ARL-TR-10409 • Aug 2026

DEVCOM ARMY RESEARCH LABORATORY

Aegis Scholar: A Lightweight Network-Centric Tool for Defense Expertise Discovery

by Mark A. Tschopp, Ethan Shafer, Steve McCoy, Adriana Canedo, Markuz Robinson, Jonathan Hale, and James R. Uplinger

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-10409 ● AUG 2026

Aegis Scholar: A Lightweight Network-Centric Tool for Defense Expertise Discovery

by Mark A. Tschopp and James R. Uplinger
DEVCOM Army Research Laboratory

Ethan Shafer, Steve McCoy, Adriana Canedo, Markuz Robinson, and Jonathan Hale
Army Artificial Intelligence Integration Center (AI2C)

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
August 2026	Technical Report	November 2025	May 2026

4. TITLE AND SUBTITLE

Aegis Scholar: A Lightweight Network-Centric Tool for Defense Expertise Discovery

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)

Mark A. Tschopp, Ethan Shafer, Steve McCoy, Adriana Canedo, Markuz Robinson, Jonathan Hale, and James R. Uplinger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

DEVCOM Army Research Laboratory
2800 Powder Mill Rd., Adelphi, MD 20783-1138
Army Artificial Intelligence Integration Center

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-10409

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

ORCID: Mark A. Tschopp, 0000-0001-8471-5035; James R. Uplinger, 0000-0002-0146-2130.

14. ABSTRACT

This report documents the design, implementation, and evaluation of Aegis Scholar, a prototype expert-discovery platform developed under the AI Technicians program at the Army Artificial Intelligence Integration Center. Addressing the challenge of identifying subject-matter experts across distributed Department of War organizations, Aegis Scholar acts as a "connection engine" by integrating semantic vector search (Milvus) with relational graph mapping (Neo4j). By ingesting metadata from the Defense Technical Information Center, the system represents authors via abstract-derived embeddings and ranks them against natural language queries using a composite relevance heuristic. Offline evaluations demonstrated the efficacy of limited-context transformer models for semantic matching, while a task-based user study showed that the targeted, network-centric interface reduced search times by up to 50% for defense-specific queries compared with general-purpose tools. The project successfully demonstrates the operational value of minimalist, visual discovery interfaces for defense applications, while simultaneously serving as a robust process development vehicle for mid-career Soldiers executing end-to-end data science projects under ARL mentorship.

15. SUBJECT TERMS

Aegis Scholar, expertise discovery, information science, artificial intelligence, semantic search, graph databases, Military Information Sciences

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	35

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Mark A. Tschopp	520-691-6297

STANDARD FORM 298 (REV. 5/2020)

Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation and Problem Statement

Large, complex organizations depend on people as much as on data. Research reports, technical notes, and program documents accumulate over time, but the ability to identify *who* has the right expertise for a new problem often lags behind the growth of that information. Personnel tend to operate inside local networks; beyond those networks, it becomes difficult to know which experts exist, where they sit in the organization, and how to reach them.

The Department of War (DoW) faces this challenge across a broad, distributed research and development enterprise. Existing information systems—bibliographic databases, internal repositories, and collaboration platforms—are effective at surfacing documents and keywords, but they offer limited support for moving from topics to people and organizational context. This is particularly acute in environments where much of the most relevant work resides in controlled-distribution or classified reports that are not indexed by public tools.

The Aegis Scholar project explores a complementary idea: a *connection engine* that helps users move from mission-relevant queries to subject-matter experts (SMEs), their collaborators, and their organizational affiliations. Rather than focusing only on document retrieval, the system is designed to highlight the people and relationships behind the work. The name "Aegis Scholar" evokes the aegis as a protective shield and guiding symbol, reflecting the goal of helping the DoW identify and connect with relevant expertise more effectively.

1.2 Objectives and Proposed Work

This report documents a prototype system and associated methods developed under the AI Technicians program at the Army Artificial Intelligence Integration Center (AI2C). Our team combined mid-career Soldiers participating in the AI Technicians program with mentors from the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL).

The work pursued two primary objectives:

1) **Technical objective:** Design, implement, and evaluate a graph- and embedding-based search system for defense-relevant research metadata. The system rep-

resents authors and organizations in a hybrid vector–graph space and ranks candidate experts for natural-language queries about technical topics or mission scenarios.

2) **Process objective:** Provide AI Technicians participants with an end-to-end experience in framing a mission-inspired problem, translating it into technical requirements, and delivering a working prototype using contemporary tools (data pipelines, vector and graph databases, containerized services, and basic experimental evaluation).

ARL mentors (James R. Uplinger and Mark A. Tschopp) formulated the initial project concept and proposal. The full capstone team then met regularly throughout the academic term to clarify requirements, align technical decisions with defense-relevant use cases, and ensure that the work remained grounded in realistic operational constraints. Faculty from Carnegie Mellon University (CMU) provided course structure and academic oversight, while the AI2C Soldiers (Ethan Shafer, Steve McCoy, Adriana Canedo, Markuz Robinson, and Jonathan Hale) executed the day-to-day design, implementation, and experimentation that produced the Aegis Scholar prototype.

As a technical record, this report makes three contributions: it documents an author-centric defense expertise-discovery prototype; captures the principal data, ranking, and interface decisions made during the capstone; and reports pilot evidence from both offline model comparison and a small task-based user study.

1.3 Scope

The scope of this work is intentionally bounded. Aegis Scholar focuses on the following:

- Ingesting and standardizing unclassified, defense-relevant research metadata, with the evaluated prototype built around a DTIC-derived snapshot rather than a complete enterprise corpus;
- Constructing a hybrid data store that combines a vector database for semantic similarity search with a graph database for authors, works, organizations, and topics;
- Representing authors through embeddings derived from the abstracts of their works and ranking candidate experts for user queries using a hybrid scoring

function; and

- Evaluating the resulting system through offline ranking experiments and a small pilot user study (n = 5) comparing Aegis Scholar to a representative public research discovery workflow.

