SYM-AM-24-095

Excerpt from the Proceedings of the Twenty-First Annual Acquisition Research Symposium

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2024

Published: April 30, 2024

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Using Natural-Language Processing and Large Language Models to Restructure DoD Comptroller Budget Materials into Portfolio Views

Jose Ramirez-Marquez—is an Associate Professor and Division Director of Enterprise Science and Engineering in the School of Systems and Enterprises at Stevens Institute of Technology. His research focuses on the development of mathematical models for the analysis and computation of system operational effectiveness research and explores the interplay between data visualization and analytical decision-making. He conducted funded research for industry and government and published more than 100 refereed manuscripts in technical journals, book chapters, and industry reports. He holds degrees from Rutgers University in industrial engineering (PhD and MSc) and statistics (MSc) and from Universidad Nacional Autonoma de Mexico in actuarial science. [jmarquez@stevens.edu, 201-216-8003]

Doug Buettner—is the Acquisition Innovation Research Center (AIRC) Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu, 201-216-8300]

Joshua Gorman—is an undergraduate at Stevens Institute of Technology pursuing a Bachelor of Science in computer science. Starting this summer, he will begin his career as a Software Engineer Associate, drawing upon his strong background in modern web development to upgrade pivotal in-house tools. Previously, Gorman garnered valuable experience as a research assistant, actively contributing to diverse application development projects ranging from meteor strike modeling to nuclear fallout predictions. He aims to continue his education by pursuing an accelerated master's program in computer science later this year, to which he is looking to concentrate in cloud computing and computer networking. [jgorman4@stevens.edu, 201-216-8003]

Aashita Patel—is a sophomore at Stevens Institute of Technology. She is majoring in business and technology with concentrations in data analytics and management, and a minor in entrepreneurship. At Stevens, she is participating in an Inclusive Leadership Certification Program to fuel her understanding of diversity, equality, and inclusiveness. She has past professional experience researching Women's Leadership, working as a Sales Associate, and Tutor. After graduation, Aashita plans to work in Business Management. apatel50@stevens.edu, 201-216-8003]

J. Matthew Mercado—is a graduate student at George Mason University's Master's in Business Administration program. He has past professional experience working as a full stack software engineer, SQL database administrator, network operations technician, IT help desk specialist, as well as a pizza shop manager, a volunteer English teacher, and a coding instructor for elementary school children! Mercado plans to join the Navy Reserve and work towards a PhD someday focusing on data science and/or emerging technology usage within the realm of federal contracting. [jmercad8@gmu.edu, 703-993-6954]

Hoong Yan See Tao—is a Research Project Manager at the Systems Engineering Research Center (SERC). She conducts and manages research in mission engineering, digital engineering, human capital development, systems engineering, and systems management transformation. She was on the Stevens Sustainability Committee and part of the student-led organizing committee for the inaugural Student Sustainability Symposium in 2016. In 2013, she was the president of the INCOSE Student Division at Stevens. She earned her BS and MS in electrical engineering from the University of Texas at El Paso and her PhD in systems engineering from Stevens Institute of Technology. [hseetao@stevens.edu, 201-216-8300]

Victoria Cuff—is an AIRC research fellow aiding OSD in research, focusing on PPBE Commission, workforce initiatives, and improving acquisition management policies. Previously, she was the Chief

Growth Officer at Sagely, a learning and development start-up focused on individualized learning for digital upskilling and reskilling. As the Senior Advisor for Agile Acquisition, she led the BA-8 Pilot for digital technology and created the Digital DNA course for acquisition professionals. Earlier, she contributed significantly to USAF's Kessel Run as the Agile Acquisition lead. Cuff holds a bachelor's in actuarial science, financial mathematics, and master's in applied mathematics from Clemson University. [vcuff@stevens.edu, 201-216-8300]

Philip Antón—is the Chief Scientist at the Acquisition Innovation Research Center (AIRC). He conducts applied research, modeling, piloting, education, and transition of innovative ideas and approaches to improve acquisition outcomes. From 1998 to 2021, Antón conducted policy research at the RAND Corporation. From 2011 to 2016, Antón directed the Acquisition Policy Analysis Center for the Under Secretary of Defense for Acquisition, Technology, and Logistics, analyzing acquisition performance and policies while crafting affordability policy. Antón's MS and PhD are from University of California, Irvine, specializing in computational neuroscience and artificial intelligence. His BS is in computer engineering from UCLA. [panton@stevens.edu, 201-216-8300]

Michael McGrath—is an AIRC Research Fellow and independent consultant with broad government and industry experience in technology, management, logistics, and acquisition. His prior senior DoD positions include DASN(RDT&E), ADUSD(Dual Use and Commercial Programs), a DARPA Program Manager, and OSD Director of the DoD Computer-aided Acquisition and Logistics Support program. He has served on multiple Defense Science Board and National Academies studies. His research interests are in manufacturing, cybersecurity, and data science. McGrath holds a doctorate in operations research from George Washington University. [mmcgrat1@stevens.edu, 703-407-3946]

Abstract

When the yearly President's Budget (PB) is submitted to Congress by the Department of Defense (DoD) Comptroller Budget Documents is complemented by "Justification Books" or "J-Books." These detailed documents provide budgetary details by individual programs, projects, and activities within individual military departments or defense agencies rather than from an integrated portfolio or mission perspective. This disjointed structure makes it difficult for non-DoD insiders (and likely congressional staff for whom these materials are intended) to understand the net operational effect of the requested investments let alone their constituent program elements. Given that final reports Section 809 Panel and subsequent statutes provide that the DoD should use data-driven portfolio management for acquisition and capability investments, we asked how the existing J-Book documents could be restructured to facilitate a portfolio view. Our paper first provides the results from our exploratory use of natural language processing (NLP) techniques to perform a key word search across multiple J-Books to extract and subsequently process the content associated with a key word. For the purposes of demonstration, we focused on using these techniques to identify disparate elements of Joint All Domain Command and Control (JADC2) in these J-Books. JADC2 was chosen as this DoD strategy spans multiple service's Research, Development, Test & Evaluation (RDT&E) J-Book volumes. This research also explored whether emerging large language models (LLMs) could be used to answer different types of portfolio or other questions about DoD spending without changing the existing layout and document delivery approach. We provide the results for our implementation of a dashboard proof-of-concept with an LLM interface from refactoring these budget materials including a temporal analysis of the J-Books content spanning multiple years. The final demonstration's use case is from the perspective of a new congressional staffer trying to understand the differences between these budget materials across the years.

