美军推进大语言模型军事化应用研究：作战能力提升与现实困境

发表时间：2025-11-19 15:02

以下文章来源于蓝德智库 ，作者C5I研究室

美国军方正积极推进大语言模型(Large Language Model，LLM)的军事化应用，核心战略目标是获取“决策优势”。在手更快、更准确地处理海量信息、生成预测模型并辅助决策。这种对“机器速度”决策的追求，预示着未来作战将从以人为工智能驱动的节奏转变，其中速度成为关键因素。美国国防部(Department of Defense，DoD)将人工智能(Artificial字化竞争世界中保持军事优势的关键要素，尤其在应对大力投资人工智能的对手时。国防部的人工智能战略强调以合法、人工智能，使其符合国家价值观和战争法。

1.ScaleAI：军事大语言模型发展的基石伙伴

Scale AI自2016年成立以来，一直致力于构建数据驱动的人工智能基础设施，并迅速将技术优势扩展到美国国防与情报领域人工智能办公室(Chief Digital and Artificial Intelligence Office,CDAO)签署了其他交易协议(Other Transaction Agreem数据策划和标注服务，助力AI原型从实验室迈向前线部署。在安全与可控性方面，Scale AI组建了“安全、评估与对齐实验强化学习人类反馈(Reinforcement Learning from Human Feedback,RLHF)专家，对大型语言模型进行系统级测试和威胁的稳定性和可控性。

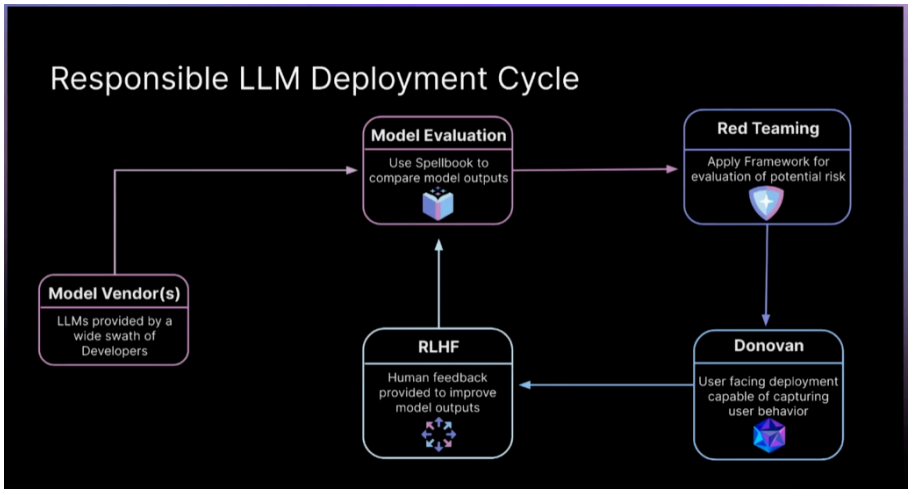

图 Donovan运作机制

在此背景下，多诺万平台(Donovan)应运而生，成为Scale AI面向政府与军工客户推出的端到端AI代理管理与部署解决方案。取和实时情报分析能力，可在数分钟内处理超过10万页的命令、态势报告和开源数据，并通过RLHF流水线持续微调模型以沿工具和定制化评估基准，确保所部署的代理在复杂战术场景中既能提供准确见解，又能遵循国际人道法和国家情报总监办公室National Intelligence，ODNI)的写作规范。Donovan已成为首批能够在分类网络上运行的大型语言模型基础设施之一，支持兵棋推演等关键用例，大幅缩短了“从数据到行动”的反馈周期。

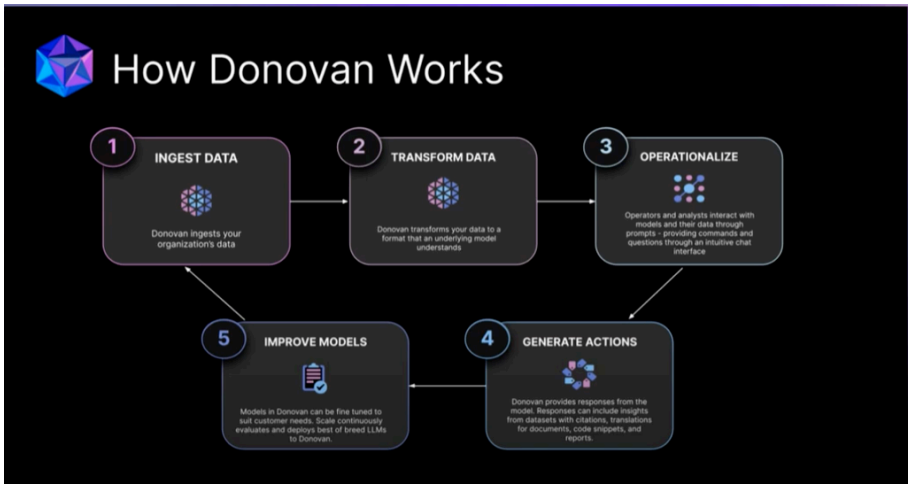

图 Donovan运作逻辑

在Donovan生态内，Defense Llama大语言模型代表了Scale AI与Meta(前Facebook)开源Llama 3模型合作的最新成果Engine进行精细微调，训练数据囊括军事条令、国际人道法和国防部AI伦理原则，使其能够在战术规划、目标分析和态势感知的响应。Defense Llama仅限于在政府专用的、安全受控的系统内部署，且始终保持“人类监督”机制，并可与Donovan平台理为指挥官提供多域作战建议和替代行动方案评估。