The project does not attempt to integrate classified holdings or repositories with controlled distribution. It also does not address enterprise deployment, authority to operate, or full integration with existing DoW platforms. In its evaluated form, the prototype emphasized author search and network exploration; broader organization-, topic-, and work-level search paths remained incomplete at the end of the capstone. Instead, the prototype uses publicly available and defense-focused open data to explore patterns, architectures, and evaluation methods that could inform future systems operating over sensitive or internal document sets.

1.4 Report Organization

The remainder of this report is organized as follows. Section 2, the background section, summarizes related discovery tools and the program context for the project. Section 3, the methods section, describes the Aegis Scholar system architecture, data pipeline, modeling approach, experimental design, and deployment environment. Section 4, the results section, presents the offline ranking experiments and the user study. Section 5, the discussion section, covers the implications of these results, technical limitations, relationships to existing tooling, and process outcomes. Section 6 concludes with key findings and outlines directions for future work.

2. Background

2.1 Existing Tools for Expert and Literature Discovery

The modern scholarly landscape is supported by a robust ecosystem of discovery platforms and metadata registries. Tools such as OpenAlex[*] and Crossref[†] provide comprehensive catalogs of global research outputs, authors, institutions, and topics.[1,2] Similarly, the Dimensions platform[‡] offers deep bibliographic and bibliometric data, including versions specifically tailored to defense and government datasets.[3] These systems are widely used for literature discovery, citation tracking,

[*] https://openalex.org

[†] https://www.crossref.org

[‡] Public version: https://app.dimensions.ai. A password-controlled federal version is also available at https://federal-collection.dimensions.ai.

and institutional benchmarking in bibliometrics and research analytics.[4]

However, many of these platforms are primarily optimized for academic or open-source environments. For large organizations like the DoW, research often spans both public and sensitive or controlled-distribution channels. In these contexts, researchers need more than just access to documents; they require actionable paths to identify and contact SMEs within the government and its partnered academic and private-sector networks. That distinction also shaped the evaluation strategy in this report: OpenAlex served as a practical public comparator for user tasks, but it did not solve the narrower problem of defense-oriented expert discovery on its own. While existing tools provide the foundational data, comparative studies also highlight systematic differences in coverage and metadata quality across fields and regions.[4] The team therefore identified a gap for specialized, lightweight systems that map this bibliographic metadata directly to defense-relevant organizational structures and collaborative networks.

2.2 Program and Capstone Context

This work was conducted as part of the AI Technicians program at AI2C in connection with CMU. The program is designed to build foundational artificial intelligence and data engineering capabilities within the force, preparing Soldiers to integrate modern AI tools into their day-to-day operations at their units.

As a capstone project for the program, ARL proposed a mission-inspired challenge: using publicly available data to develop models that quickly identify potential research partners and summarize their relevant focus areas. The project was structured across four phases over the course of an academic term, moving from initial vision and requirements definition, through data schema and modeling foundations, into integration and testing, and concluding with user experimentation.

This phased approach structured the technical development of the prototype while keeping the project aligned with the educational goals of the capstone. It allowed Soldiers to gain hands-on experience in building an end-to-end AI system—from data ingestion and database configuration to model evaluation and user interface design.

2.3 Conceptual Framing: From Knowledge to Connection Engines

Traditional discovery systems operate primarily as *knowledge engines*: a user inputs a query and receives a ranked list of relevant documents. While effective for literature reviews, this model forces the user to manually extract the social and organizational context from the results. Who are the leading experts? What organizations do they belong to? Who do they collaborate with?

Aegis Scholar is conceptually framed as a *connection engine*. By representing research metadata as a highly interconnected graph rather than a flat list of records, the system elevates authors and organizations to primary search results. When a user queries a technical topic, the system is designed to return not just what has been written, but a networked view of the people driving that research. This framing shifts the focus of discovery from finding information to building relationships and facilitating cross-functional collaboration.

3. Methods

3.1 System Architecture and Implementation

Aegis Scholar was developed as a cloud-native prototype consisting of interconnected microservices. The complete codebase, including the infrastructure-as-code and container configurations, is maintained in an open-source GitHub repository.[5] In the evaluated prototype, the primary completed user-facing path was author search and author-network exploration rather than a fully implemented multi-entity search platform.

The system relies on a central API gateway that orchestrates queries across three specialized backends:

1) A **vector database** (Milvus) that stores high-dimensional semantic embeddings of abstracts and queries.
2) A **graph database** (Neo4j) that maps the relational structures between Authors, Works, Organizations, and Topics.
3) A prototype **identity service** (OpenLDAP) for basic directory-style lookups.

Figure 1 summarizes this architecture by showing how data flows from external sites through scrapers into blob storage, through a transformation worker into the vector and graph services, and finally into the Aegis Scholar API and client inter-

faces. In the diagram, individual boxes represent modules, grouped boxes indicate aggregated modules, and the elements labeled as actors represent external sites, client interfaces, and other systems that interact with the prototype.

