

算法赋能下的进攻崇拜——美以对伊先发制人精准打击的攻防理论分析

岳汉景

摘 要：2026 年美以联军对伊朗发动先发制人打击，首次在大规模军事行动中将人工智能系统性嵌入作战链条。基于进攻—防御理论，从客观攻防平衡与主观认知平衡两个维度，分析此次行动的动因与战略结果。客观层面，人工智能通过压缩 OODA 循环、提升打击精度、实现多域协同，为美以创造显著战术进攻窗口。主观层面，人工智能系统的高置信度量化数据制造人工确定性表象，引发决策者算法权威偏见与算法确定性偏见，进而促发进攻崇拜。然而，伊朗通过分布式韧性指挥、低成本无人机消耗战及封锁霍尔木兹海峡等非对称策略，有效抵消美以的初始进攻优势；而美以决策者认知偏差导致其低估防御适应速度与战争成本，终使军事行动陷入“战术胜利、战略失败”悖论。本文揭示了智能化战争中技术优势与战略成功之间的断裂，并指出攻防理论的“主观认知平衡”变量在人工智能时代出现新的偏差机制。

关 键 词：进攻—防御理论；人工智能；先发制人打击；算法确定性偏见；进攻崇拜；攻防平衡

作者简介：岳汉景，博士，上海政法学院政府管理学院教授（上海 201701）。

文章编号：1673-5161(2026)05-0114-19

中图分类号：D815

文献标识码：A

2026年2月28日,美以对伊朗发动先发制人打击。行动首日即造成包括最高领袖哈梅内伊在内的40多名军政要员死亡。伊朗迅速展开代号为“真实承诺-4”的数十轮反击。冲突持续40余天,双方经巴基斯坦斡旋一度停火并举行谈判,但未能达成协议。5月5日美国宣布军事行动结束、转入防御态势,伊朗则表示仍处战争状态。总体而言,40余天军事冲突的初步结果是,美以在战术层面取得了包括精准打击、体系压制和快速突袭在内的显著成功,但这些胜利并未转化为战略目标的实现,因为伊朗未屈服、政权未更迭,冲突反而转入僵持。

美以为何对伊朗发动本次军事打击?美以的战术胜利又为何未能转化为战略成功?对于上述问题,现实主义宏观理论难以充分解释,因为根据该类理论,美国为了维护其唯一超级大国的地位,理应将战略资源集中于应对首要全球竞争对手,而非在次要对手身上通过极端的战争手段冒过度消耗自身实力的巨大风险。联盟理论所强调的制衡逻辑,即国家联合起来对抗主要威胁,同样无法解释美以为何在西方盟友分担成本的情况下以近乎单独行动的形式对伊发动突袭。基于国内政治的对外政策分析理论(如“选举周期论”“转移视线的战争论”或“领导人个人政治生存论”等)通常将一国对外军事行动的根本动因归结为内部政治压力(如特朗普的中期选举困境、内塔尼亚胡的腐败案)。然而,这类解释在本次冲突中忽略了技术自主性的颠覆作用,因为虽然在操作层面政治人物下达了开战指令,但具体“何时、何地、如何打”则是由人工智能系统依据实时数据和算法逻辑决定的。认知心理学理论(含前景理论)虽然可以解释本次行动决策的风险寻求特征,但它无法解释战争的实际进程和结果。比较而言,进攻—防御理论(Offensive-Defensive Theory,以下简称“攻防理论”)的优势在于:它既保留了前景理论对决策心理的关注,又弥补了基于国内政治的相关对外政策分析理论对技术自主性关注的不足(人工智能系统决定了打击节奏与目标选择)。目前而言,鲜见以攻防理论为视角,分析美以对伊朗的先发制人打击。基于此,本文的研究目标是,以攻防理论为视角,分析美以对伊朗先发制人军事打击行动的动因及结果。本文的主要研究问题因此是:攻防理论如何解释2026年美以对伊先发制人军事打击行动的动因与战略结果?围绕主要研究问题,本文提出如下子问题:攻防理论的核心要义是什么?其关键概念(攻防平衡、认知平衡等)如何界定?第二,人工智能赋能军事如何重构了美以同伊朗的攻防平衡,以及美以的攻防计算,从而助推了先发制人打击的决策?战争结果一般决定于攻防平衡的动态发展,因此本文的第三个子问题是:从攻防平衡的动态性视角看,美以的战术胜利为何未能转化为战略成功?

一、攻防理论的核心要义及其在智能化时代的适用性

攻防理论属于国际关系中层理论,其为理解战争起因与和平维持提供了独特分析视角。该理论主要通过攻防平衡这一核心变量,揭示军事技术、地理环境和组织形态等因素如何塑造国家间的战略互动。攻防理论已发展成为解释先发制人动机、军备竞赛升级以及冲突爆发概率的有力工具。在智能化战争时代,人工智能技术正在深刻重塑攻防效能。人工智能的高置信度输出,使得在传统认知偏差之上叠加了一层算法确定性幻觉,结果是决策者的主观感知进一步偏离客观平衡。美以对伊朗先发制人打击提供了一个检验该理论在新技术条件下解释力的关键案例。

(一) 理论起源与核心命题

攻防理论的经典框架由杰维斯(Robert Jervis)于1978年奠定,他提出了攻防平衡与攻防态势可辨识度两个核心变量。^①1998年,范·埃弗拉(Van Evera)将其系统化,提出:“当征服变得容易时,战争更可能爆发;攻防平衡的变化会显著提高或降低战争的风险。”^②同年,格拉泽(Charles L. Glaser)与考夫曼(Chaim Kaufmann)回应了攻防平衡无法测量的批评,将其定义为攻防成本之比,^③并指出该理论对一战、族群冲突及冷战后格局具有解释力。

综合上述学者的论述,攻防理论以攻防平衡为核心概念、以技术和军事任务执行能力为关注焦点,是现实主义的重要分支理论。其核心命题为,进攻占优则战争更可能爆发,防御占优则和平更可能维持。

(二) 客观攻防平衡与主观认知平衡

攻防理论包含两个关键维度,即客观攻防平衡与主观认知平衡。客观攻防平衡是由军事技术、地理环境、社会秩序、军队组织形态等客观因素决定的进攻与防御的相对效能。当进攻性武器(如机动性强的装甲部队、精确制导武器)相对于防御性武器(如反坦克导弹、防空系统)更具优势时,客观平衡偏向进攻;反之则偏向防御。从杰维斯到奎斯特,再经由范·埃弗拉的理论整合,客观攻防平

① Robert Jervis, “Cooperation under the Security Dilemma,” *World Politics*, Vol. 30, No. 2, 1978, p. 211.

