National Artificial Intelligence Institute

OFFICE OF THE CTO

CURRENT GUIDANCE FOR **GENERATIVE AI MODELS AT VA**

JOINT INFORMATION BULLETIN

NATIONAL ARTIFICIAL INTELLIGENCE INSTITUTE (NAII) AND
OFFICE OF THE CHIEF TECHNOLOGY OFFICER (OCTO)

The National Artificial Intelligence Institute (NAII) and the Office of the Chief Technology Officer (OCTO) have partnered to issue this joint information bulletin on generative AI at VA. This bulletin is intended to provide VA's current guidance on public generative AI services, a summary of currently known risks and pointers to learn more about the responsible use of generative AI.

WHAT IS GENERATIVE AI AND HOW IS IT USED?

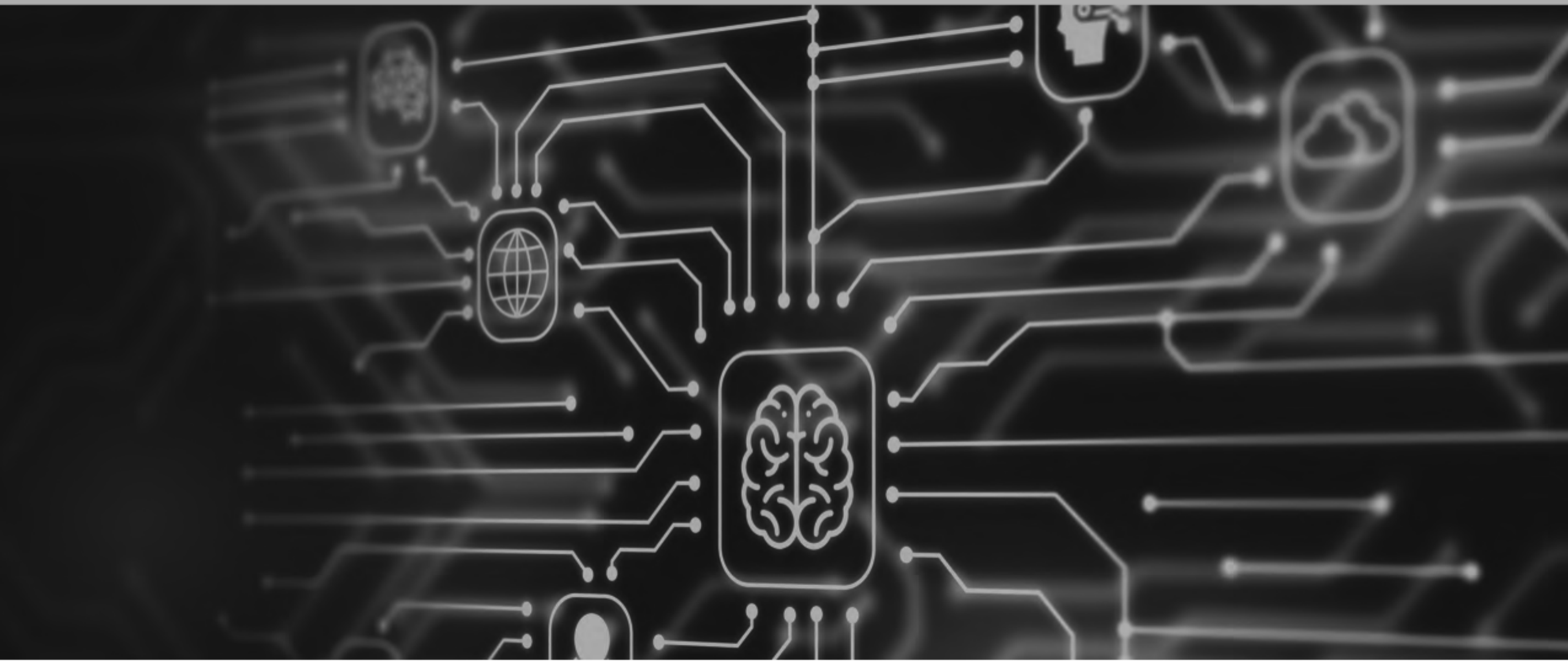
▸ AI tools collectively referred to as generative AI and Large Language Models (LLM) are systems that can rapidly create synthetic text, images, source code and other forms of media. These outputs can be realistic or stylized based on customizable prompts provided by the user. LLMs are a form of deep learning that underpin ChatGPT and other Natural Language Processing (NLP) technologies currently in the news. When used in some narrowly defined applications, generative AI tools excel at tasks, such as:

- **Producing** human readable summarization of large amounts of text
- **Generating** computer source code or suggest code completions
- **Quickly transforming** text from one format to another

WHAT ARE THE RISKS OF GENERATIVE AI?