Introduction

The Stevens Institute of Technology's Systems Engineering Research Center (SERC)/Acquisition Innovation Research Center (AIRC) was tasked by the PPBE commissioners to research the following areas,

1. Conduct case studies of technology transition.

2. Provide PPBE process research and analysis with recommendations to determine the following:
    a. if the process should be the same for programs that breach their Major Defense Acquisition (MDAP) threshold (10 USC 4201), their Major Systems threshold (10 USC 2302d) and non-major systems,
    b. examine how the Department of Defense (DoD) uses acquisition pathways,
    c. the legal foundations that drive PPBE and develop a matrix outlining how PPBE components are directed (i.e., statute, regulation, policy, or practice).
3. Explore ways to restructure budgets and artifacts (Portfolio Budgeting, J-Books, and Selected Acquisition Reports [SARs]) around threats, missions, operations, and portfolio levels rather than the level of acquisition programs.
4. Explore potential reform areas that rely on coordinated changes in all three of the PPBE, Requirements, and Acquisition communities.
5. Explore options for restructuring the DoD's President's Budget (PB) Proposal including potential groupings or combinations of budget activities (BAs) paying particular attention to how authorizers and appropriators can still understand the program's phase within potential new structures providing historical insight into how and why the current budget structure was created and evolved.
6. Explore alternative Obligation (Obs) and Expenditure target curves/profiles instead of linear targets. (This task was minimally expanded to investigate the impact of Continuing Resolutions [CRs] with the same analysis.)

Our third and fifth tasks wanted us to explore ways of restructuring budget artifacts by restructuring the PB proposal and overlapped in one significant manner, which is reviewing new streamlined methods for how they are submitted. Hence, one of our early goals was to review The DoD budget documents, also called "Justification Books" or "J-Books." These are detailed documents that justify their budgetary requests and resource allocation for specific programs, projects, or activities within the DoD. These documents are typically produced as part of the President's Budget request to Congress in the U.S. federal budgeting process and serve several essential purposes, such as:

Resource Justification: J-Books provide a comprehensive rationale for allocating resources, including funds, personnel, and equipment, for specific defense-related activities. This justification is essential for decision-makers in Congress who assess and approve the DoD budget.

Transparency: J-Books are designed to enhance transparency in the budgeting process. They help the public, Congress, and other stakeholders understand the DoD's financial needs, the objectives of various programs, and how to utilize resources.

Accountability: They hold the DoD accountable for how it spends taxpayer dollars. By detailing each program's expected outcomes and benefits, J-Books allows Congress to assess whether the proposed expenditures align with national defense priorities.

Program Evaluation: These documents with other associated materials (e.g., National Defense Authorization Acts [NDAAs], SARs, etc.) help evaluate the performance and effectiveness of DoD programs. They often include metrics and performance measures, allowing Congress to gauge the success of these programs.

Program Planning: J-Books assist in program planning by outlining goals, objectives, and expected milestones for various DoD initiatives. This information is critical for managing and tracking the progress of these programs.

Congressional Approval: J-Books play a crucial role in the congressional approval process. They are used by Congress to make informed decisions regarding the allocation of defense funds and to ensure that the DoD budget aligns with national defense and security priorities.

The next section provides a problem statement which is followed by our methodology and approach. This section is followed by our results section, and conclusions section. Our paper finishes with suggestions for future research.

Problem Statement and Our Solution Concept

A comprehensive understanding of the complexities and interdependencies of DoD programs, viewed through J-Books, ultimately advances national defense capabilities, and helps in formulating sound budgetary strategies. However, understanding the budgetary implications of various DoD acquisition programs through J-Books is daunting, especially for complex cross-domain strategies like Joint All Domain Command and Control (JADC2). JADC2, which is woven throughout the different branches of the DoD, is challenging. Tools present within the DoD's Advana[1] environment (e.g., GAMECHANGER[2]), and potentially specialized tools built external to the DoD can support human-based analysis. While Advana's capabilities are ever expanding and quite useful, we wanted to completely reconsider how analysis of budgets could be achieved via advanced presentation of the data using modern dashboards with a natural language processing (NLP) and generative artificial intelligence (AI) interface. Generative AI or large language models (LLMs) were just becoming available as we began this research and thus, we envisioned an LLM interface to support queries of documents (Star Trek™ here we come! [Paramount Pictures, 1986]).

However, first, we provide our concept of a Portfolio Analysis Dashboard using Tactical Aircraft in Figure 1.[3]

[1] Advana is the DoD's enterprise-wide, multi-domain data, analytics, and artificial intelligence (AI) platform that provides military and civilian decision makers, analysts, and builders with unprecedented access to enterprise tools and capabilities—all in a scalable, reliable, and secure environment.
[2] For additional information on GAMECHANGER, see DIA Public Affairs (2022).
[3] Note that OpenAI's ChatGPT and DALL-E terms of service can be found here, https://openai.com/policies/terms-of-use, with their publication policy found here, https://openai.com/policies/sharing-publication-policy. Stock language provided by OpenAI is amended to be the following: *The authors generated this content in part with GPT-3.5 and GPT-4, OpenAI's large-scale language-generation model. Upon generating draft language and DALL-E concept art, the authors reviewed, edited, and revised the language to their own liking and take ultimate responsibility for the content of this publication.*

Figure 1. Conceptual Portfolio Analysis Dashboard Using Notional Tactical Aircraft (TacAir)

The primary analyst workstation would allow menu access to various portfolios managed by the DoD (top dual monitor layout in Figure 1 with notional analytical charts and graphs). The arrow depicts a conceptual roller-wheel mouse interface to allow the analyst to select among the various aircraft managed for click-through deep-dives. An alternative concept is the overall Tactical Aircraft (TacAir) chart on the bottom-left ranking the aircraft by funding in the rows and the Fiscal Year (FY) in the columns with a rolled-up portfolio funding view at the bottom.