② Stephen Van Evera, “Offense, Defense, and the Causes of War,” *International Security*, Vol. 22, No. 4, 1998, p. 5.

③ Charles L. Glaser and Chaim Kaufmann, “What Is the Offense-Defense Balance and How Can We Measure It?,” *International Security*, Vol. 22, No. 4, 1998, p. 46.

衡始终是攻防理论的核心自变量。值得指出的是,该平衡并非静态。军事技术演进、战术创新以及军队组织形态变化都会持续改变攻与防的相对效能。尤其是防御方可以通过调整指挥结构、采用低成本消耗手段和实施非对称升级等策略,主动适应并抵消进攻方的初始优势。尽管攻防平衡的变化可以是真实的,也可以是被感知的,但变化本身会影响战争的爆发概率以及爆发后的过程和结果。

主观认知平衡强调决策者对攻防平衡的主观感知,往往比客观平衡本身更直接地驱动国家行为。如果决策者认为进攻占优,即便客观上并非如此,仍倾向于采取先发制人策略;反之,如果他们认为防御占优,则更倾向于维持现状。范·埃弗拉在其著作中将“关于攻防平衡的信念”列为战争爆发的关键认知变量,提出“当国家对战争结果产生错误乐观时,战争更可能发生”^①等假设。杰维斯指出,决策者实际上是根据“他们所感受到的脆弱性”来行动的,而这种感受“可能与实际情况不同”。他进一步区分了主观安全需求的两个维度:一是“他们渴望多少安全”,即愿意为安全付出多少代价;二是“对威胁的感知”,即“对对方是否会合作的估计”。^② 这些主观因素意味着决策者基于有限信息和感知形成的判断往往与实际攻防态势存在差异,而这正是理解战争起因的关键。

(三) 攻防理论在智能化时代的适用性

攻防理论形成于工业化战争时代,其核心案例主要基于一战、二战及冷战时期的历史。然而,该理论具有足够的抽象度,其核心分析框架,即攻防平衡如何塑造决策者的成本—收益计算,可以应用于不同技术时代。

人工智能正在从根本上改变攻防平衡的技术基础。智能化战争不再仅仅依赖平台性武器的数量对比,而是日益受算法质量、数据规模、算力水平和人机协同效率等新变量支配。^③ 杰维斯所揭示的技术特征与先发制人动机之间的逻辑联系,可以延伸应用于分析人工智能军事化对攻防平衡的影响。

在客观层面,人工智能对攻防平衡的影响体现在多个维度。首先,人工智能显著压缩了“观察—定向—决策—行动”(OODA)循环的时间,使进攻方能够在防御方有效反应之前完成打击,增强了进攻的突然性。^④ 有研究指出,人工智能

① Stephen Van Evera, *Causes of War: Power and the Roots of Conflict*, Ithaca, NY: Cornell University Press, 1999, p. 14.

② Robert Jervis, “Cooperation under the Security Dilemma,” pp. 174–175.

③ 参见蔡翠红:《超越“修昔底德陷阱”?——人工智能的“赋能—颠覆”效应与大国竞争新范式》,载《世界政治研究》2025年第4辑,第50–52页。

④ Michael Raska, *Mitigating the Risks of AI-driven OODA Loops in Military Decision-making*, The Hague: The Hague Centre for Strategic Studies, April 2025, p. 4.

赋能的进攻可能具有破坏战略稳定的作用,因为国家可能会发现自己越来越有能力使用武力,同时也越来越害怕被使用武力,从而展开人工智能军事化竞赛。^①其次,人工智能驱动的精准识别与打击能力改变了传统“面摧毁”的作战模式,代之以对关键节点的精准打击,大幅降低了进攻的成本。人工智能将“通过提高军事能力、影响军事战略制定和军事组织形态”等,系统性地重塑攻防平衡。^②第三,人工智能实现了跨域协同,即“认知、信息、网络、心理和物理等多个领域”^③的同步打击,使防御方在多维度同时承压。当然,人工智能也可以为防御方所用。人工智能既可以通过识别关键基础设施中的弱点扩大攻击面,也可以为防御方提供识别攻击、评估风险和制定缓解措施的先进工具。“人工智能将在总体上增强进攻或防御能力。人工智能会改善某些方面,阻碍其他方面,并使某些方面保持不变。”^④

在主观层面,人工智能深度介入决策过程带来了新的认知风险。“关于自动化偏见(automation bias)的研究表明,人类往往会对人工智能过于自信。”^⑤人工智能系统输出的高置信度量化数据(如“目标识别准确率 95%”)可能诱导决策者产生过度信任,低估算法固有局限性和战场不确定性。笔者将这一由算法输出特征所诱发的认知倾向称为“算法确定性偏见”。其核心机制是,决策者被精确数字的表象所蒙蔽,倾向于将统计概率误读为确定性结论,而忽视了算法固有的不确定性、数据偏差和模型局限。

然而,算法确定性偏见并非算法影响主观认知的唯一通道。算法还通过“算法权威偏见”发挥作用。如果说算法确定性偏见的机制是“被精确数字蒙蔽”,那么算法权威偏见的机制则是“因技术权威而被信任”。^⑥二者在认知层面形成叠

① “Artificial Intelligence and Offensive Cyber Weapons,” *Strategic Comments*, Vol. 25, No. 10, 2019, p. 10.

② 郭翠萍、刘美君:《智能化对弈:军用人工智能对现代战争攻防平衡的重塑》,载《战略决策研究》2024 年第 6 期,第 77 页。

③ Serhii Bazarnyi et al., “Information Systems Modeling,” *Advanced Information Systems*, Vol. 9, No. 3, 2025, p. 22.

④ Andrew J. Lohn, *The Impact of AI on the Cyber Offense-Defense Balance and the Character of Cyber Conflict*, Washington, D. C.: Center for Security and Emerging Technology, Georgetown University, 2025, p. 1.

⑤ Michael C. Horowitz, Lauren Kahn, “Bending the Automation Bias Curve: A Study of Human and AI-based Decision Making in National Security Contexts,” *arXiv preprint*, 2023, <https://arxiv.org/abs/2306.16507>, 上网时间:2026 年 5 月 29 日。

⑥ UNICRI and INTERPOL, *Responsible AI Innovation in Law Enforcement: AI Toolkit — Technical Reference*, Geneva: UNICRI and INTERPOL, Revised February 2024, p. 37.

