seekr

Navy GenAI Dual-Use Course of Action Solution

Gain the decision advantage by making AI trustworthy

While GenAI is promising, it remains unsafe until bias, hallucinations, and inaccuracies can be eliminated. Seekr's patented Principle Alignment AI scoring technology mitigates biases, expedites AI modeling aligned to DoD benchmarks, and lowers risks to acceptable levels.

GenAI's Ticking Time Bomb

GenAI poses risks because Large Language Models (LLMs) produce unreliable outputs, or "hallucinations," which can lead to misinterpretations and biases within DoD decision-making processes. However, as GenAI theft from U.S. companies by foreign adversaries exacerbates the GenAI arms race to stay competitive, the Navy must implement robust bias mitigation strategies at scale to ensure secure and optimized data for decision-making.

Without accelerated acquisition rules, a tipping point will be reached where the nation will lose its competitive advantage if the DoD cannot mitigate GenAI risks and gain a decision advantage before our adversaries do.

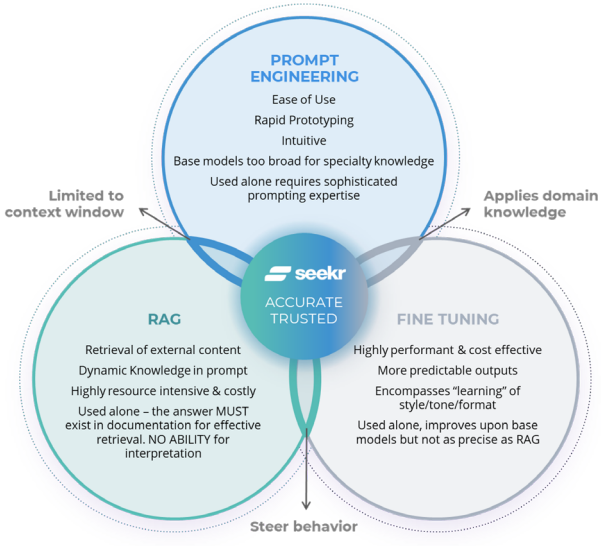
The Solution

Large Language Models rely on statistical methods for context analysis and lack human-like intuition and understanding of the values important in complex scenarios, such as warfighting. As such, oversight of LLMs, both at the individual decision level and at scale, is essential for ensuring they are mission-capable. Seekr's GenAI application applies bias and context rules more effectively than other LLMs.

Seekr enables the DoD to govern and trust LLM outputs by enhancing decision-making capabilities while supporting the intuitive judgment of military personnel. A key use case for LLMs is the development of an agile Course of Action (COA) generator, equipped with a customizable Test and Evaluation (T&E) scoring framework.

This framework leverages DoD policy guidance, documents, rules and validated insights to detect bias and errors in GenAI systems, helping the Navy make better decisions and reduce the risk of adopting unsafe LLMs that are unsafe or prone to foreign malign influence.

SeekrFlow combines prompt engineering, fine-tuning, and retrieval augmented generation (RAG) techniques to scale trusted GenAI applications rapidly (see adjacent Diagram) that mitigate hallucinations and biases in foundation models while making them suitable for Navy use. This drives AI-Enabled Decision Dominance by ensuring continual model accuracy improvement across decision-making scenarios, enabling users to rapidly analyze complex scenarios, coordinate timely responses, and disseminate critical information to distributed command staff and others.



Overcoming DoD's Domain Problems

#1 The DoD Must Align LLMs with Many Domain-Specific Rules in Documents, to Build Trust Scores

Seekr's AI application helps the DoD effectively apply Navy acquisition rules and other governance policies to optimize GenAI validation, acquisition, and adoption.

Designed for regulated industries, Seekr converts real-world benchmarks and policies into AI rules and guardrails that optimize the performance of Seekr's own LLM and other LLMs.

Seekr's patented Principle Alignment technology supports Navy IT divestment decisions per the Navy PEO Digital's Investment Horizon benchmarks, generating Course of Actions (COAs) aligned to the Navy's digital modernization strategy and broader Fleet and DoD missions.

#2 A Shortage of Acquisition Professionals Slows Digital Transformation

Seekr helps the Navy's acquisition stakeholders simplify complex IT divestment decisions by analyzing multiple factors such as costs, schedule, and performance trade-offs. Seekr's Principle Alignment technology works best where multiple courses of action exist, and human insights guide decisions.

Seekr combines AI methods like Retrieval-Augmented Generation (RAG) and fine-tuning with decision analysis to generate optimal COAs that that support Navy strategies. With decades of experience supporting DoD customers, Seekr understands when to employ RAG and when Principle Alignment offers better outcomes, helping align with Chief of Naval Operations (CNO) guidance to "Get Real, Get Better." This balance is key to navigating the uncertainty and complexity inherent in GenAI adoption decisions.

#3 Ineffective GenAI Scoring for Course of Action (COA) Generation

A program of record must manage complex rules and benchmarks, with limited staffing levels. Validating data from program documents to feed into GenAI for COA generation, remains challenging.

Ideally, if program managers trust AI, they can quickly consolidate diverse data and support decision-making. By integrating validated program data and relevant publicly available information (PAI), Seekr anticipates scenarios where adversaries outpace the DoD in acquiring advanced capabilities. In such cases, GenAI could generate COAs and contract language with human-like speed and quality. This would address current bottlenecks in acquisition and allow faster deployment of GenAI capabilities from industry.