The following list describes additional monitor-specific analytical concepts:

Left Monitor:

1. **Newly Proposed Spending Analysis:**
   - **Visual Representation:** Charts and graphs displaying the newly proposed spending.
   - **Details:** Data shows service system contributions, e.g., Air Force, Army, Navy, and Space Force.
   - **Interactivity:** Users can hover over sections of the charts to view exact figures and percentages.
2. **Lines of Effort Details:**
   - **Drill-Down Section:** Interactive selection that allows users to explore various lines of effort.
   - **Information Display:** Provides project names, services involved, and allocated budgets.
   - **User Interaction:** (L/R) clicking on a graph reveals additional information and historical data.

3. **LLM Interface:**
    - **Summarization:** An LLM interface support for requesting a quick summarization of the observed trends.
    - **New Queries:** Interface to the back-end database to create normal language data queries to populate selected charts and figures.

Right Monitor:

1. **Past Budget Submissions:**
    - **Tabular Format:** Information on past budget submissions allowing FY-after-FY comparisons.
    - **Graphical Data:** Line graph with trends in budget submissions over time.
    - **Accessibility:** Users can sort and filter this data based on different parameters.
2. **Latest (Real-Time/Monthly/Weekly) Cost/Schedule Performance Data:**
    - **Visual Indicators:** Latest data from the EVMS, cost, schedule, and performance risks.
    - **Color-Coded Alerts:** Risks are indicated using color-coded alerts, providing an immediate sense of any issues. (Further drill-down with the latest SPO risk management updates for details.)
    - **Detailed Metrics:** Information on cost overruns, schedule delays, and other metrics.
3. **User Interface and Navigation:**
    - **Navigation Panel:** Monitors feature intuitive navigation panels for easy access to different sections.
    - **Customization:** Users can customize the views and prioritize the information they find relevant.

While we just show the analytical capability concept in this example, one can also envision an LLM enabled J-Books creation capability where the LLM has been trained in the area of application or can simply be used in support of creating more succinct and easier to understand digital data constructs.

Methodology/Approach

Figure 2 shows the overall proof-of-concept flow for our proposed approach. We will discuss in this section the approach taken to implement this analysis flow.

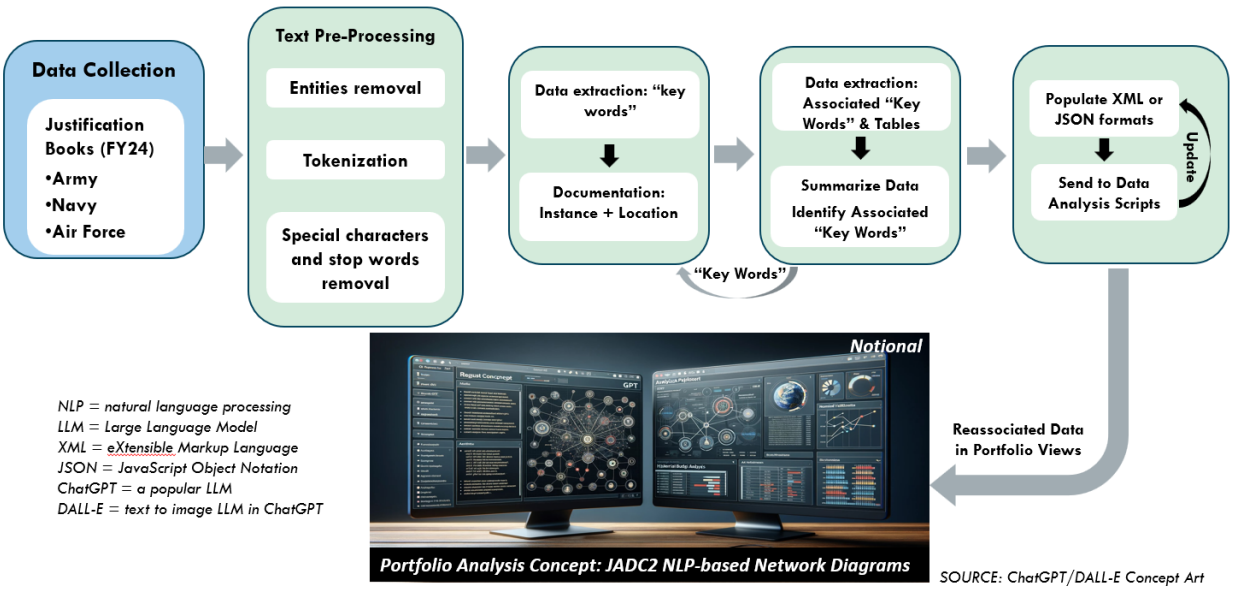

Figure 2. Proof-of-Concept Portfolio Analysis Flow Using JADC2 Keyword Search of FY24 J-Books

Our proposed approach provides an initial proof-of-concept for accumulating programs into "portfolio-like" groupings, leading to a clearer representation of budgetary implications across DoD programs. It is important to note that the information accumulated using this NLP approach on J-Books does not result in a true portfolio (as there is information that will be lacking) but is a step towards restructuring the existing content as elements in a portfolio.[4] For this reason, the approach using JADC2 demonstrates a "portfolio-like" representation of these budget materials.

