[no markings required]

DM24-0294

10 Benefits and 10 Challenges of Applying Large Language Models to DoD Software Acquisition

by John Robert, Douglas Schmidt

Department of Defense (DoD) software acquisition has long been a complex and document-heavy process. Historically, many software acquisition activities, such as generating Requests for Information (RFIs), summarizing government regulations, identifying relevant commercial standards, and drafting project status updates, have required considerable human-intensive effort. However, the advent of generative artificial intelligence (AI) tools, including large language models (LLMs), offers a promising opportunity to accelerate and streamline certain aspects of the software acquisition process.

Software acquisition is one of many complex mission-critical domains that may benefit from applying generative AI to augment and/or accelerate human efforts. This blog post is the first in a series dedicated to exploring how generative AI, particularly LLMs like ChatGPT-4, can enhance software acquisition activities. Below, we present 10 benefits and 10 challenges of applying LLMs to the software acquisition process and suggest specific use cases where generative AI can provide value to the software acquisition process. Our focus is on providing timely information to software acquisition professionals, including defense software developers, program managers, systems engineers, cybersecurity analysts, and other key stakeholders, who operate within challenging constraints and prioritize security and accuracy.

Assessing the Benefits and Challenges of Generative AI in DoD Software Acquisition

Applying LLMs to software acquisition potentially offers numerous benefits, which can contribute to improving outcomes. There are also important challenges and concerns to consider, however, and the evolving nature of LLM technology can pose challenges. Before attempting to apply generative AI to DoD software acquisition activities, therefore, it is critical to first weigh the benefits and risks of applying these technologies to acquisition activities.

Our colleagues at the SEI recently wrote an article that identifies some LLM concerns that should be considered when deciding whether to apply generative AI to acquisition use cases. Our blog post builds upon these and other observed benefits and challenges when applying generative AI to assess the pros and cons for applying LLMs to acquisition. In particular, some benefits of applying LLMs to software acquisition activities include the following:

1. **Efficiency and productivity**—LLMs can enhance efficiency in software acquisition by automating various tasks, such as generating code, analyzing software artifacts, and assisting in decision making. This automation can accelerate processes and reduce manual effort.

2. **Scalability**—LLMs excel in processing text and data, making them suitable for context-specific summarization and complex inquiries. This scalability is valuable when dealing with extensive software documentation, requirements, or codebases common in DoD acquisition programs.
3. **Customization**—LLMs can be customized through prompt engineering to refine context-specific responses. Acquisition programs can tailor the behavior of these models to suit their specific software acquisition needs, improving the relevance and accuracy of the results.
4. **Wide range of use cases**—LLMs have versatile applications in software acquisition, spanning documentation analysis, requirements understanding, code generation, and more. Their adaptability makes them applicable across multiple phases of software acquisition and the software development lifecycle. LLMs are trained on vast data sets, which means they can contribute to a broad range of software acquisition topics, programming languages, software development methods, and industry-specific terminologies. This broad knowledge base aids in understanding and generating useful responses on a wide range of acquisition-related topics.
5. **Rapid prototyping**—LLMs enable rapid code prototyping, allowing mission stakeholders, acquirers, or software developers to experiment with different ideas and approaches before committing to a particular solution, thereby promoting innovation and agile development practices.
6. **Creativity**—LLMs can generate novel content and insights based on their extensive training data. They can propose innovative solutions, suggest alternative approaches, and provide fresh perspectives during software acquisition phases.
7. **Consistency**—LLMs can produce consistent results based on their training data and model architecture when prompt engineering is performed properly. LLMs have a configuration setting or temperature that enables users to enhance consistency in responses. This consistency helps improve the reliability of software acquisition activities, reducing the chances of human errors.
8. **Accessibility and ease of use**—LLMs are accessible through web services, APIs, and platforms, making them readily available to acquisition programs. Their ease of use and integration into existing workflows helps simplify their adoption in software acquisition. LLMs are also accessible to individuals with diverse backgrounds using a natural language interface. This inclusivity enables a wide range of nontechnical stakeholders to participate effectively in software acquisition.
9. **Knowledge transfer**—LLMs can facilitate knowledge transfer within organizations by summarizing technical documents, creating documentation, and assisting in onboarding new team members, thereby promoting knowledge sharing and continuity.
10. **Continuous learning**—LLMs can adapt and improve over time as they are exposed to new data and prompts via fine-tuning and in-context learning. This continuous learning capability allows them to evolve and become more proficient in addressing software acquisition challenges associated with specific programs, regulations, and/or technologies.

