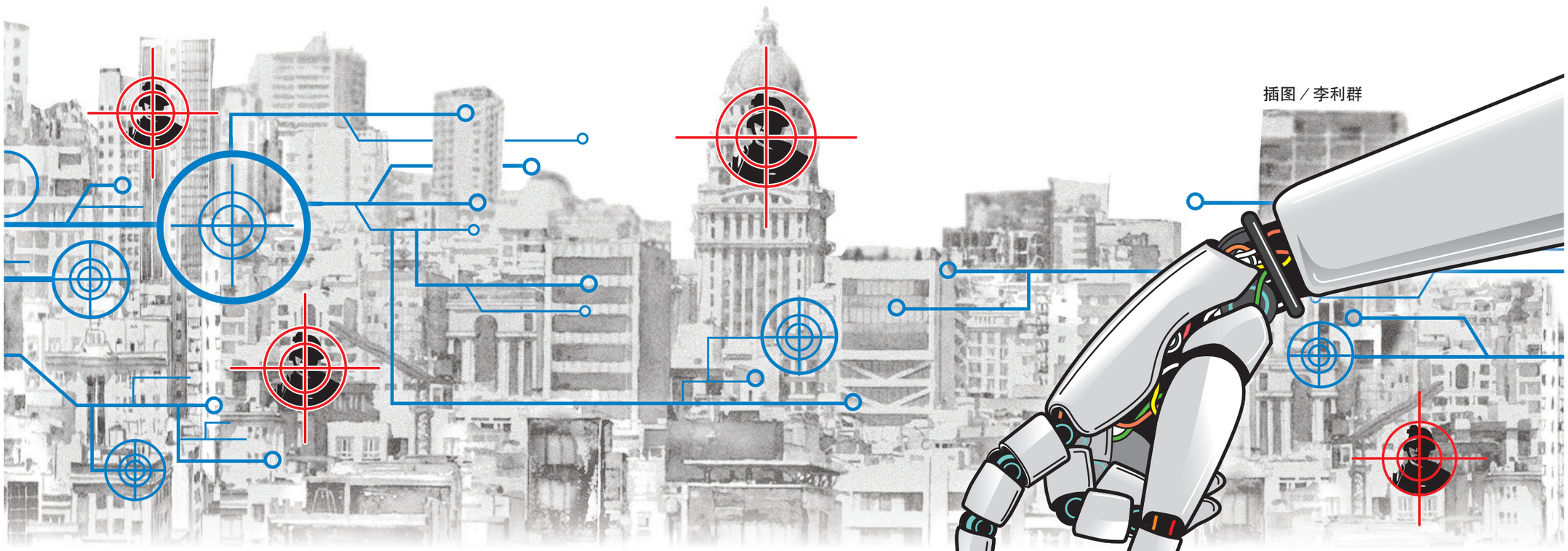
杀敌疾如雷电不讲情理法

彻底放权AI人类恐玩火自焚

在最新爆发的中东战争中，用来筛选和确认攻击目标的人工智能（AI）被美国和以色列军方派上用场，大大提高了情报处理和作战效率，但也引发对AI准确性及出现失误时责任归属的担忧。目前，最终的打击决策仍掌握在人类手中，但由于监管措施未到位，各国若为争夺军事优势而逐渐放权给AI，科幻电影中AI掌控世界与人类生杀大权的末日场景，会否终将成为现实？专家告诫，最安全的做法是永远不要把关乎人类生死的决策权交给AI。



插图／李利群

国际特稿

傅慧云 hweehoon@sph.com.sg

“它无法被说服。它不懂怜悯、悔恨或恐惧。它绝对不会停手，永远不会，直到你死去。”

这是1984年经典科幻电影《魔鬼终结者》（The Terminator）中对杀手机器人的描述。影片讲述了一个AI控制的杀手机器人穿梭时空回到过去，试图杀害未来人类抵抗军领袖之母，以终结人类反抗AI的惊悚故事。

电影上映时，片中场景似乎遥不可及，但随着AI技术被大量应用到最近爆发的中东战争中，人们开始担心影片所描述的未来正迅速逼近现实。

2月28日，美国和以色列对伊朗发动军事行动，美军在短短24小时内打击了1000个目标，速度之快前所未见。

据报道，这是2003年伊拉克战争爆发首日火力的两倍，更远高于1991年为将伊拉克军队赶出科威特而发起的“沙漠风暴”行动开始时，所打击的150个目标。

美军能做到这一点，正是因为它使用了先进AI技术。

据了解，美军主要使用由数据分析巨头Palantir于2018年开发的Maven智能系统，借助AI分析数据，以识别目标并排序优先级。

美军也将AI初创企业Anthropic开发的生成式AI大语言模型Claude，集成到Maven系统，用于处理和汇总前线情报并生成目标，共同构成一个针对伊朗行动的实时数据分析平台。