Hence, we believe it will help in the following areas:

Inter-Service Synergy: What are the fundamental mechanisms or strategies to enhance synergy and coordination among the Army, Navy, and Air Force regarding related programs to maximize their effectiveness and interoperability?

Technological Integration: Technology integration should be considered a core component of J-Books analysis. It would clarify how emerging technologies are incorporated into different programs and how they intersect across the various branches.

Resource Allocation: How do allocating resources and budgetary decisions within each branch affect the development and implementation of related programs? What are the trade-offs and challenges associated with resource allocation? How have resource allocation and budget requests changed over time? Can we correlate these temporal fluctuations with risk management behaviors of DoD decision makers?

Operational Implications: Within the J-Books framework, it's essential to understand the functional implications of programs like JADC2 on the specific missions and functions of the Army, Navy, and Air Force. Moreover, it delves into how these branches adapt to accommodate changes to ensure the success of their missions.

[4] A true portfolio would include numerous other elements that are not present in the publicly available J-Books, such as resource allocation, strategic value, risk assessments, and mitigation costs, among others.

This project aims to eventually leverage the promise of NLP techniques to accumulate data associated with programs (or instances) and strategies within J-Books. Moreover, by harnessing the power of NLP and generative AI, the project aims to enhance the understanding of budgetary implications across J-Books.

The overall objectives of this project are described in two phases:

Phase 1. Demonstrate NLP J-Books "portfolio-like" accumulation: After identifying J-Book sections associated with keywords (acronyms) through NLP pattern matching, we plan to associate other materials (e.g., NDAAs, SARs, etc.), thus demonstrating "portfolio-like" budgets and displaying their context in a network diagram.[5]

Phase 2. Add analytics: The project then prototyped an analytical front-end of the information for algorithmic approaches to facilitate a deep understanding of the portfolio's trends. The prototype includes an OpenAI LLM front-end to provide responses to chat-like questions regarding the data and text, with visual displays.

Method Used to Demonstrate J-Books Key Word Extraction

The efficient processing of large volumes of text-based documents presents a significant contemporary challenge in information management. This section introduces an inventive framework designed to simplify content extraction and the identification of content placement within documents. These processes hold immense importance across diverse domains, including academic research and corporate decision-making, as they facilitate rapid and efficient information retrieval while preparing text for decision support.

Prior to following the process as depicted in Figure 2, the researchers utilized Adobe Acrobat in a pre-processing step to apply optical character recognition (OCR) to convert each image in these documents to provide extractable data. OCR can introduce errors into this analytical process (see for example, Lamba and Madhusudhan [2023]). Requiring the Services to not use images in their budget justification documents was our first recommendation to avoid this timely and potentially error prone step.

The framework depicted in Figure 2 contains two key phases: Phase 1, which involves Text Extraction and Summarization, and Phase 2, which focuses on Analytics Extraction.

Phase 1 Text Extraction and Summarization

The framework (see Figure 3) that we implemented in our initial Phase 1 script offered a rudimentary but versatile initial toolset for keyword identification, instance location, and associated content extraction.

[5] Network diagrams are used as a graphical depiction of the data. See for example Crane et al. for some of the benefits from this kind of analysis.

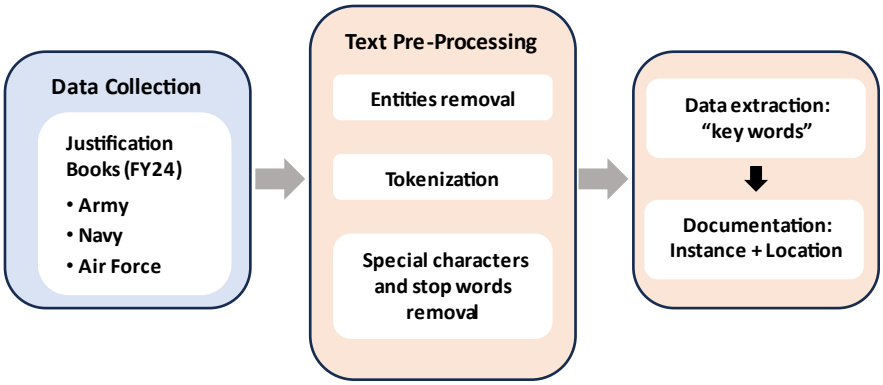

Figure 3. Instance Identification Framework For Phase 1

By automating these processes, the framework enhances the efficiency and effectiveness of information management, allowing users to quickly obtain keyword context and insights from the J-Books.

Phase 2 Prototyped Analytics Front-End

Phase 2 of the framework was intended to provide a prototype of a versatile toolkit for developing analytics, data, text, and visuals. These processes contribute significantly to information organization, retrieval, and insight generation. Figure 4 provides a graphical view of the overall demonstrated capability.