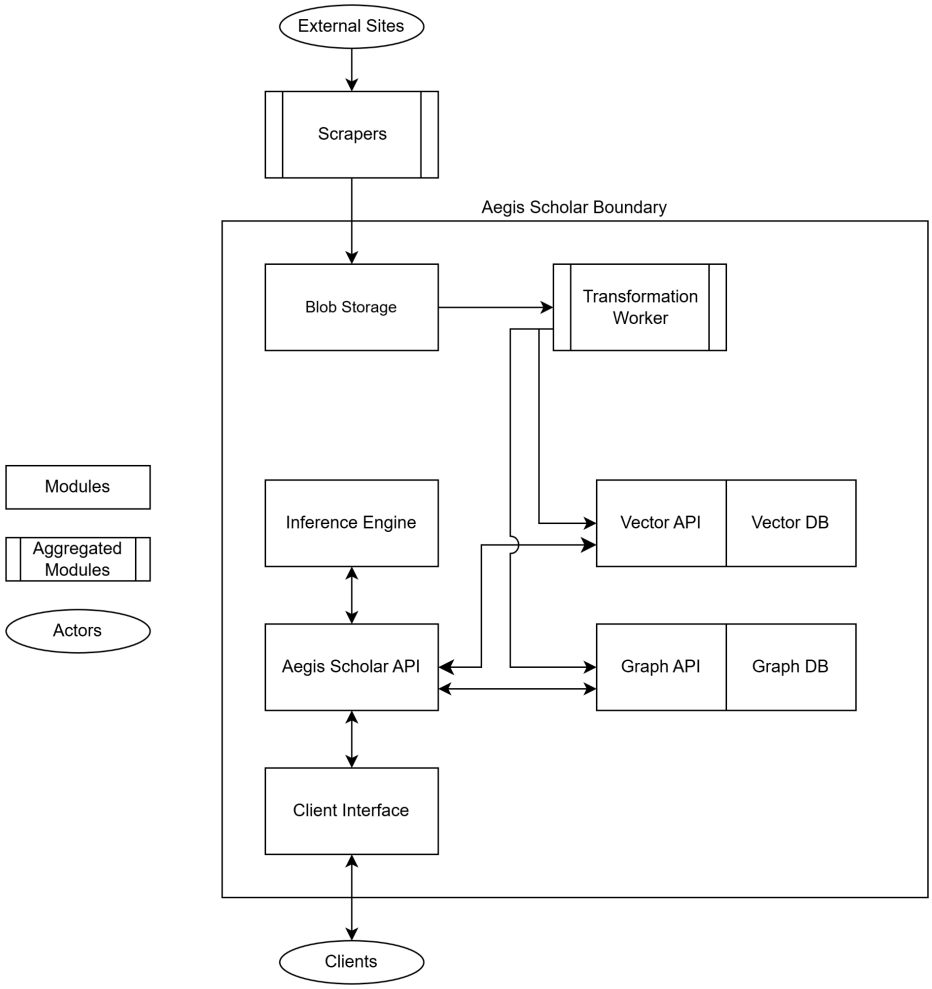

Figure 1. High-level architecture of the Aegis Scholar connection engine. Natural language queries are processed by the API gateway, which routes semantic searches to the Milvus vector database and relational traversal requests to the Neo4j graph database. The identity service provides personnel resolution.

When a user submits a natural language query, the API gateway queries the vector database for semantically similar authors, queries the graph database for the collaborative network surrounding those authors, and returns a unified, ranked list of candidate experts. Further details regarding the containerization strategy, Kubernetes deployment artifacts, and continuous integration pipelines are provided in Appendix A; these materials describe the prototype implementation and its intended deployment path rather than a fielded enterprise service.

3.2 Data Sources and Pipeline

In the initial phases of the project, we evaluated three major scholarly metadata sources: OpenAlex, Crossref, and DTIC. While OpenAlex and Crossref provided massive, global academic datasets, profiling revealed that critical fields linking authors to specific DoW affiliations were frequently sparse or missing.

Consequently, we pivoted to ingest data exclusively from DTIC for the prototype. This was a deliberate engineering decision based on data quality and mission relevance. Relying on DTIC inherently scopes the resulting network to individuals operating within the defense ecosystem: government authors publishing technical reports, academic and industry partners executing defense-funded research, and their direct co-authors. For the purposes of this capstone, however, the evaluated system was built on a DTIC-derived snapshot rather than a complete representation of all potentially relevant defense research activity.

The data pipeline utilizes a batch Extract, Transform, Load (ETL) process. Source records are extracted via web scraping because, at the time of this project, DTIC did not offer a bulk REST API. They are then normalized using consistent Universally Unique Identifiers (UUIDs), and loaded into the graph and vector databases without creating duplicate entities. Because the upstream source was accessed through scraping rather than a bulk API, dataset completeness was constrained by source behavior and project time limits. The specific field selections and Neo4j node schemas derived from this process are detailed in Appendix B. Figure 2 illustrates the unified data schema used for the prototype.

3.3 Modeling Approach and Relevance Scoring

To support semantic discovery, authors are represented not by a list of keywords, but by a unified vector embedding. The system passes the abstract of each work through a pre-trained sentence transformer to generate a dense vector. For a given author, the vectors of all their associated works are averaged element-wise and normalized, yielding a single embedding that captures their dominant research themes. In a preliminary analysis of OpenAlex data, approximately 92% of sampled authors published in fewer than five distinct topics, supporting the use of a single embedding per author as a reasonable approximation for this prototype.