加效应:前者使决策者误读了算法输出的性质(将可能是错误的概率当作确定性),后者则使决策者放弃了对此类误读进行审慎质疑的意愿。它们是自动化偏见在不同机制下的具体表现形式。自动化偏见指人类因认知惰性而过度信任自动化系统建议的倾向,其核心机制是“走最小认知努力路径”。^①

算法确定性偏见和算法权威偏见并非独立于攻防理论所强调的传统认知偏差而运作,而是为传统偏差提供了新的量化证据基础。范·埃弗拉所揭示的“错误乐观主义”^②(决策者高估自身能力、低估对手反应的倾向)在 AI 赋能场景下可能被显著强化。当决策者面对“目标识别准确率 95%”“打击成功率 98%”等量化输出时,这些精确数字为既有的过度自信提供了看似客观的佐证,使“错误乐观”获得了技术背书而更具说服力。

传统的攻防理论往往假设决策者是理性的,而认知心理学则侧重于个体认知偏差。然而,在 AI 介入的现代战争中,二者并非割裂(见图 1)。AI 系统不仅重塑了客观层面的攻防平衡(见框 1、2,由左向右排序),更通过其高置信度、量化的输出特征,充当了连接客观态势与主观认知的关键中介(见框 3)。AI 输出的高置信度数据(触发算法确定性偏见)与算法被视为“技术中立”的权威来源(触发算法权威偏见)共同作用,使决策者对进攻行动的精准度、速度和成本形成系统性的过度确信(见框 4),主观认知平衡因此显著向“进攻占优”的方向偏离。

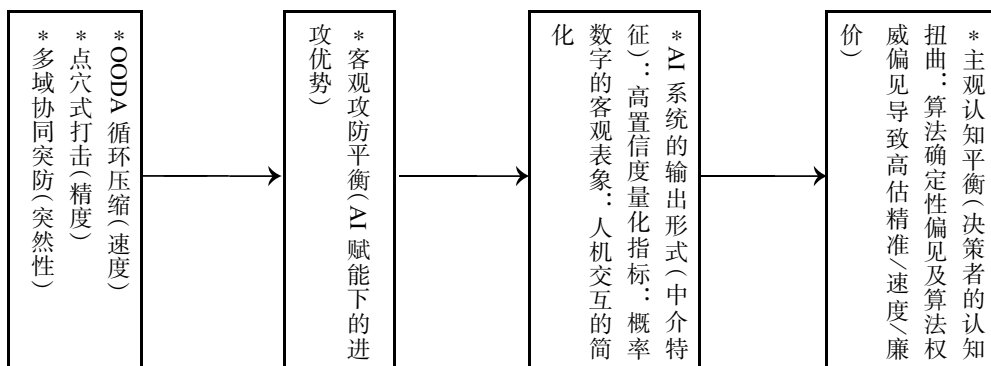


图 1 客观攻防平衡与主观认知平衡通过 AI 连接的逻辑路径

① Lauren Kahn, Emelia S. Probasco, and Ronnie Kinoshita, *AI Safety and Automation Bias: The Downside of Human-in-the-Loop*, Washington, DC: Center for Security and Emerging Technology, Georgetown University, 2024, p. 5.

② 这里需要做出一个重要区分:“错误乐观主义”是攻防理论中已被充分讨论的传统认知偏差,而“算法确定性偏见”与“算法权威偏见”则是人工智能时代特有的新偏差机制——它们并非替代传统偏差,而是直接为“错误乐观主义”提供了新的形成机制。在本文的分析框架中,后两者是核心分析概念,前者则作为背景性认知条件加以参照。

需要强调的是,客观攻防平衡的改变是主观认知偏离的前提条件。在常规军事对比中原本就占据优势的一方,经 AI 在速度、精度、协同上的实质赋能后,“进攻占优”的感知获得了更加扎实的物质基础;而传统“错误乐观主义”与两种算法偏见相叠加,进一步放大了这一感知——前者的广泛存在为过度确信提供了方向性基础,后者则在精准、速度、成本三个关键判断上为其注入了看似客观的“量化证据”,二者共同使决策者形成了“先发制人打击必胜、速决、廉价”的最终结论,促发“进攻崇拜”。^①。逻辑链条如下:作为物质基础的既有军事优势,在 AI 的客观赋能下得以显著扩大;而错误乐观主义与两类算法偏见则进一步从主观层面放大了这一优势,并由此强化了决策者对战争结果的乐观预期,最终促发进攻崇拜。

同时,攻防平衡的动态性在智能化时代表现得更为突出。防御方同样可以利用人工智能分析多源数据,提前调整部署、采用高效费比的防御策略,从而变被动反应为主动设防。当防御方成功适应时,成本天平便会逆转,作战优势就会回归防御方。人工智能对不同战略领域的攻防平衡影响仍存在高度不确定性,“网络领域太过复杂,无法用单一答案来概括人工智能是会增强进攻还是防御的总体效果”^②。这种攻防优势的快速转换使智能化战争的进程更加复杂且难以预测。

综上所述,在智能化时代攻防理论仍然构成一个有力的分析框架。该理论中的客观攻防平衡与主观认知平衡这两个核心概念,为理解人工智能技术如何影响战争决策和战略结果提供了坚实的理论基础。因此本文将攻防理论应用于 2026 年美以对伊先发制人打击的分析,检验该理论在人工智能赋能军事的条件下的解释力。

二、算法赋能下的进攻窗口：美以先发制人打击的动因逻辑

2026 年 2 月 28 日,美以联军对伊朗发动先发制人的精准打击。从攻防理论的视角看,这一决策并不偶然,一定程度上是人工智能技术深刻重构攻防计算后的结果。本部分将从客观与主观两个层面,分析人工智能如何改变了美以对攻防平衡的认知,进而增强了其发动这场高风险军事行动的意愿。促使美以采取

① “进攻崇拜”是攻防理论的一个重要概念,其指“一种认为征服容易的信念”。参见 Stephen Van Evera, *Causes of War: Power and the Roots of Conflict*, p. 237.

② Andrew J. Lohn, *The Impact of AI on the Cyber Offense-Defense Balance and the Character of Cyber Conflict*, p. 1.