In addition, Seekr plans to adapt the Joint Planning Process (JPP) to improve COA generation speed and quality, overcoming the labor-intensive nature of current processes. Unlike other models, Seekr's Principle Alignment technology makes such use cases feasible and cost-effective.

#4 Limitations of Existing GenAI Technologies

Studies show that LLM hallucinations are rooted in biases built in pre-training tasks, because they rely on isolated data sources and lack domain-specific expertise. These biases mean that regulated industries filled with domain-specific expertise can only leverage LLMs with oversight through detailed decision-making processes like JPP-that require collaboration from diverse stakeholders. This comprehensive approach to GenAI T&E and governance helps ensure that LLM applications are grounded and informed by a range of perspectives.

- **Lack of Tailored Benchmarks for GenAI:** Existing AI benchmarks focus on traditional AI models, not GenAI's complex needs in DoD settings and thus lack contextual accuracy and robustness under adversarial conditions.
- **Limited Real-World Testing:** Evaluation processes often fail to replicate real-world conditions, where GenAI models may perform well in controlled settings but struggle in unpredictable military and intelligence environments.
- **Inadequate Risk Management:** Existing AI risk frameworks, like the NIST do not address GenAI-specific risks such as hallucinations, biases, and explainability gaps, which are even more critical in defense applications.
- **Lack of Explainability:** Many GenAI models are black boxes, making it hard to understand their decision-making. This lack of transparency undermines trust, especially in high-stakes military or intelligence contexts.
- **Slow Adaptation to AI Advancements:** Government platforms often lag behind rapidly evolving AI technologies, delaying integration of cutting-edge solutions and hindering timely benefits from AI advancements.
- **Limited Automation in Evaluation:** GenAI testing relies on manual processes, limiting scalability. Only automated tools can handle large-scale evaluations, continuous monitoring, and adaptive risk management, making AI model testing and evaluation itself challenging.

#5 Challenges Scaling Navy IT Acquisition and Divestment Decision Support

PEO Digital manages over 130 IT services for 670,000 users, using its Investment Horizon process for divestment decision analysis. This process, based on Technology Business Management (TBM) and other frameworks, is ideal for integrating GenAI to streamline decision-making. The challenge is harmonizing these standards to address differences, overlaps, and ambiguities, with the Joint Planning Process (JPP) serving as an overarching framework adaptable for smaller staffing levels. Decision analysis best practices, proven in defense, can democratize JPP through GenAI, but this requires validation. Trustworthy AI is needed to govern these processes, which involve multiple vendors with contracts that define IT services and benchmarks. Seekr can ingest contract data to strengthen LLM accuracy, and improve decision support.

Solving the GenAI Problem with Trust and Transparency

Seekr's GenAI trust scores for Navy acquisition decisions can be tailored and scaled across DoD to combine and augment human-like "fast" and "slow" decision processes. Seekr has built two tools: one for COA generation and one for GenAI scoring; no other GenAI platform can deliver and mange both effectively. Seekr's solution offers:

- **Tailored and Scalable GenAI Trust Scores:** Seekr combines "fast" and "slow" decision-making processes. It offers tailored trust scores for COA generation and scoring, scalable across DoD, ensuring both human judgment and AI capabilities work in tandem.
- **Addressing Ambiguities:** Seekr highlights the Volatility, Uncertainty, Complexity, and Ambiguity (VUCA) environment, focusing on the ambiguities and risks in AI acquisition decisions. These challenges present an opportunity to train AI to complement human intuition and experience, improving COA generation by bringing more information and insight to the table.
- **Governance and Continuous Improvement:** Seekr helps customers refine their AI applications through a structured governance process, emphasizing key decision analysis benchmarks like clarity of goals, risk characterization, and value-focused thinking. This process allows for continuous testing, evaluation, and improvement of LLMs.
- **GenAI for DoD and T&E Benchmarking:** Seekr advances DoD GenAI by deploying commercial use cases, validating solutions, and comparing time and quality benchmarks between manual contract generation and LLM outputs. Interviews with stakeholders will assess relevance, ultimately saving time and enhancing decision-making in strategic scenarios.

Enabling the Observe-Orient-Decide-Act (OODA) Loop

SeekrFlow optimizes LLMs for human intuition and decision-making tasks in alignment to any given framework. For example, SeekrFlow can instantiate the Observe-Orient-Decide-Act (OODA) Loop with agents collecting, comparing, and critiquing information in decision-specific spatiotemporal parameters as follows:

- **Observe:** Use LLM agents to collect and pre-process information from documents, sensors and other sources relevant to the decision at hand.
- **Orient:** Apply RAG and GraphRAG to classify entities, establish relationships, and analyze dependencies for the given decision scenario.
- **Decide:** Simulate COAs by analyzing causal relationships and predicting outcomes based upon proven patterns discerned agentically from other edge and enterprise information sources.
- **Act:** Deploy COA(s) and continuously monitor effectiveness through agentic feedback loops that optimize human learning, as environments and competition changes.

About Seekr

Seekr is a privately held AI technology company founded in 2021 to deliver safe, trustworthy, and mission-capable AI and GenAI to the DoD. Seekr allows organizations to develop, tune, deploy and monitor trusted LLMs specific to your mission, powered by your data. The company was named one of Fast Company's "Most Innovative Companies" of 2024.

Request a private demo at www.seekr.com