据报道，这个平台在策划对伊朗的打击行动时，生成了包含上千个选项的潜在目标数据库，并根据战略重要性对目标进行优先级排序。平台还将特定的军事单位和最合适的弹药，例如针对地下设施的钻地弹或针对建筑物的卫星定位制导炸弹，与具体目标自动匹配。它还能模拟战术场景、评估打击的合法性及协助进行战损评估。

美国中央司令部司令库珀（Brad Cooper）3月11日承认美军在与伊朗的战争中使用“多种”AI工具，帮助美军士兵处理海量数据，但他没有透露具体使用了哪些工具。

AI擅长高速处理庞大数据 但若信息缺失会凭空捏造

以色列军方也用了AI，据报它的主要系统为薰衣草（Lavender）与福音（Gospel）。在加沙战争中，以色列的大规模监控系统把加沙等地区居民的数据输入薰衣草系统，由它负责识别和生成人员目标；福音系统则主要负责分析建筑物和地形。

目前不清楚伊朗是否在作战体系中部署了AI，以及具体部署了哪些系统。伊朗在2025年声称将在导弹瞄准系统中应用AI，但据《卫报》分析，受国际制裁影响，伊朗的AI项目与美国和中国这两个AI超级大国相比，规模似乎微不足道。

AI应用到军事上的好处显而易见。它擅长高速处理庞大数据，大幅缩短“杀伤链”（Kill Chain），即把发现目标、人类审批到执行打击的整个决策和规划流程，从过去的几小时或几天，极大地压缩到几分钟甚至几秒钟。这不仅提高效率，也减少人力，把过去可能需要2000名分析师才能完成的工作量缩减至约20人，让人类可专注于更高层级的决策、后勤保障和任务规划上。

AI也已部署在用于侦察和打击任务的无人机、处理爆炸物的地面机器人等，由机器来执行危险任务，进而提升士兵的安全。

不过，事情总有正反两面，AI在军事应用中虽然前景广阔，但在安全性和不可控方面却引发担忧。

首先，AI驱动的系统容易受到黑客攻击和篡改数据，一旦AI平台遭入侵，敏感信息可能泄露，甚至被敌人利用来提供虚假情报。

其次，任何使用过AI聊天机器人的人都知道，它们有时会犯下错误，但又会把错误或误导性的分析说得头头是道，极具说服力。在紧张的战争时期，这尤其可能让指挥官和分析人员更容易采纳它们的建议。

英国伦敦大学学院电脑科学系荣誉教授本特利（Peter Bentley）回答《联合早报》询问时指出，AI大语言模型的最大问题是，当面临信息缺失时，它们会“产生幻觉”，即凭空捏造不存在的结果。

他说：“我们无法判断它们的结论是虚构的还是基于现实，因为AI非常擅长呈现看似合理且逼真的结果。也就是说，如果你根据AI的分析结果来瞄准目标，很可能无法正确打中。在最糟糕的情况下，这可能意味着无辜平民的丧命，而这仅仅是因为有人太相信AI。”

AI只看数据不讲伦理 人命关天人类责无旁贷

新加坡南洋理工大学拉惹勒南国际研究院（RSIS）研究员哈贾尼（Manoj Harjani）也告诉《联合早报》，在军事领域使用AI的一大担忧在于难以预测AI驱动系统的行为方式，以及难以理解其行为背后的原因。“这给问责制和责任承担带来了挑战。”

换句话说，如果用AI生成的错误结果导致了误判，那应该是由AI还是人类来承担责任？

伊朗战争开打首日，一枚导弹击中伊朗南部米纳布市一所女子小学，据报造成约170名孩童死亡。美国媒体引述知情者说，原因是美方使用了过时的数据，从而制定错误的打击坐标。

美方至今未承认是其导弹击中学校，只表明正在调查。库珀在承认美军使用AI作战时强调，美军是否以及何时发动军事打击，最终决定权始终掌握在人类手中。若以此为依据，无论美方调查结果如何，造成人员伤亡的责任方将是人类而非AI。

本特利在谈到如何负责任地使用AI时指出，归根结底，人类还是应对人类的生命负责。“即使我们使用工具来帮助了解局势，也须亲自核实这些工具的有效度；如果我们真的必须相互厮杀，那应该由人类自己来执行，而不是把责任推卸给机器。”

展望未来，在军事领域运用AI的趋势已势不可挡，而分析认为，在AI领域取得决定性优势的国家将掌控未来冲突的节奏，因此，21世纪军事竞争的核心将是争夺AI主导权。

但与此同时，人们必须在作战效率和伦理责任之间寻求平衡，除了加大对AI技术的投资，更须建立保障措施、法规和故障保护机制，以确保AI的使用是负责任的。

AI军用监管尚无定案 负责任使用挑战巨大

就在美军采用Anthropic开发的AI模型打击伊朗之际，五角大楼竟然与Anthropic就军事AI的安全红线起了争执。

争端起因是五角大楼要求Anthropic签署合同，允许军方“不受限制”地把Anthropic技术用于“所有合法用途”。

但Anthropic首席执行官阿莫代伊（Dario Amodei）坚持，合同需列明两个例外情况：不得对美国公民进行大规模监控，以及不得开发完全自主且无人监管的武器。