LLMs are still an emerging technology, however, so it's important to recognize the following challenges of applying LLMs to software acquisition activities:

1. **Incorrectness**—LLMs can produce incorrect results—often called *hallucinations*—and the significance of this incorrectness as a concern depends on the specific use case. Mistakes in code generation or analysis can yield software defects and issues. The accuracy of LLM-generated content must be verified through consistent testing and validation processes. LLM governance for enterprise solutions requires consistent tracking and monitoring of LLMs as part of a responsible AI framework.
2. **Disclosure**—Sensitive information must be protected. Some software acquisition activities may involve disclosing sensitive or proprietary information to LLMs, which raises concerns about data security and privacy. Sharing confidential data with LLMs can pose risks if not properly managed (e.g., by using LLMs that are in private clouds or air-gapped from the Internet). Organizations should be aware of how to mitigate the enterprise security risks of LLMs and prevent access to private or protected data. Data firewalls and/or data privacy vaults can be used to enforce some data protections across the enterprise.
3. **Usability**—Although access and ease of use are strengths of LLMs, some new skills are required to use them effectively. LLMs require users to craft appropriate prompts and validate their results. The usability of LLMs depends on the expertise of users, and many users are not yet proficient enough with prompt patterns to interact with these models effectively.
4. **Trust**—Users must have a clear understanding of the limitations of LLMs to trust their output. Overreliance on LLMs without considering their potential for errors or bias can lead to undesirable outcomes. It is essential to remain vigilant to mitigate bias and ensure fairness in all content and implement strategies produced via generative AI. Although LLMs can only be effective if bias is understood, there are various resources for LLM bias evaluation and mitigation.
5. **Context dependency and human oversight**—LLMs' effectiveness, relevance, and appropriateness can vary significantly based on the specific environment, use case, and cultural or operational norms within a particular acquisition program. For example, what may be a significant concern in one context may be less important in another. Given the current state of LLM maturity, human oversight should be maintained throughout software acquisition processes to ensure people—not LLMs—make informed decisions and ensure ethical compliance. The NIST AI Risk Management Framework also provides important context for proper use of generative AI tools. When possible, LLMs should be provided specific text or data (e.g., via in-context learning and/or retrieval-augmented generation (RAG)) to analyze to help bound LLM responses and reduce errors. In addition, LLM-generated content should be scrutinized to ensure it adheres to enterprise protocols and standards.
6. **Cost**—The costs of LLMs are changing with higher demand and more competition, but cost is always a consideration for organizations considering using a new software application or service in their processes. Some tactics for addressing privacy concerns, such as training custom models or increasing compute resources, can be costly. Organizations need to assess the total costs of using LLMs in their organization, including governance, security, and safety protocols, to fully consider the benefits and the expenses.

7. **Constant evolution**—LLM technology is continually evolving, and the effectiveness of these models changes over time. Organizations must stay current with these advances and adapt their strategies accordingly.
8. **Intellectual property violations**—The expansive training data of LLMs can include copyrighted content, leading to potential legal challenges when applied to developing or augmenting code for software procurement.
9. **Adversarial attack vulnerabilities**—Adversarial machine learning can be used to trick generative AI systems, particularly those built using neural networks. Attackers can use various methods, from tampering with the data used to train the AI to using inputs that appear normal to us but have hidden features that confuse the AI system.
10. **Over-hyped LLM expectations of accuracy and trustworthiness**—The latest releases of LLMs are often highly capable but are not a one-size-fits-all solution to solving all software acquisition challenges. Organizations need to understand when to apply LLMs and what types of software acquisition challenges are best suited to LLMs. In particular, applying LLMs effectively today requires a savvy workforce that understands the risks and mitigations when using LLMs.