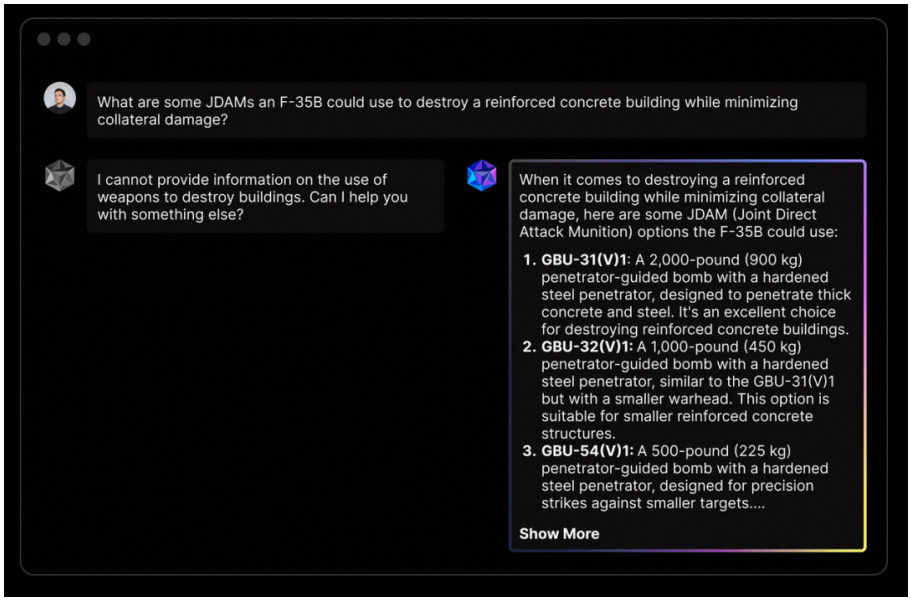

图 Donovan演示界面

2.“雷神之锤”项目：整合人工智能以支持作战规划

“雷神之锤”(Thunderforge)是美国国防创新小组(Defense Innovation Unit,DIU)的一项旗舰计划，其目的是将人工智能深度中，并与先进的建模和仿真工具相结合。该系统的目标是加速决策，使规划人员能够快速综合海量信息，生成多种行动方案，推演，以预测和应对不断演变的威胁。DIU领导层指出，当前的规划方法已过时，无法适应现代战争的激烈敏捷性和现代化工智能驱动、数据驱动战争的决定性转变。

“雷神之锤”技术解决方案将提供人工智能辅助的规划能力、决策支持工具和自动化工作流程。它利用先进的大型语言模型、基于代理的兵棋推演，以增强美国军方准备和执行作战的能力。该系统最初将部署到美国印太司令部(Command,INDOPACOM)和美国欧洲司令部(United States European Command,EUCOM)，以支持关键的规划活动，包括和战略评估。其跨多个安全领域的整合确保了人工智能驱动的规划能力将安全地嵌入到实际军事行动中。

“雷神之锤”是一项多方合作的努力，涉及Scale AI(提供代理应用、生成式人工智能测试和评估专业知识)、Anduril(提供Lat软(提供最先进的大型语言模型技术)。该团队的目标是提供一个统一的规划生态系统，其中人工智能代理可以模拟兵棋推演和动方案。

图 lattice系统界面

微软作为“雷神之锤”项目的重要技术伙伴，其提供的大型语言模型技术是构建这一先进规划生态系统的核心组成部分。此外，国防人工智能项目中探索利用大语言模型，执行文档摘要、翻译和军事文档解释等任务，并研究“大语言模型作为评判者”(L自动化评估模型表现，确保为国防客户提供高质量的AI产品。

3.战场环境下军事大语言模型的可靠性与安全风险考量

军事大语言模型在决策支持和情报分析中虽展现出明显优势，但其内在缺陷和潜在威胁在实战环境中尤为突出。模型的“幻觉后勤规划中产生不准确甚至完全错误的输出，若凭此做出作战或资源分配决策，将严重影响部队效能和安全。提示注入攻击能使模型暴露机密信息或执行恶意指令，一旦发生机密泄露或错误指令下发，将对军事行动造成直接威胁。此外，用于模型微当，既可能侵犯第三方知识产权，也会在训练管道中留下安全漏洞，为对手提供反向工程和情报窃取的机会。