Related work on academic expert finding has shown that well-constructed textual

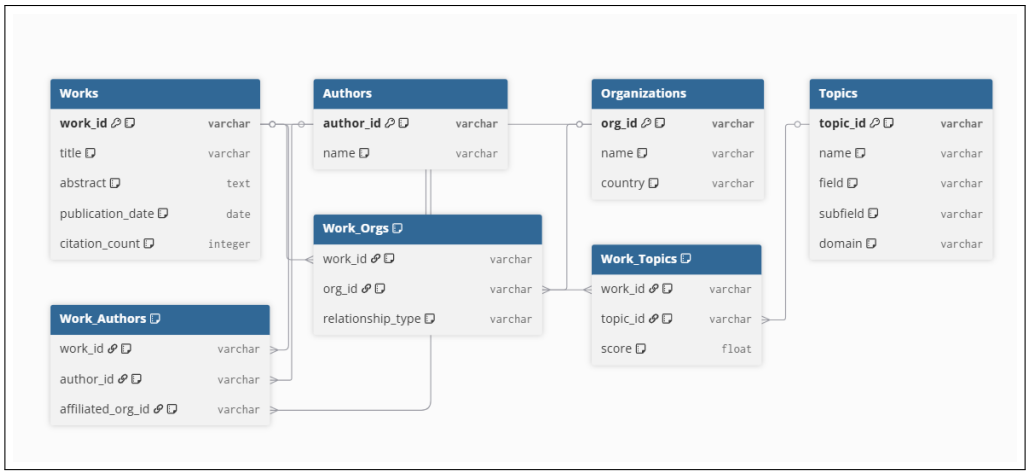

Figure 2. The unified data schema integrating DTIC metadata. Authors, Works, Organizations, and Topics are modeled as discrete nodes. This structure allows multi-hop relational traversal, ensuring that semantic vector searches can be grounded in verified organizational hierarchies.

profiles remain strong baselines in large expert spaces. De Campos et al. compare information-retrieval-style expert profiles with machine-learning classifiers on biomedical corpora and find that profile-based ranking is often more robust when there are many potential experts and noisy labels.[6] In Aegis Scholar, the author embeddings and associated metadata play a similar role, providing a compact representation of topical focus that can be combined with network context.

We evaluated six candidate embedding models (comprising both bidirectional encoders and autoregressive decoders) using Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) against a curated set of defense and control queries. MRR evaluates how quickly the first relevant result appears in the ranked list (e.g., $1/R$, where R is the position of the first relevant expert) and produces scores between 0 and 1, with higher values indicating that relevant authors appear earlier in the ranking. NDCG measures the overall ranking quality by penalizing highly relevant results that appear lower in the list and comparing the system's output to an ideal ranking; its scores also range from 0 (worst) to 1 (best), with higher values reflecting rankings that place relevant experts near the top of the list.

The candidate models were drawn from public model hubs and varied in parameter count and context window. The evaluated models are listed here:

- `all-MiniLM-L6-v2`[7]

- `rubert-tiny2`[8]
- `gte-base-en-v1.5`[9]
- `USER-bge-m3`[10]
- `Qwen3-Embedding-0.6B`[11]
- `jina-code-embeddings-1.5b`[12]

The offline comparison highlighted a trade space between retrieval quality and deployment cost. In the implemented prototype, the running stack used the `all-MiniLM-L6-v2` checkpoint from the `sentence-transformers` library for author-search embeddings, while the broader six-model comparison remained part of the offline evaluation.

After the vector service returns a bounded set of candidate authors by semantic similarity, the API applies a custom, multi-variate heuristic to rerank those returned candidates. The relevance score combines semantic distance (x), the author's total citation count (y), and the recency of their publication activity measured in decades (t):

$$score(x, y, t) = w_1(1-x) + w_2\,\sigma\big(0.005(y-y_0)\big) + w_3\left[\frac{1}{2}\big(1-\tanh(t-t_0)\big)\right] \quad (1)$$

Here, $y_0 = 100$ serves as a citation baseline offset, and $t_0 = 2$ decades provides a recency offset. The functions σ (the sigmoid function) and $\tanh$ (the hyperbolic tangent) normalize the citation and recency variables to fall between 0 and 1. The factor of $\frac{1}{2}$ combined with $\tanh$ ensures the third term is correctly scaled between 0 and 1. This normalization ensures that no single factor (e.g., a massive citation count) disproportionately dominates the ranking, given the same weighting factors. The 0.005 multiplier scales the citation growth so the sigmoid transitions smoothly around the y_0 baseline.

The coefficients w_1, w_2, and w_3 act as weighting functions. For this initial prototype development, we chose $w_1 = w_2 = w_3 = 1/3$ to evenly split the weight between semantic similarity, historical impact, and recency. We also recognize that grouping recency by decades ($t_0 = 2$) was an initial design choice; scaling by individual years might prove more granular for prioritizing contemporary research. In

future iterations, these weights and offsets can be empirically tuned based on user feedback to optimize result quality.

3.4 Experimental Design

To validate the hypothesis that a targeted, graph-and-vector approach improves SME discovery over general-purpose tools, we designed a small, task-based pilot user study. The study recruited five experienced technical researchers, ensuring a baseline where the majority had no prior exposure to Aegis Scholar.

Participants were asked to execute discovery tasks using both Aegis Scholar and OpenAlex across three distinct scenarios:

1) **Unmanned Aerial Systems (UAS) Deception** (highly DoW-specific).
2) **Medical Imagery** (interdisciplinary).
3) **Bioluminescent Algae** (niche/oceanography control).

The experimental framework measured both quantitative efficiency (time required to identify three relevant experts per scenario) and qualitative user experience. Participants provided ratings on a five-point Likert scale (ranging from 1 = strongly disagree to 5 = strongly agree) regarding their confidence in the results, the perceived relevance to defense missions, and the utility of the visual interfaces. The study was intended as a prototype-oriented comparison rather than a definitive validation of production readiness. The outcomes of both the offline model evaluations and this user study are presented in Section 4.

4. Results

4.1 Offline Model Evaluation

Prior to the user study, we conducted offline evaluations to select the optimal embedding model. Six candidate models were evaluated against 12 queries categorized into Artificial Intelligence (AI), Military Deception (MilDec), Interdisciplinary (AI and MilDec), and Unrelated Controls.