▸ While there is wide interest in the benefits of generative AI models, this technology introduces new risks and unknown consequences that can have a significantly negative impact on the privacy and safety of Veterans. Major risks include:

- **Misinformation:** Output from LLM tools, like ChatGPT, may contain errors or fabricated information
- **Bias and discrimination:** The manner of training LLMs introduces and amplifies biases in the training data
- **Threats to data privacy and security:** Control and provenance of data and queries are lost once submitted to public generative AI services
- **Abuse and fraud:** Guardrails and safeguards in common generative AI tools can be bypassed to allow intentional misuse of the model's output and violation of responsible use guidelines

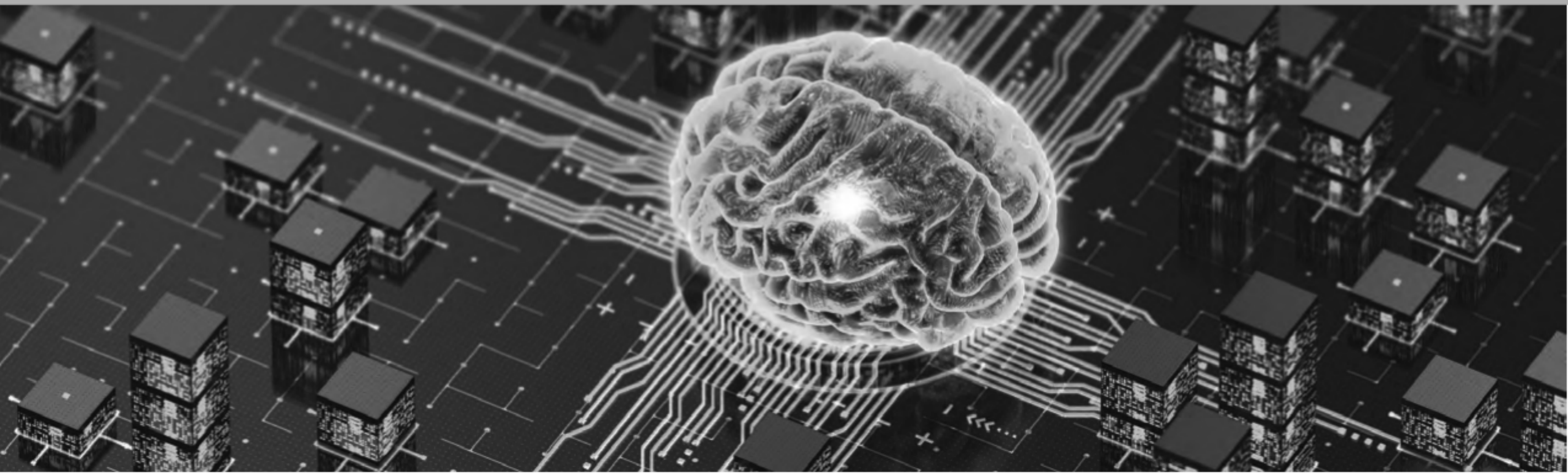


WHAT IS VA'S CURRENT GUIDANCE FOR GENERATIVE AI TECHNOLOGIES?