Anthropic的“不识时务”，引发白宫强烈不满。2月27日，也是美军袭击伊朗前一天，美国总统特朗普指示所有联邦机构立即停止使用Anthropic技术；战争部长赫格塞斯也要求把Anthropic及其产品，列入对国家安全构成威胁的“供应链风险”名单。这个极其罕见的措施，历来只用于外国对手的企业。一旦被列入名单，战争部将禁止它的所有承包商与Anthropic进行商业交易。

Anthropic和五角大楼的争执核心之一是自主武器。

自主武器面世只是迟早问题 人类是掌控它还是反被掌控？

自主武器（Lethal Autonomous Weapon Systems）是指无需人类干预或操控，即可通过AI独立搜寻、识别和攻击目标的武器系统。批评者也把自主武器称为“杀手机器人”，这也是人们对《魔鬼终结者》里主角阿诺·施瓦辛格饰演的T-800型终结者机器人的称呼。

目前人类已有很多具备某些自主功能的武器，例如：可自行识别、选择和攻击目标的导弹；可自行排雷的无人潜艇；可组成蜂群网络并独立执行任务的无人机。当中，无人机已大量应用到乌克兰战场上，但这些无人机并非完全自主，另一端仍有人类在远程操控。

Anthropic坚持美国战争部不得在完全自主且无人监管的武器系统上使用它的AI，谈判破局后遭受严厉惩罚。这张示意图显示了Anthropic与美国战争部的标志。（路透社）

随着AI和机器人领域的迅速发展，完全自主的武器系统面世，只是时间问题。但令人不安的是，目前尚无专门的国际条约规范在武装冲突中使用AI，更别说是管制自主武器了。

不过，国际科学期刊《自然》网站在3月10日发表的社论指出，国际人道主义法明确规定，武器不得滥用。此外，作战人员必须采取预防措施，核实目标，并将平民伤亡风险降至最低。这些要求同样适用于AI，如同适用于任何其他军事技术一样。

RSIS军事转型项目副研究员刘美均回答《联合早报》询问时指出，国际社会其实已讨论规范军事AI应用的问题10多年。“自2016年以来，联合国致命性自主武器系统政府专家组一直是这些讨论的主要平台。”

2026年3月3日伊朗南部米纳布一所小学遭美军空袭，导致至少175人死亡，当中绝大多数是学生，震惊世界。这张由伊朗新闻中心发布的航拍照片显示，哀悼者正在为遇难儿童挖掘坟墓，准备下葬。（法新社）

刘美均说，专家组目前获授权制定一套框架文本，以解决自主武器引发的问题。虽然文本的具体内容未定，但授权范围并不包括针对一项具法律约束力的条约进行谈判。专家组的任期将于今年底届满，到时将由《特定常规武器公约》缔约国决定是否延长任期，以及延长后的具体授权范围。可以肯定的是，国际社会无法在今年达成一项具法律约束力的协议。

刘美均透露，AI在军事领域的更广泛应用目前也在其他论坛上进行讨论，包括负责任AI全球委员会（REAIM）和联合国大会第一委员会。当中第三届REAIM峰会于2月4日至5日在西班牙举行，峰会成果文件《行动路径》仅获得约40个国家认可，认可率远低于前两届。

刘美均说：“这份成果文件不具法律约束力，但发出重要的政治信号，表明各国承诺遵守在军事领域开发和使用AI的具体原则。支持率下降确实令人担忧。”

哈贾尼则认为，国际社会关于军事AI治理的讨论，即使在约束国家行为方面可能显得力度不足，却依然是各国以及私营部门等各方交流观点的重要途径。“这些交流的目的应该是增进相互理解，并鼓励制定负责任地在军事领域使用AI的规范。”

> 若不想失去对AI的控制，最好的做法是不要把我们的生命托付给这些非人类的智能体。在AI军备竞赛中，即使看到其他国家以更高速度推进，也不应放弃对AI的掌控，更必须在必要时介入甚至踩下刹车。
>
> ——英国伦敦大学学院电脑科学系荣誉教授本特利

专家吁求各国坚持对AI掌控

管制措施尚未到位，但各国为了争夺军事优势，想来也不会停下开发更强大和更自主武器的脚步。到了科技成熟时，如果将决定人类生死的最终权力交给缺乏情感和道德观的机器，是否会引发不必要的杀戮？到时，人类是否还能驾驭AI，抑或被AI掌控？

若不想失去对AI的控制，本特利告诫，最好的做法是不要“把我们的生命托付给这些非人类的智能体”。

他以驾驶火车为比喻指出，在AI军备竞赛中，即使看到其他国家以更高速度推进，也不应放弃对AI的掌控，更必须在必要时介入甚至踩下刹车。

他强调：“不要跳下驾驶座，继续驾驶，也不要害怕踩刹车。仅仅因为你看到其他AI‘失控列车’以惊人的速度行驶，并不意味着你必须效仿它们。”