军事行动的因素众多,其中,AI 赋能之外的美以对伊朗的既有力量优势构成了行动的客观基础(抛开 AI,双方在客观攻防上本就不平衡)。但本文聚焦于 AI 赋能如何在这一既有优势基础上进一步推动了美以的行动决策。

(一) 客观层面的攻防计算:人工智能重构的美以进攻可行性

在传统战争中,进攻方是否采取行动,往往取决于其对自身力量、敌方防御能力以及战争成本的综合评估。人工智能的介入,在三个关键维度上显著改变了这一计算。这三个维度是速度、精度与协同。

1. 速度的绝对优势:OODA 循环的压缩

军事理论家约翰·博伊德(John Boyd)提出的“观察—定向—决策—行动”循环是理解作战节奏的核心框架。谁能更快地完成这一循环,谁就能获得战场主动权。^①在本次行动中,人工智能将美以的 OODA 循环从传统的小时级压缩至分钟级,甚至秒级。“美以联军的人工智能指挥系统可在数秒内生成多套打击方案并推演效果,一个 20 人的情报小组借助人工智能,即可产出等同于过去 2,000 名参谋的工作量。”^②

这一速度优势在实战中得到了极致体现。“阿联酋关联智库哈布图尔研究中心指出,此次空袭前,美以军方整合分析了海量卫星情报及各类数据。这项工作若由人工完成,需要 328 名分析师耗时 100 天,而人工智能仅用约 90 分钟就全部完成。”^③“美以联军引入克劳德(Claude)大模型,在 90 分钟内处理了 2.3PB 异构数据(相当于 1.2 亿条碎片),效率较人工提升近 800 倍。该系统能以 98.7% 的准确率实时转写波斯语通话,快速提炼出伊朗特种部队总部、导弹阵地等核心目标,支撑多轮精准空袭。”^④

以军部署的人工智能系统更是展现了人工智能在持续监控中的巨大效能。以色列使用了“薰衣草”与“福音”两套人工智能系统,“前者用于生成人员目标清单,一度录入 3.7 万个潜在目标,后者主要筛选建筑物、基础设施等实体打击目标并划分优先级。”^⑤“在哈梅内伊遇害前,以色列情报部门就已深度渗透伊朗首都德黑兰的监控网络。通过入侵交通监控摄像头,结合人工智能算法分析,以军

① Frans P. B. Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd*, London and New York: Routledge, 2007, pp. 234-257.

② 李明海、刘伟:《AI 军事化颠覆战争范式》,环球网,2026 年 3 月 17 日, <https://3w.huanqiu.com/a/de583b/4QmD8MEJZWV>, 上网时间:2026 年 8 月 17 日。

③ 张全:《伊朗战事:被 AI 压缩的“杀伤链”在吞噬什么?》,上观新闻,2026 年 5 月 13 日, <https://www.shobserver.com/news/detail?id=1106832>, 上网时间:2026 年 5 月 16 日。

④ 李明海、刘伟:《AI 军事化颠覆战争范式》。

⑤ 张全:《伊朗战事:被 AI 压缩的“杀伤链”在吞噬什么?》。

对伊朗政权核心人物的生活轨迹了如指掌。”^①

这种“秒级响应”的能力,使伊朗传统的预警和反应机制几乎失效。随着对伊朗发动袭击的准备工作逐步展开,配备了安特罗匹克(Anthropic)公司开发的克劳德人工智能模型的梅文(Maven)智能系统列出了数百个可能的目标地点,提供了这些目标的精确地理坐标,并根据其战略重要性对这些目标进行了排序。梅文与克劳德的结合大大提升了军事规划的效率。指挥官们可以在近乎实时的情况下做出决策,而无需像过去那样花费数周时间来准备战场事宜。^②在伊朗决策层尚未确认威胁性质时,打击已经降临。这正是“速战速决”战略意图的技术基础。

从攻防理论的视角看,这种速度优势直接增强了进攻方行动的突然性。杰维斯曾指出,当进攻方能够通过技术手段获得先发制人的显著优势时,战争的爆发概率会显著上升。人工智能对OODA循环的压缩,正是这种技术优势的集中体现。

2. 精度的致命诱惑:从“面摧毁”到“点穴瘫痪”

传统空袭往往需要大规模轰炸才能摧毁一个目标,因此结果往往是附带损伤多、成本高、舆论风险大。人工智能赋能的精准识别与打击能力,彻底改变了这一模式。通过构建伊朗作战体系的“数字化镜像”,人工智能能够识别出关键节点,如指挥中枢、通信中继、雷达站、导弹发射车等,并推荐最优打击序列。

在此次行动中,“在行动最初 24 小时内,美军就在人工智能协助下打击了伊朗境内超过 1,000 个目标。截至 3 月 25 日,据美军中央司令部司令库珀称,这一数字已经突破 1 万。”^③这一打击行动的背后,是一条由人工智能主导的精准情报链。从人力情报锁定时间窗口,到卫星和信号情报交叉验证目标位置,再到隐身战斗机执行精确打击,人工智能贯穿了从情报收集到火力打击的全流程。这种“点穴式”打击使得进攻方可以用极少的弹药实现体系瘫痪,大幅降低了行动成本与风险。此次先发制人打击中美以“使用的导弹数量还不到(2024 年 9 月)杀

① 晁晶、齐旭:《新民环球|谁在主导战场? 算法还是人类?》,上观新闻,2026 年 4 月 12 日, <https://www.shobserver.com/news/detail?id=1094980>,上网时间:2026 年 5 月 23 日。

② Esra Tekin, “US Used AI-powered System to Identify Targets in Iran: Report,” *Anadolu Agency*, March 5, 2026, <https://www.aa.com.tr/en/americas/us-used-ai-powered-system-to-identify-targets-in-iran-report/3851248>,上网时间:2026 年 5 月 23 日。

③ 晁晶、齐旭:《新民环球 | 谁在主导战场? 算法还是人类?》。

死藏身地堡深处的黎巴嫩真主党领导人纳斯鲁拉时所用弹量的一半”^①。这体现了人工智能精准打击在弹药成本上的巨大优势。

从攻防理论的角度看,精度提升直接降低了进攻的相对成本。人工智能通过大幅降低进攻方的力量成本(更少的弹药、更低的伤亡风险),使这一比率向进攻方倾斜。

3. 多域协同的突防能力:认知战与物理打击的结合

此次行动突破了传统战场界限,实现了物理域、信息域、认知域和网络域的不同步打击。在硬杀伤之前,美军通过逻辑过载、虚假目标注入等算法攻击,使伊朗防空系统瞬间瘫痪,廉价无人机在人工智能调度下实现高效作战。在认知域,生成式人工智能被用于制造高保真的虚假信息,实施精准心理战,扰乱伊朗指挥层的判断。

这种“多域混合”作战模式,使得防御方在多个维度同时承压,反应能力被系统性削弱。从攻防理论视角看,人工智能放大了进攻方的协同优势。这是传统攻防理论较少讨论的维度,但在智能化战争中已成为关键变量。