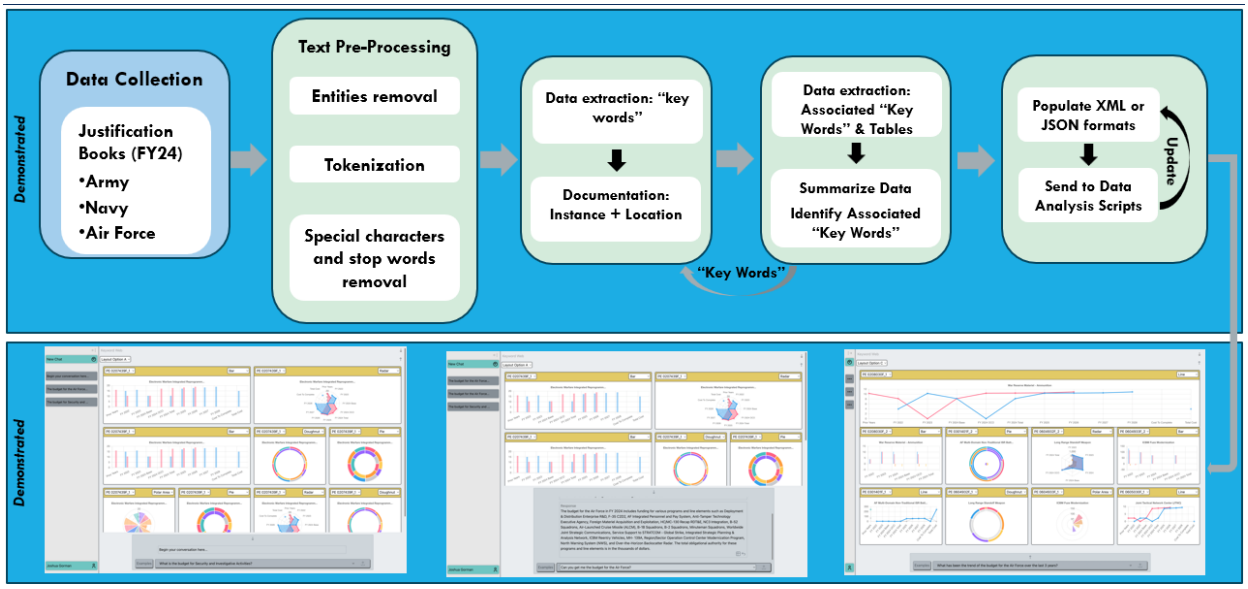

Figure 4. Completed Prototype Work from Phase 1 and Phase 2

Shown in Figure 4 are the completed/demonstrated aspects of the prototype. The top blue box in the figure demonstrates scripts that can extract key words (Phase 1). We were then able to send those results into ChatGPT to extract additional keywords to search for, illustrated as the "Key Words" feedback loop in the top-row of the figure. The output was processed to extract tables from the PDF J-Books files, and successfully associated in the feedback loop of key word search results to create an associated network diagram (Figure 5). We also successfully populated JSON files using the extracted tables from these search results and subsequently used those JSON files in a prototyped front-end (the bottom blue box in Figure 4).

The prototyped front-end (the bottom half of Figure 4) includes a ChatGPT interface to query a back-end database server containing the J-Books extracted JSON information. The current implementation only includes FY24 Air Force RDT&E volumes I, II, IIIa, and IIIb, but could easily be extended to include all the J-Books' contents. Further work is required to fully integrate the NLP processing in the top half of Figure 4 with the front-end and back-end implementation with the ChatGPT/LLM interface shown in the bottom half. Figure 5 is a screengrab of the network diagram results of our JADC2 keyword association. Future iterations of this diagram will include symbols, line type variations and colors to provide visual meta-information regarding the association.

