. I think on paper, the CDAO looks good, but I haven't seen the action here.

I'll be very direct about the CDAO, as I worked at the Joint Artificial Intelligence Center before that. There's been a lot of talk about the Defense Innovation Unit, innovation/technology tourism, and a lot of paperwork and press releases. But I have not seen any action when it comes to bringing generative AI into DoD or the Intelligence Community. They want to have the defense contractors just build a custom version, rather than bring in true commercial technology, and then adapt it.

John Zangardi: You get into a question of what works better, the 'stick' or the 'carrot'. I think the 'carrot' works best in a lot of cases. You want to recognize people and programs that are using the tools effectively and doing the right things, and then making sure that people understand and quantify, whether it's dollars saved or improvements. A lot of these things are very quantifiable. DoD does these awards every year in information technology; we need to do the same thing for artificial intelligence. If you've done well as an industry or academic partner, you need recognition to encourage people to go in the right direction. Flowing money to contracts is another positive 'carrot', of course.

There are pathways coming up, trade-winds and other methods, and other contracting authorities (OTAs) providing pathways, but it is very hard in some of these processes to actually get selected. Reducing the barriers to entry is really important. It has to be balanced against the ability of the source selection team and the source selection authority to have enough information to judge whatever has been submitted on a proposal. Balancing the ability to have enough information to judge, with a low enough threshold that smaller startup companies can propose something is very important.

Dennis Chornenky: I've followed the discussion around DoD seeking to speed up the acquisition process. There has also been a lot of talk about generative AI over the last couple of years, especially with key leaders from the Pentagon, Silicon Valley, and other parts of the national security ecosystem. The Defense Innovation Unit is a good example of accelerating collaboration in this area.

The challenge is that, for any federal institution, there tend to be multiple layers of what you can do and what you can't, in terms of your own internal policy, congressional mandates, laws, and authorities. You would need some process and broad consensus on how we can extract that information from the DoD components to clearly articulate what it is that they can change internally, at what levels, or where Congress is going to have to act.

The pace of technology is becoming a forcing function. Generative AI is evolving so quickly that people are recognizing and feeling the pressure from all sides that we have to get better at this. We have to work more closely with legislators, Silicon Valley, and the private sector in creating the necessary pathways.

1030 15th St. NW · Suite 200 East
Washington, DC 20005
www.BENS.org | X: @BENS_org | 202.296.2125