(二) 主观认知层面的扭曲:美以“进攻崇拜”的形成

前文从客观层面分析了人工智能如何在速度、精度、协同三个维度上为美以创造了显著的战术进攻窗口。然而,客观攻防平衡的改变并不自动催生“进攻崇拜”。从“进攻优势”到“进攻崇拜”的跨越,还需要一个主观层面的认知转化机制。客观层面(AI实质性地提升了进攻效能)与主观层面(算法权威偏见与算法确定性偏见将这种效能放大为“尽在掌控”的确信)的双重作用,推动了美以对伊先发制人打击的决策。在此需要澄清两种偏见的分工与关联:算法权威偏见作用于决策链条的输入端口,使决策者不加批判地接受AI系统提供的各类信息;算法确定性偏见则作用于信息处理环节,使决策者将量化信息误读为确定性结论。前者关闭了质疑通道,后者对数字进行误读,二者叠加才构成完整的认知偏差链条。其他传统认知偏差(如过度自信或错误乐观主义)虽存在,但算法的主导作用在于其高置信度输出为这些偏差提供了新的、量化的“证据基础”,从而使传统认知偏差得到了进一步强化。在这一链条中,算法权威偏见是贯穿三重迷思输入端的前置条件,而算法确定性偏见则直接决定了具体的战略误判内容:精准度被高估、速度被高估、成本被低估。正是这三重误判,构成了“进攻崇拜”的

^① 曹然:《“斩首”行动细节揭秘:美以战机同地起飞,直扑哈梅内伊》,中国新闻周刊网,2026年3月3日,<https://www.inewsweek.cn/world/2026-03-03/29103.shtml>,上网时间:2026年5月23日。

核心认知内容。以下将重点讨论在美以对伊军事行动决策中,算法确定性偏见如何通过这三个具体认知维度显现。每个维度都以战前 AI 系统输出的某一类高置信度数据为起点,经由该偏见的中介,最终形成对该维度的过度确信。这三重过度确信共同强化了对进攻的偏好。

第一,美以的精度迷信。在行动筹划阶段,AI 系统输出的目标识别数据构成了决策层精度信心的首要来源。帕兰提尔(Palantir)公司的哥谭(Gotham)平台对伊朗导弹发射车等目标的识别准确率被标注为“95%以上”,梅文(Maven)系统辅助识别并排序的目标数量在 24 小时内即超过 1,000 个。^①然而,这一“95%”本质上是在特定测试集上得出的统计概率,其可靠性高度依赖训练数据的质量与时效性。但算法确定性偏见使决策者将“95%”理解为“100%精确”,将“统计上大概率正确”等同于“事实上确定无疑”。在这一认知滤镜下,美以决策者很可能相信 AI 赋能的打击体系已达极高精度。事实上,梅文系统的总体准确率仅徘徊在 60%左右,而人类分析师的准确率为 84%。^②即便如此,决策层很可能仍将“95%的识别准确率”等同于“100%的战略成功”,正是算法确定性偏见在精准确度维度的直接产物。2026 年 3 月 12 日,美国国务院发布的中文官方声明开篇即称:“在唐纳德·J·特朗普总统坚定不移的领导下,美国武装部队正在以压倒性的力量、置敌于死地的精准度以及坚不可摧的决心执行‘史诗狂怒行动’。”^③“置敌于死地的精准度”这一表述将“精准”从战术层面的技术特征提升为战略层面的行动保证,暗示“因为足够精准,所以必定成功”。这正是“精度迷信”的话语体现。

更为隐蔽的一重误推在于:即便物理打击做到了“精准”,其政治后果也无法被算法预测,毕竟针对一国最高精神领袖的“定点清除”,其冲击绝非仅限于目标个体的物理清除,而是会触发整个政治体系的重组与报复的升级。

第二,美以的速度迷思。AI 系统输出的另一类高置信度数据是“决策—打击”链路的响应时间。“人工智能大大缩短了从最初发现目标到将其摧毁所需的时间。作为对比:在第二次世界大战期间,从比如通过空中侦察收集情报,到实

① Omar Ahmed, “Integration of AI in Iran War Fuels Ethical Debate amid Heavy Human Toll,” *Rudaw*, April 22, 2026, <https://www.rudaw.net/english/categories/opinion/952886>, 上网时间:2026 年 7 月 10 日。

② 《美媒文章:伊朗战争成“第一场人工智能战争”》,参考消息网,2026 年 4 月 4 日,上网时间:2026 年 5 月 10 日。

③ 《史诗狂怒行动:美国以果敢威力粉碎伊朗恐怖政权》,美国国务院官网,2026 年 3 月 12 日, <https://www.state.gov/translations/chinese/20260312-operation-epic-fury-decisive-american-power-to-crush-irans-terror-regime/>, 上网时间:2026 年 7 月 19 日。

施轰炸任务,往往需要长达六个月的时间。”^①AI 将 OODA 循环,“从小时级压缩至秒级”。美以联军智能指挥系统可在数秒内生成多套打击方案并推演效果。^②对于识别目标与打击目标之间的时间差,历史上首次从数小时压缩至数秒。然而,这些“秒级响应”数据同样是在受控条件下测得的技术指标,其前提是通信链路畅通、目标处于持续追踪状态。美以决策者在算法确定性偏见作用下,将这一战术层面的速度指标理解为确定事实,仿佛战争进程的每一个环节都可以被精确计算和准时执行。在这种“速度迷思”的支配下,决策层相信 AI 赋予的响应速度足以使伊朗陷入“尚未反应即已被击败”的境地。行动初期美以官方释放出“行动很快结束”的信号正是这一迷思的直接体现。^③事实证明,即使拥有最先进的智能化打击体系,也无法阻止一个有韧性的国家在遭受核心打击后 12 小时内组织起大规模导弹反击。算法确定性偏见的要害在于,它将一个受控条件下的战术响应数据误读为确定事实,进而使决策者产生了对整个战争进程的虚假掌控感。

第三,美以的成本幻觉。AI 系统在战前输出的第三类高置信度数据,是与精确打击方案相关的弹药消耗量、打击效能比等量化指标。AI 能够精确估算“打击 N 个目标需要 M 枚弹药”,但它无法量化战争的政治成本、经济代价以及长期战略消耗。算法确定性偏见使决策者将这些局部量化数据误读为完整的成本图景,将“弹药账单”当作了“战争总账”。美以领导人在战争初期对战争总成本的预估均显著低于后期显示的实际成本。^④当伊朗用 2 万美元的无人机消耗美军数百万美元的拦截弹时,^⑤“廉价”的幻觉便暴露无遗。算法确定性偏见在此的运作逻辑是:它将 AI 能够量化的部分包装为“全部成本”的确定性结论,而将 AI 无法量化的部分排除在决策者的认知视野之外。

上述三种倾向并非简单的并列关系,而是在算法确定性偏见的驱动下形成

① Israel Wullman, “The First AI War? How Algorithms and Data Are Reshaping the War with Iran,” *Ynetnews*, March 5, 2026, <https://www.ynetnews.com/tech-and-digital/article/sjksd1likbg>, 上网时间:2026 年 7 月 11 日。