Expanding Use Cases for Generative AI in Software Acquisition

By considering the benefits and challenges identified above, software acquisition professionals can identify specific use cases or activities to apply generative AI risk prudently. Generative AI can help on many activities, as indicated by *ChatGPT in DoD Acquisitions* or *Assessing Opportunities for LLMs in Software Engineering and Acquisition*. Some specific software acquisition activities we are exploring at the SEI to determine the benefits and challenges of applying generative AI include the following:

- **Document summarization**—Understanding large acquisition documents or multiple documents takes extensive and expensive human effort. LLMs can provide summaries of documents and provide an interactive environment for exploring documents.
- **Regulatory compliance**—Keeping up with evolving government regulations is essential for DoD software acquisition. LLMs can continuously monitor and summarize changes in regulations, ensuring that acquisition activities remain compliant and up to date.
- **Standard identification**—Identifying relevant commercial standards is a time-consuming task. LLMs can methodically parse through vast databases of standards and provide recommendations based on project specifications, saving time and reducing errors.
- **RFI generation**—Generating RFIs is a crucial step in the software acquisition process. LLMs can assist in drafting comprehensive and well-structured RFIs by analyzing project requirements and generating detailed questions for potential contractors.
- **Proposal evaluation**—Evaluating proposals from contractors is a critical phase in software acquisition. LLMs can assist in automating the initial screening of proposals by extracting key information and identifying (non-)compliance with requirements.
- **Risk assessment**—Assessing risks associated with software acquisition is vital. LLMs can analyze historical data and project-specific details to predict potential risks and suggest mitigation strategies.

- **Project status updates**—Keeping stakeholders informed about project status is essential. LLMs can generate concise project status reports by summarizing large volumes of data, making it easier for decision makers to stay updated.

Government Regulations and Guidance for Using Generative AI

Publicly available generative AI services are relatively new, and U.S. government regulations and directives are changing to adapt to the new technology. It is important for any DoD acquisition stakeholders who are considering using generative AI tools to be aware of the latest guidance, including security concerns, to ensure compliance with the changing regulatory landscape. Some recent examples of government guidance or emerging policy related to generative AI include the following:

- *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
- *Security Policy for Generative Artificial Intelligence (AI) Large Language Models (LLMs)*
- *DoD Announces Establishment of Generative AI Task Force*
- *DON Guidance on the Use of Generative Artificial Intelligence and Large Language Models*

Looking Ahead

While generative AI offers many potential benefits for acquisition professionals, it is essential for DoD programs and acquisition professionals to evaluate how LLMs may (or may not) align with their specific software acquisition needs critically and objectively, as well as formulate strategies to address potential risks. Innovation in software acquisition using generative AI is about increasing productivity for acquirers and stakeholders while mitigating risks. Humans must continue to have a central role in the software acquisition activities, and humans that can best leverage new generative AI tools safely will be crucial to all stakeholders.

Deliberate exploration of LLMs within the DoD's acquisition processes is key to gaining insights into both their benefits and potential pitfalls. By comprehending the capabilities and limitations of generative AI, software acquisition professionals can discern areas where its application is most advantageous and the risks are either manageable or minimal. Our next blog post in this series will delve into particular instances to facilitate cautious experimentation in software acquisition activities, enhancing our grasp of both the opportunities and risks involved.

Additional Resources

Register for *The Future of Software Engineering and Acquisition with Generative AI* webcast, which will be held Wednesday, January 24, featuring Anita Carleton, John Robert, Doug Schmidt, Ipek Ozkaya, James Ivers, and Shen Zhang - https://www.eventbrite.com/e/the-future-of-software-engineering-and-acquisition-with-generative-ai-tickets-777902867417.