DEFENSE INNOVATION UNIT

RESPONSIBLE ARTIFICIAL INTELLIGENCE

2022 IN REVIEW

WWW.DIU.MIL

EXECUTIVE SUMMARY

As part of its mission to accelerate adoption of commercial technology within the Department of Defense (DoD), the Defense Innovation Unit (DIU) launched a strategic initiative in March 2020 to integrate the DoD's Ethical Principles for Artificial Intelligence (AI) into its commercial prototyping and acquisition programs. Drawing upon best practices from government, non-profit, academic, industry, and international partners, DIU explored methods for implementing these principles in several of its AI prototype projects. The result is a set of Responsible Artificial Intelligence (RAI) Guidelines, originally published in November 2021.

DIU's RAI Guidelines aim to provide a clear, efficient process of inquiry for personnel involved in AI system development (e.g., program managers, commercial vendors, or government partners) to achieve the following goals:

- Ensure that the DoD's Ethical Principles for AI are integrated into the planning, development, and deployment phases of the technical lifecycle;
- Effectively examine, test, and validate that all programs and prototypes align with DoD's Ethical Principles for AI; and,
- Leverage a process that is reliable, replicable, and scalable across a variety of AI programs.

DIU's RAI Guidelines are presented in the form of detailed worksheets that instruct and guide AI vendors, DoD stakeholders, and DIU program managers on how to properly scope AI problem statements. These worksheets also provide detailed guidance on the considerations that each of these stakeholders should keep in mind as they proceed through each phase of AI system development.

The purpose of this report is to capture key observations and recommendations based upon DIU's RAI initiative in the 2022 calendar year. In 2022, 11 technology vendors engaged in active prototypes at DIU piloted the RAI Guidelines. The RAI Guidelines process comprises three distinct stages: Planning, Development, and Deployment. This update contains an overview on the status of active RAI projects, training activities, as well as observations on progress and recommendations for improvement.

TABLE OF CONTENTS

OVERVIEW,

As part of its mission to accelerate adoption of commercial technology within the Department of Defense , the Defense Innovation Unit launched a strategic initiative in March 2020 to integrate the DoD's Ethical Principles for Artificial Intelligence into its commercial prototyping and acquisition programs. Drawing upon best practices from government, non-profit, academic, industry, and international partners, DIU explored methods for implementing these principles in several of its AI prototype projects. The result is a set of Responsible Artificial Intelligence Guidelines, originally published in November 2021. The Guidelines consist of a set of questions, captured in worksheets, that must be addressed by technology vendors, DoD partners and DIU Program Managers (PMs) at the planning, development and deployment stages within a project lifecycle.

The purpose of this report is to capture observations and recommendations based upon RAI work completed to date. This report's recommendations are divided into four categories: legal, procurement, technical, and operational.

LEGAL CONSIDERATIONS

Protecting vendor intellectual property.

Observation: Some vendors raised concerns that the worksheets sought potential proprietary materials or information. In one instance, a vendor was reluctant to share information about how they were checking for unintended bias out of concern that this would reveal commercially sensitive information about their algorithm. This was overcome in two ways. First, DIU assured the vendor that information provided in response to the worksheets would be considered confidential and that the company would be contacted to obtain permission before any information was shared (barring a legal requirement to disclose). Second, DIU clarified that disclosure of source code was neither necessary nor sufficient to address the relevant question. Rather, the vendor was asked to provide a description of metrics utilized during training and deployment. These measures were able to resolve all of the vendor's concerns.

Recommendation(s): Create incentives and safeguards for companies to disclose RAI-related issues without incurring negative legal or contractual consequences. Clarify the level of technical disclosure required and ensure that proprietary information remains protected throughout the RAI process. Only request data directly relevant to ensuring compliance.

Clarifying legal requirements vs. program guidelines.

Observation: Some companies were confused as to the legal status of the RAI guidelines which led to unnecessary back-and-forths with company counsel to clarify. In general, it is important to distinguish between legal requirements and program recommendations; DIU's guidelines very much fall in the latter category. Companies may be contractually obligated to comply with the guideline – usually as evidenced by DIU acceptance of a completed RAI Guidelines worksheet – but successful compliance is not defined by a particular statute.

Recommendation: Clarify to vendors that the RAI process is meant to be collaborative, iterative and focused less on compliance than with surfacing issues for discussion. A successful implementation of the RAI process does not mean that issues have been resolved but that they have at least been surfaced and documented. Vendors should feel comfortable sharing system vulnerabilities and risks without fear of legal or contractual downside.

PROCUREMENT CONSIDERATIONS

Integrating RAI workplan into Statement of Work (SOW).

Observation: Vendors' abilities to plan and price RAI work depend upon establishing an up-front estimate of the level of effort required.

Recommendation: RAI should be included in the initial statement of work so vendors can price and plan appropriately. Completing RAI worksheets to an acceptable standard takes time and resources. This necessitates advanced planning to make sure that appropriate funding is available, and that timelines accommodate both vendors completing worksheets and meetings to review worksheets once completed. An RAI plan that covers responsibilities, timelines and costs should be agreed upon and incorporated into the statement of work prior to project commencement.

RAI as the leading indicator for company competence.

Observation: In the projects carried out over 2022, there was a correlation between companies that possess a solid grasp of RAI requirements and overall performance. Thus, presence or absence of an RAI strategy within a request for proposal (RFP) can be a valuable indicator of general technical competency.

Recommendation: RAI should be addressed during the RFP so companies have an opportunity to present their strategy; companies that are differentiated along an RAI axis should be able to translate that differentiation into an advantage during downselection.

TECHNICAL CONSIDERATIONS

Clarify technical level of detail required in worksheets.