② 李明海、刘伟:《AI 军事化颠覆战争范式》。

③ “Supreme Leader of Iran Khamenei Dead Following Israel’s Strike on Iran,” *Fox News*, March 1, 2026, <https://www.foxnews.com/live-news/israel-us-launch-attack-on-iran-amid-escalating-protests>, 上网时间:2026 年 7 月 11 日。

④ 李佳:《哈佛教授警告:美伊战争或让纳税人背负 1 万亿美元“沉重账单”》,华尔街见闻,2026 年 4 月 14 日, <https://wallstreetcn.com/articles/3769950>, 上网时间:2026 年 7 月 19 日。

⑤ Michael Brown, “The First AI War: How the Iran Conflict Is Reshaping Warfare,” *Forbes*, March 30, 2026, <https://www.forbes.com/sites/mikebrown/2026/03/30/the-first-ai-war-how-the-iran-conflict-is-reshaping-warfare/>, 上网时间:2026 年 7 月 12 日。

了一个紧密咬合、互为前提的强化回路。精准度迷信消除了“打不准”的担忧,为速度迷思提供了信心基础;速度迷思反过来强化了精准度迷信的战略价值;成本幻觉则为前两者提供了“可行性”的经济背书。三者互馈的结果,便是促发“进攻崇拜”。

综上,算法确定性偏见构成了美以决策层从“必须行动”的紧迫感走向“如此行动”的关键认知桥梁。在伊朗内部骚乱、地缘变局、军事部署等多重因素共同营造的“不能再等”的决策氛围中,美以决策者已陷入“不行动即损失”的认知;而算法确定性偏见则在此基础上,回答了“为何敢于如此行动”的问题。它将大规模先发制人打击这一高风险选择,在主观认知中重新编码为“精准、快速、廉价”的技术执行方案。当“不得不做”的紧迫感与“做得到且做得起”的确信感相遇时,审慎考量的空间便被彻底压缩。这一因果链条表明:前沿军事技术本身并不必然导致冒险,但技术能力经由算法确定性偏见的中介所促发的“进攻崇拜”,却可能成为推动国家走向战争的关键心理动力。

（三）攻防计算的重构与先发制人打击的决策逻辑

综合以上分析,人工智能对美以攻防平衡的重构体现在客观与主观两个层面(见表 1)。

表 1 人工智能对美以攻防平衡的重构:客观赋能与主观放大

层面	维度	传统计算	人工智能赋能后的计算	对决策的影响
客观	速度	情报到打击周期为数天至数月	压缩至分钟级,秒级生成方案	增强突然性,压缩对手反应窗口
	精度	“面摧毁”模式,成本高、附带损伤大	“点穴式”打击,精准识别关键节点	降低行动风险与弹药成本
	协同	以单域作战为主	物理域、信息域、认知域、网络域同步打击	多维度施压,削弱防御方反应能力
主观	精度	基于经验判断,预期存在误差	基于“95%准确率”等数据,将概率视为确定性	形成精度迷信:高估打击效果,低估政治后果
	速度	预期战争节奏缓慢、反应周期长	基于“秒级响应时间”等数据,将技术指标视为战略优势	形成速度迷思:高估速胜可能,低估对手适应能力
	成本	预期战争成本高昂、难以承受	基于弹药消耗量等量化数据,将局部成本视为全部成本	形成成本幻觉:低估战争总代价,忽视非对称消耗

资料来源：作者自制。

从攻防理论视角看,人工智能在客观上为美以创造了显著的“进攻优势窗口”;同时,人工智能系统输出的高置信度数据诱导决策者产生了算法确定性偏见,使其对人工智能赋能下进攻行动的精准度、速度和低成本产生了非理性崇拜。由于这些技术特征降低了进攻的门槛与代价,决策者逐渐形成了这种信念:进攻是解决战略问题的最优甚至唯一手段。本文将这种基于客观进攻优势,由技术赋能诱发的、对进攻行动的过度偏好与依赖,称为“进攻崇拜”。这两方面因素促发了美以发动先发制人打击的决策。然而,这种崇拜建立在对技术优势的过度自信和对战争复杂性的系统性低估之上。这正是后续战略困境的根源。

三、战略困境的生成：防御适应与认知偏差的叠加

AI 赋能的美以联军在客观上获得了显著的进攻优势窗口,并在主观上催生了“进攻崇拜”的决策倾向,这共同驱动了先发制人打击行动。然而,战术层面的压倒性胜利并未导向战略层面的预期结局——伊朗政权未更迭、国家未屈服,冲突持续四十余天后,美以被迫转入防御态势。按照攻防理论的核心逻辑,当进攻方取得显著优势时,先发制人往往有望实现速胜。美以决策层显然抱有这种预期。AI 赋能的精准打击似乎又提供了一种“廉价”的战争手段。但实际结果与预期严重偏离:从时间维度看,冲突持续 40 余天,至 4 月初才达成临时停火;从成本维度看,美以陷入了“拦得起、耗不起”的消耗困境;从谈判维度看,首轮谈判不欢而散,战局一度陷入“打不赢、退不出”的僵局。这一“战术胜利、战略失败”的悖论究竟是如何生成的?本部分将运用攻防理论中“攻防平衡的动态性”与“主观认知偏差”两个核心概念,从客观与主观两个维度加以解释。

(一) 攻防平衡的动态性：防御适应抵消进攻优势

攻防理论的一个重要洞见是:攻防平衡并非静态。当进攻方发展出新技术时,防御方会相应调整策略,从而抵消初始优势。伊朗在此次冲突中展现的防御适应策略,从三个层面验证了这一机制。

1. 分布式韧性：去中心化的指挥体系

美以先发制人打击的核心假设是:清除最高领导层将使伊朗陷入指挥瘫痪。然而,伊朗政权的制度设计早已对此有所准备。伊朗体系内部“内嵌着某种韧性”,关键领导岗位预设了多级继任机制,即便顶层遭受打击,体系运转亦

有预案。^① 已故最高领袖哈梅内伊生前曾指示,“每个关键军事和文职职位都要预设多个继任者”,这一制度化安排构成了伊朗“马赛克防御战略”(mosaic defensive strategy)的核心支柱。^② 伊朗的“马赛克防御”学说形成于 2005 年,2008 年重组为 31 个省级指挥体系,每个省级单位都拥有独立的情报能力、武器库和指挥控制设施。该学说源于 2003 年伊拉克战争的教训——萨达姆中央集权式指挥结构在 26 天内崩溃。^③ 战前哈梅内伊已为每个关键职位制定了四层继任计划,每个职位有多达四名指定接替者。^④