Table 1 presents the overall performance of the six evaluated models, measured by MRR and NDCG. The overall MRR represents the arithmetic mean of the model's MRR across all tested query categories. A complete table of all evaluated models is provided in Appendix C.

Table 1. Overall embedding model performance.

Model	Overall MRR	Overall NDCG
deepvk/USER-bge-m3	0.667	0.900
all-MiniLM-L6-v2	0.653	0.914
jina-code-embeddings-1.5b	0.653	0.913
Qwen3-Embedding-0.6B	0.625	0.884
gte-base-en-v1.5	0.568	0.766
rubert-tiny2	0.386	0.770

While overall metrics were comparable among the top contenders, analyzing the performance by domain category revealed significant biases inherent to the models' training data (see Table 2). For instance, the gte-base-en-v1.5 model achieved a perfect MRR of 1.000 on Military Deception queries but scored only 0.270 on general AI queries. Conversely, USER-bge-m3 demonstrated high performance across both AI and MilDec domains. All models also returned an MRR of 0.0 for the unrelated control queries (e.g., oceanography). Because MRR measures how quickly a *relevant* expert appears in the results, this indicates that no clearly relevant expert surfaced near the top of those unrelated rankings. In that limited sense, the controls behaved as intended by separating the tested defense-oriented tasks from obviously out-of-scope topics.

Table 2. MRR performance by query domain.

Model	AI	MilDec	AI+MilDec	Unrelated
deepvk/USER-bge-m3	1.000	0.667	1.000	0.000
all-MiniLM-L6-v2	0.778	0.833	1.000	0.000
jina-code-embeddings-1.5b	0.833	0.778	1.000	0.000
Qwen3-Embedding-0.6B	1.000	0.833	0.667	0.000
gte-base-en-v1.5	0.270	1.000	1.000	0.000
rubert-tiny2	0.583	0.700	0.261	0.000

These results informed the decision to deploy all-MiniLM-L6-v2 in the running prototype, favoring a strong balance of performance, implementation simplicity, and computational efficiency for the capstone environment.

4.2 User Study: Efficiency

The pilot user study measured the time required for five participants to identify three relevant experts across three mission scenarios using both Aegis Scholar and

OpenAlex. Table 3 displays the raw average completion times.

Table 3. Average query resolution time (MM:SS).

Scenario	OpenAlex	Aegis Scholar	Reduction
1) UAS Deception	06:05	02:57	-51.5%
2) Medical Imagery	10:12	12:52	+26.1%[a]
3) Bioluminescent Algae	17:29	16:16	-6.9%
Average	**11:15**	**10:01**	**-10.9%**

[a] Scenario 2 average includes a single 35-min outlier due to an interface navigation issue.

In highly DoW-specific tasks (Scenario 1), Aegis Scholar reduced the average search time by over 50%. Scenario 2 resulted in a longer average time for Aegis Scholar due to a single 35-min outlier caused by an interface limitation (specifically, the inability to "pivot" the network graph by clicking on a co-author node). Including this outlier, the aggregate average time reduction across all tasks was 10.9%. If this isolated outlier is removed, the Aegis Scholar average for Scenario 2 drops to 07:15, representing a total aggregate time reduction of 27.6%. Because this was a small pilot study, these results should be interpreted as early evidence of promise rather than as a definitive benchmark.

4.3 User Study: Qualitative Feedback and System Comparison

Following the timed tasks, participants evaluated the systems based on confidence, relevance, and usability.

Relevance and Visualization: Aegis Scholar performed strongly in this small study regarding perceived mission relevance and data presentation. 80% of participants agreed or strongly agreed that the results returned by Aegis Scholar felt more relevant to DoW and government research needs than those from OpenAlex. Furthermore, 100% of participants agreed that Aegis Scholar's network visualization was superior for illustrating the complex relationships between authors, works, and organizations. As shown in Figure 3, the interface simultaneously displays nodes for individual experts, their authored papers (with full titles), and at least one node representing organizational affiliation, tying people, works, and institutions into a single graph view. A lightweight network-centric tool designed specifically for finding experts within the DoW (see Figure 3) was noted as being highly effective for quickly isolating 5 to 10 relevant individuals within a specific defense domain.