Observation: In a number of instances, the initial responses from vendors to the worksheets consisted of very short answers that did not provide enough information to meaningfully assess whether or not the program was in compliance.

Recommendation: Program managers and their technical support should state up-front the level of effort required, and ideally provide examples where possible. A rough ballpark is 2-6 hours per worksheet / per stage, however this will range considerably depending on the overall risk profile of the project. Some questions, such as harm modeling or plans for mitigating errors, may require extensive elaboration in order to achieve a satisfactory answer, while others, such as identifying appropriate metrics, *should* require a more concise response.

Need to improve post-deployment monitoring.

Observation: In many cases, DoD does not have appropriate personnel or tools to monitor vendors' performance and are dependent on vendors to self-monitor. This problem is compounded by the fact that DoD customers rarely set aside money for the purposes of funding post-deployment model monitoring.

Recommendation: DoD needs to establish appropriate technical standards for audits / post deployment monitoring. At a program level, projects should have access to tools for machine learning observability, explainability, etc. Ideally, these tools can be made available at little or no cost to programs and are sustained at a centralized repository like the Chief Digital and Artificial Intelligence Office (CDAO). This recommendation to establish appropriate technical standards for audits and monitoring aligns with the recent Joint Requirements Oversight Council Memorandum (JROCM), "Creating a Federated Artificial Intelligence Enterprise." In addition, DoD personnel require training on how to conduct an AI audit – to our knowledge, this has not yet been done within the defense enterprise for a deployed system.

Competing technical standards for RAI.

Observation: The International Organization for Standardization (ISO), Institute of Electrical and Electronics Engineers (IEEE), National Institute of Standards and Technology (NIST), Responsible AI Institute, and others have all produced various technical standards for implementing RAI. However, to-date, the DoD has not released any official guidance on which of these standards are more or less aligned with internal guidance. Consequently, vendors may be confused which, if any, of these standards should be adopted.

Recommendation: DoD (likely via CDAO) should release guidance on whether and to what extent third-party standards should be adopted, treated as interchangeable, etc.

OPERATIONAL CONSIDERATIONS

Human resources required.

Observation: Despite trying to make the RAI guidelines as lightweight as possible, the human resources demand is considerable. At DIU, we are fortunate to have a number of technical subject matter experts who are familiar with RAI principles. Their expertise was integral to the success of applying the guidelines. Each project that went through the process was led by a DIU program manager, responsible for overall project management, a DIU technical subject matter expert (SME), and support from a dedicated DIU RAI team composed of both technical and ethical SMEs. On the vendor side, successful engagements relied upon support from the engineering teams to complete the required documentation.

Recommendation: While not all programs require a dedicated RAI SME, all programs that undergo RAI review *do* require access to both technical and RAI specific expertise. Projects tended to work best when RAI was delegated to a technical SME, who could then draw upon external resources as needed, with overall coordination from the program manager. Clarifying to the vendor that technical SMEs should be the points of contact for RAI artifacts helped them gather the proper resources before and during RAI execution.

Need to triage projects through preliminary risk assessment.

Observation: As is evident from the breadth of projects that underwent DIU's RAI guidelines, the risk profiles varied enormously. At present, there is no systematic way for triaging projects in terms of how much RAI resources are required. While a risk assessment is included in the Planning phase of the DIU RAI guidelines, this is really meant to create a common understanding among the project team members of risks associated with the process. It is *not* designed to decide what resources should be allocated for RAI review since it is part of that review.

Recommendation: DoD (likely via CDAO) should release a lightweight, risk assessment tool for pre-screening AI projects. The tool should systematically evaluate different aspects of a program to determine the appropriate type / level of resources required to comply with RAI guidelines.

RAI needs to be integrated into Machine Learning Operations (MLOps).

Observation: The questions and design of the RAI worksheets display links between MLOps and commercial vendors' development and deployment strategies. However, most vendors that have reached the Deployment stage of the RAI process have been hesitant and/or delayed in sharing their MLOps strategy with the government, unless this is specifically required. Further, AI observability and explainability techniques have often been deprioritized by vendors due to the complexity of building those tools into existing MLOps pipelines or a lack of internal resources to support such efforts.

Recommendation: The DoD needs to proactively understand vendors' internal MLOps pipeline and ensure that RAI tools are or will be integrated. In particular, companies should either have their own or be provided with tools that offer model monitoring (for data drift, bias, etc.) and explainability. Preferably such tools are hosted and funded by a centrally located program office, (e.g., within the CDAO).

CONCLUSIONS & NEXT STEPS

DIU intends to continue and expand upon the RAI work accomplished to-date. In the near term, DIU is focused on making sure that DoD partners and vendors have access to the best available tools for implementing the technical requirements of RAI. As an example, DIU is working with the U.S. Navy to procure various MLOps tools for model monitoring and explainability that will ideally be integrated into the service-wide Overmatch Software Armory.

DIU is also engaged in representing the Undersecretary of Defense Research and Engineering (USD R&E) on the CDAO-led RAI Working Council charged with operationalizing the DoD RAI Strategy and Implementation Pathway. In this capacity we are focused on creating standardized documentation, maintaining active dialogue with industry on the state of RAI, and establishing structures for reporting and managing RAI issues as they arise across both USD(R&E) and the broader Department.

APPENDIX A: Leading Responsible AI Workshop Materials

During the Workshop

- Description of the Pizza Delivery Drones
- Planning Worksheet for Pizza Delivery Drones

Workshop Resources

- Presentation
- Facilitators Guide
- Session Planning Form

DIU RAI Report and Updated Worksheets

- DIU Responsible AI Guidelines Report
- Planning Worksheet
- Development Worksheet
- Deployment Worksheet