2026 年 2 月 28 日,哈梅内伊在首轮打击中身亡。然而,空袭结束不到 170 分钟,伊朗伊斯兰革命卫队即启动了代号“真实承诺—4”的全域反击指令。当天下午,革命卫队即向以色列及中东多处美军基地发射了导弹和无人机。这一反应速度表明,伊朗的指挥体系具备高度的去中心化特征:决策权分散在多个节点,中下层指挥官被赋予独立实施报复性打击的权限。正如有分析指出,伊朗的“马赛克防御”通过“去中心化、预授权、非对称”三大机制,实现了“抗精准打击、防瘫痪、可持续反击”的战略目标。^⑤

从攻防理论视角看,这种分布式韧性构成了对精准打击的典型防御适应。传统防御依赖中心化指挥结构,一旦指挥节点被摧毁便陷入瘫痪;而伊朗的去中心化设计使防御体系对节点损失的容忍度大幅提高,单次精准打击行动难以达到预期效果。

2. 低成本消耗战术:用数量优势消耗对手的高价值资源

伊朗防御适应的第二个层面,体现在其精心设计的低成本消耗战术上。伊

① 黄粤涵、李怡彬:《深度|最高领袖办公室附近遭袭,伊朗权力体系如何“去中心化”强化韧性?》,澎湃新闻,2026 年 2 月 28 日, https://www.thepaper.cn/newsDetail_forward_32671097, 上网时间:2026 年 5 月 20 日。

② Frud Bezhan, “With Top Brass Dead, Iran Deploys Decentralized ‘Mosaic’ Strategy to Boost Defenses,” *Radio Free Europe/Radio Liberty*, March 7, 2026, <https://www.globalsecurity.org/wmd/library/news/iran/2026/03/iran-260307-rferl01.htm>, 上网时间:2026 年 5 月 21 日。

③ Maj. (res.) Alexander Grinberg, “Iran’s Mosaic Defense Faces Its Real Test,” *Jerusalem Institute for Strategy and Security*, March 24, 2026, <https://jiss.org.il/en/grinberg-irans-mosaic-defense-faces-its-real-test/>, 上网时间:2026 年 7 月 8 日。

④ *Jerusalem Post* Staff, “Khamenei Takes Steps to Ensure Succession amid Rising Fears of US Strikes, Assassination-NYT,” *The Jerusalem Post*, February 22, 2026, <https://www.jpost.com/middle-east/iran-news/article-887492>, 上网时间:2026 年 7 月 8 日。

⑤ Asad Mirza, “Iran’s Decentralised ‘Mosaic Defence’ and ‘Fourth Successor’ Leadership Models,” *Raksha Anirveda*, March 14, 2026, <https://raksha-anirveda.com/irans-decentralised-mosaic-defence-and-fourth-successor-leadership-models/>, 上网时间:2026 年 5 月 22 日。

朗大量使用“见证者—136”(Shahed-136)自杀式无人机对美以目标进行饱和攻击。这种无人机的造价极低,据估计每架仅约2万美元。而美以用于拦截这些无人机的“爱国者”导弹约380万美元,“标准—6”约400万美元,“萨德”高达1,500万美元——与之形成了悬殊的代价差异。^①

这种战术的逻辑在于:用廉价武器大量消耗对手的高价值防御资源。美以虽然拥有技术优势,但每一次有效拦截都是一次经济上的亏损。当伊朗持续发动数十轮乃至近百轮打击时,这种经济不对称便转化为战略困境——“拦得起、耗不起”。从攻防理论视角看,这是一种精巧的非对称博弈策略。

3. 非对称升级:将代价外溢至全球经济

伊朗防御适应的第三个层面,体现在其将战争代价外溢至全球经济的能力——封锁霍尔木兹海峡。伊朗通过威胁和实际控制这一全球能源咽喉,将冲突的成本扩展至国际社会。

冲突爆发后,伊朗对霍尔木兹海峡的控制急剧收紧。霍尔木兹海峡承担着全球约五分之一的石油运输。伊朗对霍尔木兹海峡的有效管控,使科威特、卡塔尔等国的原油出口受阻。在此情况下,全球原油价格暴涨,全球石油市场剧烈波动。与此同时,伊朗的军事实力并未如美方宣称般被“彻底摧毁”,仍保有约70%的战前导弹库存。而且,伊朗已恢复了其在霍尔木兹海峡沿线33个导弹基地中30个基地的操作访问权限。^②

这种将冲突升级至经济与能源领域的策略,大幅提高了美以的战争成本,不仅在军事上,更在国际支持与政治合法性层面。从攻防理论视角看,霍尔木兹海峡的封锁体现了防御方通过“扩大冲突维度”来抵消进攻方技术优势的非对称策略。这一策略的运用,使美以面临一个困境:继续军事行动意味着承担不断上升的经济代价和国际压力。

(二) 主观认知偏差:美以三重低估及其代价

如果说伊朗的防御适应是客观层面的抵消机制,那么美以决策者在战前对其反应速度与反击烈度的系统性低估,则进一步加深了战略困境。第二部分已分析了算法确定性偏见如何使决策者对精准、速度、成本形成三重过度确信。过

① 《2万美元无人机消耗400万美元拦截弹 战局或看谁先耗尽弹药》,联合早报网,2026年3月3日, <https://www.zaobao.com.sg/news/world/story20260303-8672922>, 上网时间:2026年5月23日。

② 南博一:《情报显示伊朗仍保留七成导弹水平,美国战争成本已达290亿美元》,澎湃新闻,2026年5月13日, https://m.thepaper.cn/newsDetail_forward_33171067, 上网时间:2026年5月25日。

度确信的另一面,恰恰是对同一组问题的低估(详见表 1)。本部分不再重复机制,而是聚焦于低估的具体表现及后果。

一是对政治后果的低估。算法确定性偏见使决策者将“95%的识别准确率”等同于“100%的战略成功”,仿佛精准命中目标就等同于解决了问题。然而战争的逻辑远非如此。针对最高领袖的“定点清除”,非但未能如预期般瓦解伊朗政权,反而在九天内促成了新领袖的继任,^①并激发了更强硬的报复意志。更关键的是,物理打击的“精准”无法延伸到政治领域:斩首策略非但换不来指挥瘫痪,反而激活了一个去中心化体系的全面反击,更强化了伊朗军民的抵抗意志。精度迷信的代价在于:技术上命中了目标,战略上却产生了更大的麻烦。