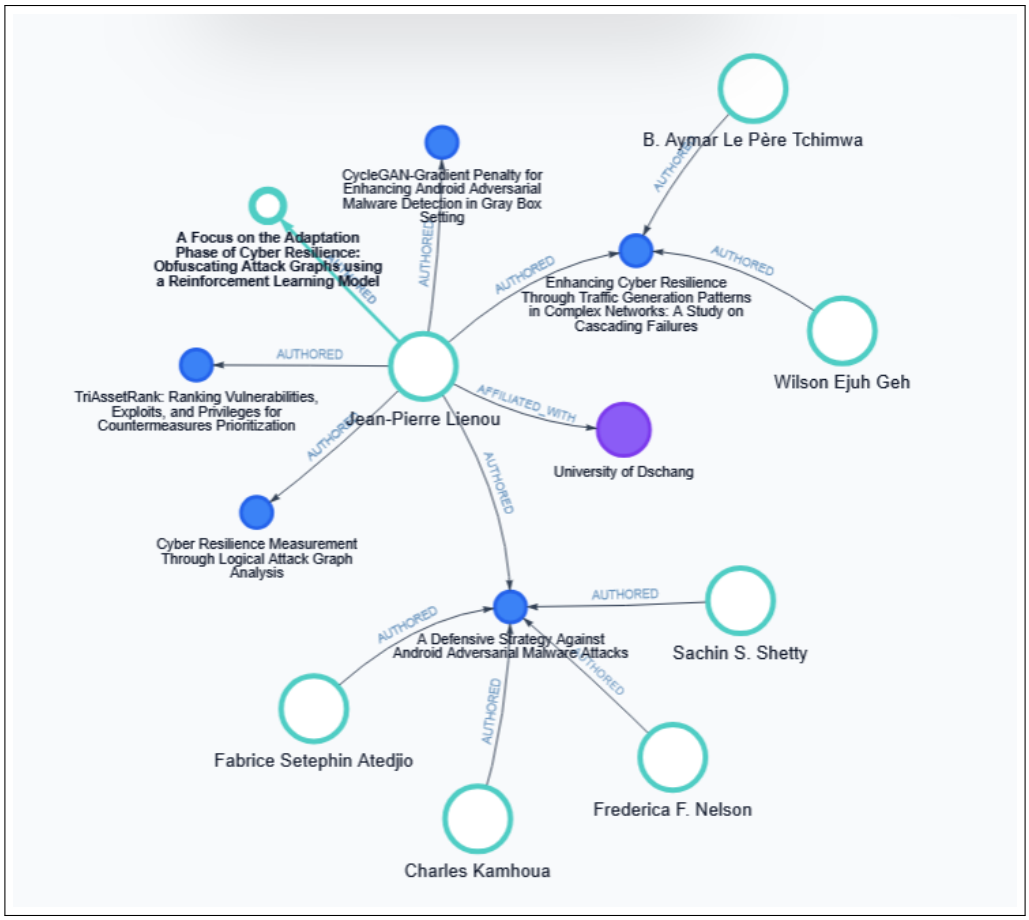

Figure 3. Aegis Scholar user interface illustrating the network-centric approach to expert discovery. Unlike traditional document-list interfaces, the system visualizes the collaborative networks connecting researchers, their authored works, and their organizational affiliations, allowing users to rapidly identify cross-functional teams and institutional clusters.

Confidence and Verification: Despite the higher perceived relevance of Aegis Scholar's results, participants reported slightly higher overall *confidence* when using OpenAlex (average rating of 3.6/5.0 vs. 3.2/5.0). Qualitative feedback revealed that this confidence gap stemmed from verifiability. OpenAlex provided direct, clickable links to ORCID profiles, external papers, and DOI records, allowing expert users to manually validate institutional affiliations and publication history. Aegis Scholar, as a prototype, did not yet provide equally strong in-tool verification or authoritative contact validation, leaving users with less direct support for confirming the suggested experts.

Usability Trade-offs: The evaluation highlighted a clear divergence in user needs based on expertise. OpenAlex is a highly mature, data-dense platform; while this

density is valuable for expert data analysts, novice users frequently found the interface overwhelming and "clunky." Conversely, Aegis Scholar provided a streamlined, intuitive entry point for finding networked personnel, though expert users desired deeper drill-down capabilities (i.e., verifying current email addresses and confirming active government employment status). Taken as pilot evidence, these findings suggest that while mature enterprise systems remain necessary for deep bibliometric analysis, a lightweight, network-centric tool may add value when the immediate goal is to find experts within a topic area.

5. Discussion

5.1 Interpretation of UX Results and Iterative Design

A core tenet of modern software engineering is to collect user feedback early and often. The user study highlighted certain constraints in the Aegis Scholar prototype—most notably, the inability to click on a co-author node to pivot the graph and explore that individual's extended network, as well as the absence of direct links to external papers. These constraints were largely a function of the project's time limits and the prototype's early technology readiness level, rather than fundamental design flaws. They should also be interpreted alongside two broader limitations of the present work: the evaluated system was built on a DTIC-derived snapshot rather than a complete enterprise corpus, and the user study was intentionally small and prototype-oriented. Robust author-name disambiguation is also a known prerequisite for high-fidelity expertise graphs.[13–15] The prototype did not attempt to solve this problem; names were treated at the string level, and verifying that a suggested expert corresponds to a unique person remains future work.

From an operational perspective, the feedback regarding the node pivot actually validates the core *connection engine* hypothesis. When users evaluate research ecosystems, they naturally think in terms of networked relationships. Just as atomistic simulations analyze first-nearest-neighbor and second-nearest-neighbor interactions, researchers naturally want to explore the orbit of an identified expert to validate their relevance. Providing visibility into first- and second-nearest neighbors gives users the surrounding context needed to be satisfied with their initial choices. However, as noted during the study, there is a practical limit to this exploration; unrestricted clicking away from the initial search can lead to distraction. Capturing this feedback early ensures that future iterations can implement controlled pivoting

(allowing users to see the immediate network emanating from an expert) to build confidence without overwhelming the core discovery task.

5.2 Relationship to Existing Tooling and the Value of Minimalism

The scholarly discovery landscape is already served by highly mature platforms such as OpenAlex, Crossref, and Dimensions (which features capabilities tailored to the Department of War and advanced nodal exploration). Initially, the team examined these highly engineered products and observed that their breadth of features could make them difficult to use for specific, targeted queries. By analyzing how these systems operated, the team identified a gap they could exploit: the need for a focused tool tailored explicitly to defense networks without the overhead of a general-purpose academic platform.

The user study suggests a distinct niche for the Aegis Scholar approach. High-end, data-dense tools are exceptionally powerful for expert data analysts and bibliometricians, but that complexity often overwhelms novice or occasional users. In this pilot setting, Aegis Scholar showed that a minimalist, stripped-down tool can provide real practical value. If a user's objective is simply to find 5 to 10 defense-related experts within a specific domain, a targeted interface may be faster and less intimidating than a comprehensive enterprise platform. At the same time, the current prototype should not be mistaken for a replacement for those richer systems: it remained author-centric, lacked stronger in-tool verification, and did not yet provide equally mature search paths for organizations, topics, or works.