- **No web-based, publicly available generative AI service has been approved for use with VA sensitive data.** Examples of these include OpenAI's ChatGPT and GPT4, Google's Bard, Anthropic's Claude, and Microsoft's new Bing Search. VA follows existing federal requirements and processes to ensure VA data is protected. When users enter information into an unapproved web-based tool, VA loses control of the data. Some public LLM web services have terms of service that explicitly allow them to use the data entered into the tool for other purposes.
- **No Personally Identifiable Information (PII), Protected Health Information (PHI) or VA sensitive data should be entered into these unapproved services**. VA sensitive data includes: financial, budgetary, research, quality assurance, confidential commercial, critical infrastructure, and investigatory and law enforcement information.
- **Where possible, limit the sharing and saving of data in unapproved services.**
- **VA staff should carefully evaluate the output of any LLM tool for accuracy before using the output in VA work products.** LLMs are known to generate inaccurate information that sounds plausibly true, and VA staff are responsible for the accuracy of their work products.

WHAT IS VA DOING TODAY TO ASSESS THE VALUE OF GENERATIVE AI?

NAII and OCTO formed a partnership to closely track and communicate with one another on updates in generative AI technologies and risks across government and in industry. The VA Office of Information & Technology (OIT) is coordinating carefully scoped, low-risk generative AI pilot projects. For example, OCTO is exploring if and how generative AI models can improve the reliability and trustworthiness of chatbot responses to Veteran questions on VA services and benefits. Due to privacy and security concerns of LLMs, all evaluation processes are internal to VA and under very controlled environments. Another example is an internal pilot of GitHub Copilot, an AI pair programmer that offers code suggestions. In parallel, the VA Chief AI Officer (CAIO) and National Artificial Intelligence Institute (NAII) Director, in coordination with the VA Data Governance Council (DGC) AI Working Group (AIWG), are currently developing trustworthy AI implementation guidance for VA AI system owners, including generative AI.



TIPS FOR ASSESSING GENERATIVE AI TECHNOLOGIES

In the meantime, here are some steps you can take to help VA assess and use these technologies in a responsible manner:

- **Purpose:** Consider the purpose of any decisions to use generative AI technologies and ensure that the scope of application is well-defined so that the benefits of use outweigh the risks.
- **Safety & Efficacy:** Understand the limits of generative model accuracy. If your group is considering generative AI technologies, it is imperative to develop a process to verify the accuracy of the output and ensure that they are not used without human oversight or for automated decision making.
- **Security & Privacy:** As in other situations, do not share VA sensitive information, including PII or pre-decisional documents, with public services that pose privacy or security risks, including third-party APIs or interfaces for AI tools. It is also advised to not bypass VA security controls to access public LLM services, such as ChatGPT and Bard.
- **Fairness & Equity:** VA policies and federal requirements regarding trustworthy AI, including the VA *Principle-Based Ethics Framework for Access to and Use of Veteran Data*, VA cybersecurity and privacy guidance in VA Handbook 6500, and Executive Order 13960 apply to any use of generative AI models.
- **Transparency & Explainability:** Generative AI models do not provide explanations of how they produce their outputs. To avoid violating transparency rules, laws and policies, exercise extreme caution when using generative models as support for higher-risk scenarios and understand when use of AI tools needs to be disclosed.
- **Accountability & Monitoring:** VA personnel do not abdicate accountability for actions, decisions, or outcomes informed or produced by generative AI tools. At this stage, actions influenced by generative tools should be safeguarded with full human oversight.

WHERE CAN YOU GO FOR MORE INFORMATION?

- Generative AI technologies, and especially LLMs, are evolving rapidly. This bulletin will be updated as new information becomes available. In the meantime, here are resources for more information.
  - Join the **AI@VA Community Hub** run by the NAII **(NAII@va.gov)** for collaborations and information sharing about AI within VA.
  - Reach out to the Office of the CTO's data team in OIT (POC: **Kimberly.Mcmanus@va.gov**) who are exploring small-scale pilots of generative AI at VA.
  - Review **OIS Information Security Do's and Don'ts** and **Privacy Do's and Don'ts.**
  - Review section 5 of the **VA Directive 6500** and its definitions of VA sensitive data, PII and PHI.
  - NAII maintains an **AI use case inventory**, which will capture upcoming generative AI projects.
  - Generative AI pilots are also tracked in the **VHA Innovation Ecosystem Registry.** Connect with the **Office of Healthcare Innovation and Learning (OHIL)** on innovative projects currently underway at VA.

Last Updated: July 3, 2023