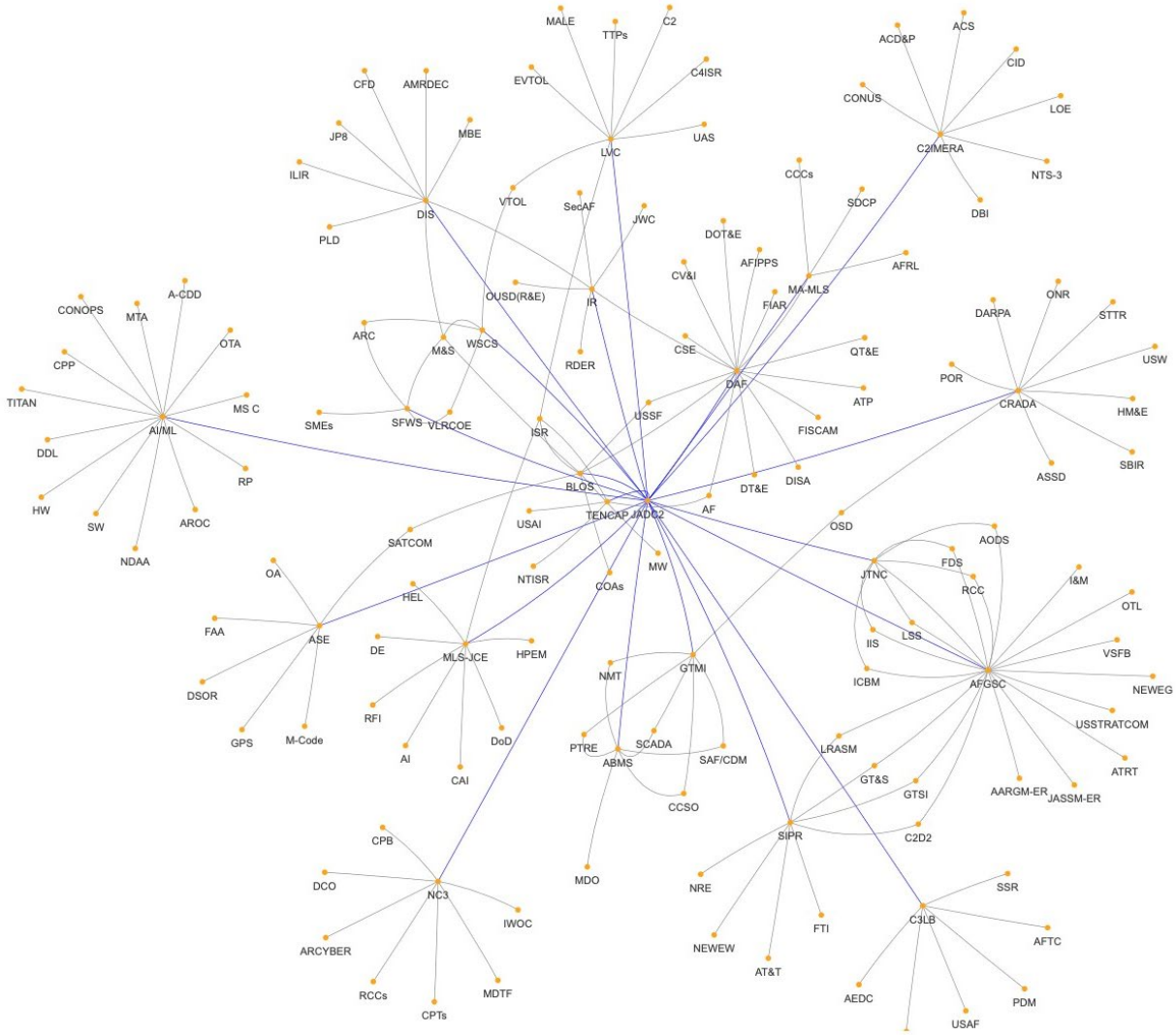

Figure5. Network Diagram Results from Key Word Processing FY24 J-Books for "JADC2" and Associated Acronyms

Note that we made no attempt to identify all the acronyms found by the feedback loop for this proof-of-concept report (i.e., all these acronyms will not be found in the acronym definition section of our final reports [Buettner et al., forthcoming; Ramirez-Marquez et al., forthcoming]). Future functionality should define acronym categories and use other contextual syntax-based "key words" for correlating concepts and terms for association. Figure 6 provides a larger

screengrab demonstrating a changed layout with new tables feeding into the graphics based on the most recent LLM query with its response.

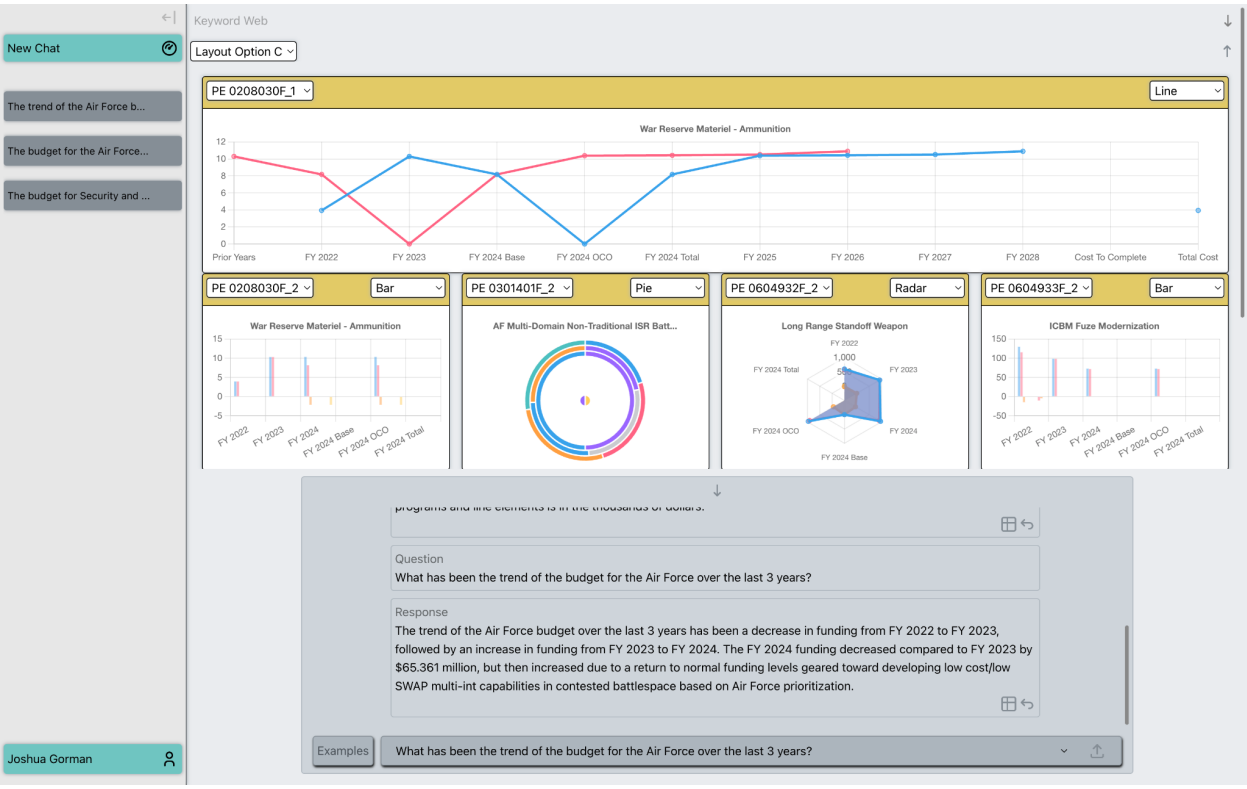

Figure6. Screengrab Demonstrating a Changed Layout with New Tables Based on the Recent LLM Query with the Response

Finally, our Phase 2 implementation was provided directly to the PPBE Commissioners with a video of the primary capabilities implemented by the end of February 2024 and documented in our final reports (Buettner et al., forthcoming; Ramirez-Marquez et al., forthcoming). Table 1 documents the final functionality that was included in our dashboard prototype.

Table 1. Completed Dashboard Prototype Functionality

User Story Name	Story Acceptance
As a User, I want to log in to the dashboard.	The dashboard should have a secure login system. Users should be authenticated against a user database.
As a User, I want to query the database using LLM.	There should be a text input box for users to input their queries. The language model should process the query.
As a User, I want to submit my query and receive relevant data.	Upon submitting the query, the dashboard should send a request to the language model. The language model should interpret the query and generate a query. The database should return relevant data based on the query.
As a User, I want to see the queried data displayed in a table format on the dashboard.	The dashboard should dynamically generate a table to display the queried data. The table should include relevant columns and rows based on the query.