5.3 Process Outcomes and Return on Investment

While the specific codebase developed during this project is best understood as a prototype reference architecture rather than as an enterprise-ready system, much of the value for the DoW lies in the process.

The AI Technicians program at AI2C focuses on building technical capability within the military workforce. This capstone project required a team of mid-career Soldiers to navigate the entire lifecycle of an AI project. They did not simply write code; they were required to take a mission-oriented problem, scope the technical requirements, scrape and standardize unstructured data, evaluate competing embedding models, containerize the architecture, and design an experiment that tested user satisfaction.

ARL mentors played a structured role in this process, meeting with the team weekly to steer the project, answer domain-specific questions, and ensure the development remained grounded in real-world constraints. Taken on its own terms, this pilot suggests a repeatable model: pairing operational Soldiers with experienced research mentors can produce useful prototype systems while also developing personnel who are better equipped to procure, manage, and evaluate future AI systems for the DoW.

6. Conclusions and Future Work

The Aegis Scholar project provided a credible pilot demonstration of a specialized, defense-focused "connection engine." By shifting emphasis away from traditional document retrieval and toward authors, collaborators, and organizational context, the prototype met its capstone technical and process objectives within the limits of a training-oriented implementation.

Key outcomes of this project include the following:

- **Hybrid Discovery Architecture:** The prototype integrated semantic vector embeddings (via Milvus) with relational mapping (via Neo4j) to process natural language queries over a DTIC-derived metadata snapshot.
- **Value of Minimalist Design:** Pilot user evaluations suggested that while mature, data-dense enterprise systems remain necessary for exhaustive bibliometric analysis, lightweight, targeted tools may reduce the effort required for some users to identify a curated list of defense-focused SMEs.
- **Process and Capability Development:** The project served as an effective vehicle for the AI Technicians program at AI2C. Mid-career Soldiers successfully navigated the entire lifecycle of an AI project, from problem scoping and data engineering to model evaluation and containerized deployment. This end-to-end execution, guided by ARL mentors, suggests a strong and repeatable model for building technical proficiency within the force.

While the current codebase serves as a useful pilot system and an open-source reference architecture for educational purposes, transitioning these capabilities into an operational environment would require significant structural integration. Rather

than scaling Aegis Scholar as a standalone application in its present form, the concepts explored here—especially author-centric search, network visualization of collaborators, and defense-oriented ranking heuristics—are best viewed as features that could be integrated into existing enterprise knowledge platforms. Future efforts should focus on more complete data ingestion, robust name disambiguation, stronger verification of current affiliation and contact data, and secure access to controlled-distribution documents. More broadly, this work sits alongside emerging science-of-science systems that use concept-level knowledge graphs and machine learning to study how research topics evolve over time.[16] Ultimately, the lessons learned from this prototype provide a useful blueprint for how large organizations might better connect their personnel to the expertise hidden within their institutional data.

Note on AI-assisted tools. During the preparation of this report, the authors used large language models (e.g., Google Gemini-3.1-Pro, OpenAI GPT-5.1) as a writing aid for editing, summarizing, and rephrasing text. The authors subsequently reviewed and edited all output and take full responsibility for the content of the published report.

7. References

1. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint 2022. arXiv:2205.01833.

2. Hendricks G, Tkaczyk D, Lin J, Feeney P. Crossref: The sustainable source of community-owned scholarly metadata. Quantitative Science Studies. 2020;1(1):414–427.

3. Hook DW, Porter SJ, Herzog C. Dimensions: Building context for search and evaluation. Frontiers in Research Metrics and Analytics. 2018;3:23.

4. Visser MS, van Eck NJ, Waltman L. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Quantitative Science Studies. 2021;2(1):20–41.

5. Shafer E, McCoy S, Canedo A, Robinson M, Hale J. Aegis scholar: Author-centric expertise discovery prototype. GitHub repository, 2025.

6. de Campos LM, Fernández-Luna JM, Huete JF, Ribadas-Pena FJ, Bolaños N. Information retrieval and machine learning methods for academic expert finding. Algorithms. 2024;17(2):51.

7. Wang W, Wei F, Dong L, Bao H, Yang N, Zhou M. MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. arXiv preprint 2020. arXiv:2002.10957.

8. Dale D. rubert-tiny2: A lightweight russian BERT model. 2021.

9. Li Z, Zhang X, Zhang Y, Long D, Xie P, Zhang M. Towards general text embeddings with multi-stage contrastive learning. arXiv preprint 2023. arXiv:2308.03281.

10. Chen J, Xiao S, Zhang P, Luo K, Lian D, Liu Z. M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. arXiv preprint 2024. arXiv:2402.03216.

11. Bai J et al. Qwen technical report. arXiv preprint 2023. arXiv:2309.16609.

12. Günther M et al. Jina embeddings 2: 8192-token general-purpose text embeddings for long documents. arXiv preprint 2023. arXiv:2310.19923.

13. Han H, Giles CL, Zha H, Li C, Tsioutsiouliklis K. Two supervised learning approaches for name disambiguation in author citations. In Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries. 2004;296–305.

14. Torvik VI, Smalheiser NR. Author name disambiguation in MEDLINE. ACM Transactions on Knowledge Discovery from Data. 2009;3(3):11:1–29.

15. Rodrigues NS, Mariano AM, Ralha CG. Author name disambiguation literature review with consolidated meta-analytic approach. International Journal on Digital Libraries. 2024;25:765–785.

16. Gu X, Krenn M. Forecasting high-impact research topics via machine learning on evolving knowledge graphs. Machine Learning: Science and Technology. 2025;6(2):025041.