二是对适应速度的低估。“秒级响应”是 AI 使用者的优势,美以决策者据此相信伊朗来不及反应就会被击溃。但战争的节奏从来不是单方面决定的。事实证明,空袭开始不到三小时,伊朗的反击已经上路。十二小时内,大规模导弹和无人机扑向以色列及多处美军基地。^②速决战变成了消耗战:AI 压缩的是“从传感器到射手”的时间,然而,令美以没有想到的是,对手从受挫到报复的时间窗口也同样短得惊人。

三是对战争总代价的低估。AI 能精确算出“打掉 N 个目标需要 M 枚弹药”,于是战争看起来“很划算”。但这个账本漏掉了太多东西:政治信誉的贬值、全球能源市场的震荡、供应链的中断、战略资源的持续消耗。当伊朗用两万美元的无人机去消耗数百万美元的拦截弹时,算法意义上的“低成本”在真实的消耗面前显得苍白无力。而这种消耗战的逻辑在于:如果一方能低成本大量生产武器,同时迫使对手花费更多进行防御,那么延长战争本身就是一种施压方式。而当霍尔木兹海峡被封锁、油价暴涨时,战争的经济和政治成本通过全球经济链条回流,最终落在了发起者自己头上。

(三) 攻防理论对“战术胜利、战略失败”悖论的解释

综合以上分析,攻防理论为“战术胜利与战略失败并存”这一悖论提供了系统的理论解释(见表 2)。

① Samriddhi Vij, Kabir Taneja and Akram Zaoui, “The Iran War’s Implications for the Gulf: Three Expert Perspectives,” *ORF Middle East*, March 10, 2026, <https://orfamerica.org/orf-america-comments/the-iran-wars-implications-for-the-gulf-three-expert-perspectives>, 上网时间:2026 年 7 月 11 日。

② “Iran’s Missile Math: \$ 20,000 Drones Take on \$ 4 Million Patriots,” *Business Standard*, March 4, 2026, https://www.business-standard.com/world-news/iran-s-missile-math-20-000-drones-take-on-4-million-patriots-126030401035_1.html, 上网时间:2026 年 7 月 12 日。

表 2 攻防理论对美以“战术胜利、战略失败”的解释

攻防理论维度	具体机制	对结果的影响
客观攻防平衡的动态性	防御适应:分布式韧性(去中心化指挥)、低成本消耗(无人机饱和攻击)、非对称升级(霍尔木兹海峡封锁)	初始进攻优势被抵消
主观认知平衡	算法确定性偏见使决策者对精准、速度、成本形成三重过度确信——由此导致对政治后果、适应速度、战争总代价的系统性低估	低估伊朗的防御适应速度与反击烈度,使战略困境进一步加深

资料来源:作者自制。

从攻防理论视角看,美以的“战术胜利”之所以未能转化为“战略成功”,根源在于:一方面,人工智能赋予的初始进攻优势被伊朗的防御适应所抵消,体现了攻防平衡的动态性,即防御方能在承受初始打击后发展出抵消进攻优势的策略;另一方面,美以决策者的算法确定性偏见导致其对战争进程产生了系统性误判,体现了主观认知平衡的关键作用,即决策者对攻防平衡的感知与实际态势之间的偏离是理解战略困境的关键变量。

这一案例验证了攻防理论的一个核心洞见:技术优势只有在决策者准确评估其边界、并充分考虑对手适应能力时,才能转化为战略成功。人工智能带来的优势可能是短暂的,防御方的适应速度可能比预期更快。在智能化战争中,这一洞见尤其值得注意。

四、结语

本文以 2026 年 2 月 28 日美以联军对伊朗发动的先发制人打击为分析对象,运用攻防理论系统分析了此次智能化战争的发动动因与战略结果。

关于发动动因。本文发现,美以对伊朗的既有军事优势经人工智能赋能后,通过压缩 OODA 循环、提升打击精度、实现多域协同,进一步放大了进攻优势,为美以创造了显著的战术进攻窗口;在主观上,人工智能系统输出的高置信度量化数据制造了“人工确定性”的表象,诱导决策者产生了算法确定性偏见,形成了对精准、速度、廉价的迷思。这两方面因素助推了美以发动先发制人打击的决策。这一发现验证了攻防理论的核心命题:当进攻占优且被决策者主观感知时,先发制人的动机显著增强。AI 赋能军事的情况下,攻防理论和认知心理学相结合所形成的美以决策因果链条如下:既有军事优势(物质基础)→AI 客观赋能(放大固有优势)→错误乐观主义+算法偏见(主观放大)→对战争结果的乐观(由精

准、速度、成本三方面的系统性确信所强化)→“进攻崇拜”→先发制人决策。

关于战略结果。本文发现,美以的“战术胜利”之所以未能转化为“战略成功”,根源在于两个机制的叠加:其一,攻防理论中“攻防平衡的动态性”所揭示的机制,即伊朗通过分布式韧性(去中心化指挥)、低成本消耗战术(无人机饱和攻击)和非对称升级(霍尔木兹海峡封锁)等防御适应策略,有效抵消了美以的初始进攻优势;其二,美以决策者的算法确定性偏见导致其对政治后果、防御适应速度、战争总代价的系统性低估,加剧了战略困境。这一发现揭示了攻防理论的另一个重要洞见:技术优势只有在决策者准确评估其边界、并充分考虑对手适应能力时,才能转化为战略成功。

本文将攻防理论应用于单一案例的分析。研究表明,攻防理论的核心框架即客观攻防平衡与主观认知平衡,在智能化战争场景中仍然有效,且能够解释“战术胜利、战略失败”这一看似矛盾的现象。然而,本文的分析主要依赖公开报道、智库报告和学术文献,无法获取美以决策层完整的内部决策记录。因此,关于算法确定性偏见在美以决策过程中实际作用程度的论证,仍有待更直接的证据支持。

本文在实践上具有启示意义。第一,决策者应警惕对人工智能技术优势的过度自信,避免陷入算法确定性偏见的陷阱。技术优势不等于战略成功,唯有与审慎的战略判断相结合,才能服务于可持续的安全目标。第二,应重视防御方创新的动态性,在军事规划中预判对手可能采取的非对称反制手段。第三,应建立人工智能决策的人机协同纠错机制,确保人类对致命性武力的最终控制权。第四,国际社会应加强对人工智能军事化的风险管控,探索建立相关行为准则与沟通机制,防止技术竞赛引发不必要的冲突升级。

基于本文的研究发现与局限,未来研究可在以下方向深入:第一,若未来有更多解密材料或内部决策记录公开,可进一步检验本文关于算法确定性偏见的核心假设。第二,可将攻防理论应用于其他智能化冲突案例(如无人系统冲突、网络战等),检验该理论在不同技术条件和战略环境下的普适性。

(责任编辑:赵 军 责任校对:李 意)