User Story Name	Story Acceptance
As a User, I want to save my queries for future use.	The dashboard should have a feature to save user queries. Users should be able to retrieve and reuse saved queries.
As a User, I want to visualize the queried data with charts and graphs.	The dashboard should provide options to generate visualizations (e.g., bar charts, pie charts) based on the queried data. Users should be able to customize and interact with the visualizations.
As a User, I want to be able to upload new tables to the database.	A portal to upload CSV/XML/JSON files to database. All files if not in JSON need to be converted to JSON.
As a User, I want to be able to use an LLM for my queries and data display.	A User can use an LLM to query the database and plot data.
As a user of the system, I need it to accept an extracted pdf table and populate XML or JSON format.	To achieve this goal, I have written a code to ask the user for a JBOOK and program element. This code then extracts the specific table.
As a User I want to convert tables to XML or JSON format.	The user can convert a PDF table into either XML or JSON.
As a User I want some examples to choose from/general overview.	Create a way to save “templates” which have built in tables and layout. Should be able to choose from a dropdown a “query” sentence which has the desired results. Make sure this is new user friendly, aka intuitive.
As a User I would like to search for additional keywords related to the original search.	
As a User I want to visualize the network of associated keywords.	
Integrate network diagram.	Look for ways to include user options

Conclusions

Trying to use an LLM on a much larger document did not provide satisfactory results, as could be seen in Appendix A of Buettner et al. (forthcoming). Appendix B in that same reference demonstrated the variability of an LLM which could be expected to provide without fine-tuning it to this domain. To support this claim, we provided a quote from an article that claimed this could be overcome.

> *By fine-tuning the model on specific domains or topics, users can enhance the relevance and uniqueness of the generated responses. Customizing ChatGPT’s behavior by providing explicit instructions or preferences also helps in reducing repetition and tailoring the output to specific requirements.*

Despite our overall lack of satisfactory results in the larger analytical context, our results section demonstrated an implementable alternative to obtain satisfactory short summaries of

sub-sections that were extracted from the much larger document.[6] Hence, we postulate that this proof-of-concept successfully demonstrated a satisfactory LLM summarization of 19 pages from a DoD J-Book, as the existing formatting provided by the USAF easily allows automatic sub-sectioning[7] of PEs for analysis by an LLM. Hence, we integrated an LLM into our prototype dashboard and successfully had it analyze sections of data from these same AF J-Books using our database implementation.

Importantly, our research can be used as a first step to build a capability to associate programs and their elements into portfolios and demonstrated an implementation of our Phase 2 concept. We felt our concept should provide the capability of reorganizing existing documents into a portfolio budget view. Once that reorganization has occurred, we could have assessed what data is missing for this view to be considered a portfolio budget for PEO management as envisioned by the 809 Panel. This research provided a foundation for refactoring budget materials into a more user-friendly product with links to various parts of the larger portfolio.

Ultimately, this research empowers the DoD and Congress to confidently invest in alternative approaches to understand the complexities and interdependencies of our large DoD programs, how they can begin to associate programs into portfolios, and contribute to a more comprehensive understanding. Such insights enhance the capabilities and effectiveness of military programs and inform more efficient resource allocation, bolstering the overall decision-making processes. This research can significantly influence program development and implementation, ultimately advancing national defense and military strategies.

Initially, we felt a follow-on effort should attempt to characterize the various ChatGPT PDF reader plugins, however, recent updates to ChatGPT-4 have phased these plugins out. Other LLM technologies should be investigated to determine if they can be appropriately trained/tuned to provide summarizations that directly align with congressional staff needs. Further, using a specifically trained LLM in a "common data analytics platform" (PPBE Reform Commission, 2024) could also be a future research step.

Our research team recommended:

- **Funding further research into LLM support of budgetary analysis:** Noting that Advana[8] has implemented the "GAMECHANGER"[9] capability, and a blog indicates that at least one tool vendor may have resolved the issues noted in this report[10], further research should assess various LLMs and confirm that an LLM can be trained on the DoD's corpus of data. The goal is to confirm that a ChatGPT text-based query interface can reliably support and enhance analysts with their tasks.
- **Funding further efforts to reorganize budget documents**: Providing portfolio-like budget views from AI/ML reorganizations without having to make drastic changes to the existing documentation format used by the various services would simply add a processing step to the existing delivery flow. Once completed, these results can support collaborative decisions on what changes and additions should be promulgated to the

[6] Recent changes to ChatGPT4 have made it easier to analyze large documents in addition to phasing out the use of plugins.
[7] Looking through the larger document, one sees that the PEs are provided in sections that our pattern matching algorithm can use to extract the entire section that the key word search can help identify.
[8] Advana is the DoD's enterprise-wide, multi-domain data, analytics, and AI platform that provides military and civilian decision makers, analysts, and builders with unprecedented access to enterprise tools and capabilities—all in a scalable, reliable, and secure environment.
[9] For additional information on GAMECHANGER, see DIA Public Affairs (2022).
[10] An Associate Professor of Computer Science, from Carnegie Mellon University blog (Neubig, 2023), indicates that many of the issues identified here have been resolved by at least one LLM vendor.

services to fully enact a portfolio management approach that includes a portfolio budget view during the accumulation of the data.

Further Research and Capabilities

To address the questions specified in the Problem Statement section, future work on this research would address the following:

Integrate the NLP Python and Front-End: The proof-of-concept text extraction script leveraged well-formatted J-Books to extract important contextual information, such as cost data, and used other acronyms found in these sections to identify connections. We would demonstrate additional concepts to integrate the NLP scripting approach into the implemented front-end to populate the back-end database for additional analytical capabilities.

Expand Front-End and Back-End Functionality: The existing capabilities would be expanded to provide additional useful analytical processing.