Appendix A. System Infrastructure and Deployment

This appendix outlines the infrastructure and deployment configurations used to containerize and orchestrate the Aegis Scholar prototype. It is a combination of implemented prototype components and deployment-path artifacts developed during the capstone.

A.1 Containerization and Services

The platform operates as a Docker Compose monorepo, consisting of five core microservices communicating over an internal bridge network:

1) **Aegis Scholar API:** A FastAPI application serving as the primary gateway. It orchestrates downstream calls and computes the hybrid relevance scores.
2) **Vector DB:** A Milvus instance storing the 384-D `all-MiniLM-L6-v2` abstract embeddings.
3) **Graph DB:** A Neo4j 5.x instance maintaining the relational network.
4) **Identity Service:** An OpenLDAP-backed prototype directory layer handling fuzzy name resolution and lookup-style identity support.
5) **Frontend:** A React/Vite web application providing the user interface and network visualizations.

A.2 Kubernetes Deployment Strategy

For cloud deployment, the system targets Azure Kubernetes Service (AKS). The infrastructure-as-code is managed via Terraform, which provisions the necessary namespaces, persistent volume claims (PVCs), and load balancers. These materials document the intended operational path for the prototype, even though the evaluated capstone system was not a fully deployed enterprise service.

Key deployment patterns include the following:

- **Horizontal Scaling:** The FastAPI gateway is configured with a Horizontal Pod Autoscaler (HPA) triggered at 70% CPU utilization.
- **Secrets Management:** To adhere to Department of War security guidelines, no plaintext passwords are used in environment variables; all database credentials and API keys are injected at runtime via Kubernetes Secrets.
- **Health Monitoring:** Liveness and readiness probes are configured for all pods. Logs are aggregated via Azure Log Analytics, enabling Kusto Query Language (KQL) analysis of search latency and error rates.

Appendix B. Data Schema and Field Mappings

This appendix details the unified logical schema developed during Phase 2 to map data from Defense Technical Information Center (DTIC) into the graph and vector databases. In the evaluated prototype, this schema was applied to a DTIC-derived snapshot assembled through the project's extraction pipeline rather than to a complete enterprise-scale corpus.

B.1 Field Selection and Normalization

The schema design process prioritized fields with high empirical fill rates (the proportion of records containing data for a given field) to ensure system reliability. Profiling of a 10,000-record DTIC sample revealed that core identifiers such as Digital Object Identifiers (DOIs) had a 99.4% fill rate, and abstracts were present in 96.8% of the records. Conversely, some fields defined in external Application Programming Interfaces (APIs) (e.g., specific organization types) exhibited a 0% fill rate in the sample and were subsequently excluded from the pipeline.

Data normalization rules were established to ensure consistency across the graph database:

- **Identifiers:** Universally Unique Identifiers (UUIDs) were generated deterministically based on source DOIs or standardized author names to prevent duplicate node creation during idempotent `MERGE` operations.
- **Entities:** The flat JSON responses were decomposed into four distinct entity types: `Author`, `Work`, `Organization`, and `Topic`.
- **Text Processing:** Abstracts were cleaned of HTML/XML artifacts prior to being passed to the sentence transformer models for embedding generation.

B.2 Graph Relationships

The normalized data populates Neo4j using the following node labels and directed relationships:

- `(:Author)-[:AUTHORED]->(:Work)`
- `(:Author)-[:AFFILIATED_WITH]->(:Organization)`
- `(:Work)-[:COVERS_TOPIC]->(:Topic)`

Appendix C. Model Evaluation Tables

This appendix provides the extended evaluation results for the six embedding models tested during Phase 3. The appendix documents the offline comparison set; the running prototype described in the main text used `all-MiniLM-L6-v2`.

The evaluation utilized 12 queries categorized into Artificial Intelligence (AI), Military Deception (MilDec), Interdisciplinary (AI and MilDec), and Unrelated Controls. Thirty-two authors were manually rated for relevance to these queries on a scale of 1 to 5.

Table C-1 presents the complete Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) scores for all tested models. The results highlight the trade-off between absolute retrieval speed (MRR) and overall ranking quality (NDCG), as well as the varying degrees of domain bias present in models such as `gte-base-en-v1.5`.

Table C-1. Complete embedding model performance metrics.

Model	Max Tokens	Overall MRR	Overall NDCG
`deepvk/USER-bge-m3`	8192	0.667	0.900
`all-MiniLM-L6-v2`	256	0.653	0.914
`jina-code-embeddings-1.5b`	32768	0.653	0.913
`Qwen3-Embedding-0.6B`	32768	0.625	0.884
`gte-base-en-v1.5`	8192	0.568	0.766
`rubert-tiny2`	2048	0.386	0.770

List of Symbols, Abbreviations, and Acronyms

AI2C	Army Artificial Intelligence Integration Center
AKS	Azure Kubernetes Service
API	Application Programming Interface
ARL	Army Research Laboratory
CMU	Carnegie Mellon University
CPU	Central processing unit
DEVCOM	U.S. Army Combat Capabilities Development Command
DOI	Digital Object Identifier
DoW	Department of War
DTIC	Defense Technical Information Center
ETL	Extract, Transform, Load
HPA	Horizontal Pod Autoscaler
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
KQL	Kusto Query Language
LDAP	Lightweight Directory Access Protocol
LLM	Large Language Model
MilDec	Military Deception
MRR	Mean Reciprocal Rank
NDCG	Normalized Discounted Cumulative Gain
ORCID	Open Researcher and Contributor ID

PVC	Persistent Volume Claim
SME	Subject-Matter Expert
UAS	Unmanned Aerial Systems
UUID	Universally Unique Identifier
UX	User Experience
XML	eXtensible Markup Language