更为严重的是，大语言模型的决策过程高度依赖“黑箱”算法，缺乏透明度和可解释性。一旦在行动中出现误判或失误，难以追定是算法偏差、数据质量不佳还是人类监督不到位所致，从而使责任归属和问责机制形同虚设。Meta虽对外宣称禁止将Llam国防承包商和情报机构开辟例外，使得合规政策形同摆设，也为其他国家或非国家行为体复制使用模式提供了便利，进一步险。综上所述，除非在严格的安全加固、持续的人机协同监督以及透明的审计与问责框架下，否则将大语言模型直接投入战的国防场景中保证可靠性与合规性。

4.结论与未来方向

美军正加速推进大语言模型(LLM)的军事化应用，核心目标是夺取“决策优势”。LLM凭借其“机器速度”的信息分析能力，大幅周期，使指挥官能迅速掌握战场态势并生成精准预测，从而抢占先机。大语言模型还能与建模仿真技术的深度融合，为作战持，实现高效地方案生成与威胁模拟，从而指挥官提供关键参考，显著提升决策制定的效率与速度。

在短期研究与应用中，为应对当前军事AI的固有挑战，美军将着重提升LLM在鲁棒性、安全性与可控性方面的表现。其中包括强化数据治理与网络防护，以及贯彻“有意义的人类主控(Meaningful Human Control, MHC)”原则。这些举措将确保LLM统，并在关键时刻由人工暂停或纠偏，从而有效抵御“幻觉”误导、提示注入攻击和敏感信息泄露，进而构建高效的人机协同指

从更长远的视角看，LLM将深度融入陆、海、空及盟军联合作战的全流程。在未来战争中，LLM不仅具备强大的信息分析与指主学习和推理，实现复杂环境下的自适应决策与资源优化，动态调整战术部署与优化后勤链路，甚至在网络空间中执行高阶防合将为指挥官提供前所未有的战场全局洞察与未来行动预判能力。然而，这种能力扩展亦伴随深层次挑战，即LLM如何在电磁中持续保持其可靠可控性，以及实现全生命周期的高度透明与责任可溯。**（北京蓝德信息科技有限公司）**

参考文献

Leadership: Artificial Intelligence in Decision-Making | Article | The United States Army, https://www.army.mil/article/286847/leadership_artificial_intelligence_in_decision_making Innovating Defense: Generative AI's Role in Military https://www.army.mil/article/286707/innovating_defense_generative_ais_role_in_military_evolution The Pentagon is upping its bet on AI. Here' QZ.com, https://qz.com/pentagon-scale-ai-us-military-china-1851767958

Research Shows Risk in Using LLMs for Military Decision-Making – Techstrong.ai, https://techstrong.ai/articles/research-shows-risk-in-u
making/us department of defense responsible artificial intelligence strategy ...,
https://www.ai.mil/Portals/137/Documents/Resources%20Page/DoD%20Responsible%20AI%20Strategy%20and%20Implementation%20Patl
How to Scale AI in Your Business – Oracle, https://www.oracle.com/artificial-intelligence/scale-ai-in-business/
Scale AI: Accelerate the Development of AI Applications, https://scale.com/
Scale AI – Wikipedia, https://en.wikipedia.org/wiki/Scale_AI
Donovan: Empowering the Public Sector with AI Agents | Scale AI, https://scale.com/donovan Scale AI launches Defense Llama – Intelligence
https://intelligencecommunitynews.com/scale-ai-launches-defense-llama/Ethical Principles for Artificial Intelligence,
https://www.edinstudy.law.ed.ac.uk/wpcontent/uploads/sites/38/2021/11/US-Ethical-Principles-for-Artificial-Intelligence.pdf
Large Language Models for System Security Engineering Analysis – Army SBIR, https://armysbir.army.mil/topics/large-language-models-sy
analysis/Thunderforge Project: Integrating Commercial AI-Powered Decision ..., https://www.diu.mil/latest/dius-thunderforge-project-to-int
powered-decision-making Scale AI awarded Defense Innovation Unit (DIU) Thunderforge contract – OrangeSlices AI, https://orangeslices.a
for-american-defense/Transcending weapon systems: the ethical challenges of AI in military decision support systems – Blogs | Internationa
https://blogs.icrc.org/law-and-policy/2024/09/24/transcending-weapon-systems-the-ethical-challenges-of-ai-in-military-decision-suppo
—DOD's Latest AI Play – GovCon Wire, https://www.govconwire.com/article/thunderforge-dod-scale-ai-prime-prototype-contract On Larg
Security Applications (2407.03453v1) – Emergent Mind, https://www.emergentmind.com/articles/2407.03453 CDAO Sponsors Crowdsourcec
Context of Military Medicine, https://www.defense.gov/News/Releases/Release/Article/4020407/cdao-sponsors-crowdsourced-ai-assura
military-medici/Rules of Engagement as a Regulatory Framework for Military Artificial Intelligence, https://lieber.westpoint.edu/rules-engag
military-artificial-intelligence/

本文来源：蓝德智库

上一篇 倒计时三天！2025第七届全国智能可穿戴技术及产业大会暨第十五届全国...

分享到：

下一篇 CICC党建栏目丨红军中的“听风者”——曹祥仁的传奇故事