Extend Processed Capabilities: Include processed J-Books across fiscal year boundaries with the addition of other resources, such as including the National Defense Authorization Act (NDAA) in the files processed.

Full-Context Extraction and Portfolio Reformatting: Expand this proof-of-concept text extraction script to leverage these well-formatted J-Books to extract other important contextual information, such as cost data. We would then demonstrate restricting the data into representative portfolios.

Inter-Service Synergy: Implement semantic network analysis to identify critical nodes and their relationships.[11] Analyze links between these nodes based on their proximity within the text.

Apply community detection algorithms like the Louvain or Leiden algorithms[12] to identify clusters of related content. This approach can reveal the synergy and coordination among the Army, Navy, and Air Force regarding associated programs. Figure 7 illustrates this type of analysis for a project to identify skills within the acquisition community.

[11] Regarding a definition of critical nodes, Lalou et al. (2018) defines them as “those the deletion of which disconnects the network according to some predefined connectivity metrics, such as: maximizing the number of connected components, minimizing pairwise connectivity in the network, minimizing the largest component size, etc.”

[12] The benefits of the Leiden cluster identification algorithm are found in Tragg et al. (2019). The researchers have made their code available on a GitHub site.

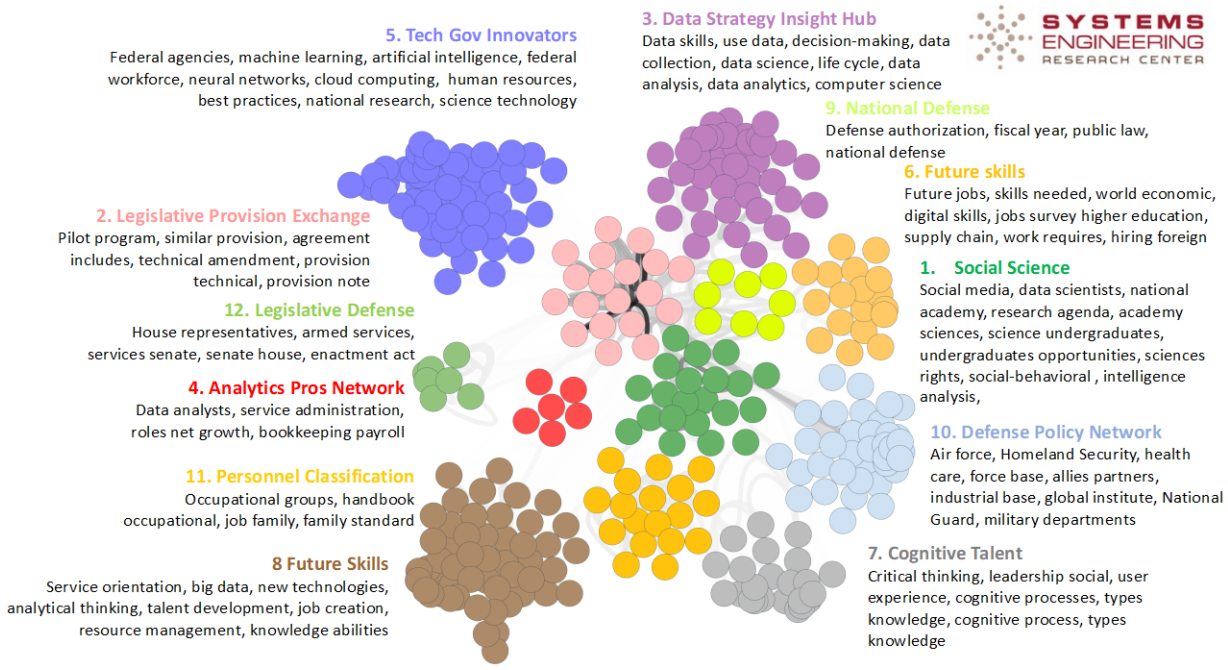

Figure 7. Network Analytics Description Based on Text Extraction

Technological Integration: Analyze text data to identify trends (trend analysis) and patterns in technological integration within the Army, Navy, and Air Force programs through time and programs. Investigate the adoption of emerging technologies across different programs and how these technologies intersect across various branches.

Resource Allocation: Develop models to simulate and analyze resource allocation scenarios within each branch and assess their impact on program development and implementation. Investigate the trade-offs and challenges of different resource allocation strategies.

Operational Implications: Analyze the text data to understand the operational implications of programs like JADC2 on the specific missions and functions of the Army, Navy, and Air Force. Investigate how these branches adapt to accommodate changes and ensure the success of their missions.

Temporal Association and Differencing: Analyze the data across multiple years to identify trends in changes to the budget values and an ability to analyze the specific syntactic meaning of the wording associated with key areas from year to year.

References

Buettner, D. J., Ramirez-Marquez, J. E., & Anton, P. S. (forthcoming). *ChatGPT budget justification book program element summarization: A proof-of-concept demonstration.* Stevens Institute of Technology, Acquisition Innovation Research Center.

DIA Public Affairs. (2022, Feb. 7). *Gamechanger: Where policy meets AI*. Defense Intelligence Agency. https://www.dia.mil/News-Features/Articles/Article-View/Article/2926343/gamechanger-where-policy-meets-ai/

Lalou, M., Tahraoui, M. A., & Kheddouci, H. (2018). The critical node detection problem in networks: A survey. *Computer Science Review*, 92–117.

Neubig, G. (2023). *C3 generative AI: Getting the most out of enterprise data*. C3.ai. https://c3.ai/c3-generative-ai-getting-the-most-out-of-enterprise-data/

Ramirez-Marquez, J. E., Buettner, D. J., Anton, P. S., Gorman, J., & Patel, A. (forthcoming). *DoD budget justification book keyword search association: A proof-of-concept demonstration.* Stevens Institute of Technology, Acquisition Innovation Research Center.

Tragg, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 1–12.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net