探索人工智能和自主系统的合成数据

入门指南

HARRY DENG

致谢

裁研所核心资助者的支持为裁研所的所有活动奠定了基础。本出版物由欧盟资助，是裁研所安全与技术计划的一部分，该计划得到了捷克、法国、德国、意大利、荷兰、挪威、瑞士和联合王国政府以及微软公司的支持。作者感谢Giacomo Persi Paoli博士和Ioana Puscas博士为本文提供的建议和帮助，感谢Tim Watson教授和Leslie Sikos博士为本研究做出的宝贵贡献。

关于裁研所

联合国裁军研究所（裁研所）是联合国内部一个由自愿捐款供资的自主机构。裁研所是世界上为数不多专注于裁军的政策研究所之一，提供裁军与安全方面的知识，并促进这方面的对话和行动。裁研所总部设在日内瓦，协助国际社会提出切实可行的创新想法，以找到解决重大安全问题的办法。

引文

H. Deng，《探索人工智能和自主系统的合成：入门指南》，瑞士日内瓦：裁研所，2023年。

注

本出版物所使用的名称和材料的编排方式并不意味着联合国秘书处对任何国家、领土、城市或地区或其当局的法律地位，或对其边界或界线的划分表示任何意见。出版物中表达的观点仅由作者本人负责，不一定反映联合国、裁研所、其工作人员或赞助者的观点或意见。

目录

关于安全与技术方案

当代科学技术的发展为国际安全与裁军带来了新的机遇和挑战。裁研所的安全与技术方案（SecTec）力求建立对特定技术创新的国际安全影响和风险的了解和认识。该计划召集利益攸关方探讨各种想法，并就如何应对这些想法形成新思维。

关于作者

Harry Deng 是裁研所安全与技术方案的顾问，他的工作重点是新兴技术对国际安全的影响。他持有滑铁卢大学全球治理硕士学位，目前是该校的博士生。请在推特上关注 Harry @hwrdeng。

缩写与缩略语

AI	人工智能
GAN	生成式对抗网络
GGE	政府专家组
ICT	信息和通信技术
IoT	物联网
OEWG	不限成员名额工作组
VAE	变分自动编码器

内容提要

近年来，人工智能（AI）和机器学习领域的进步为增强人类能力和改善各种自主系统的功能，包括在国际安全领域，带来了前所未有的机会。然而，在防卫领域，用于训练日益复杂的人工智能系统的高质量、高度多样化和相关真实世界的数据集却十分稀缺。因此，合成数据正逐渐成为开发和训练人工智能系统的数据工具箱中必不可少的工具。合成数据的特点和潜在优势，以及该技术在各个领域的成熟应用，使其成为围绕在国际安全背景下使用人工智能展开辩论的一个重要话题。

本文简要概述了合成数据，包括其特征、生成方式、附加值、风险以及其在防卫组织和军事行动中的潜在用例。此外，本文还概述了现有数据面临的挑战和限制，这些挑战和限制促使合成数据成为开发日益复杂的人工智能系统的重要工具。

迄今为止，合成数据在国际安全领域的使用大多停留在实验和探索阶段。然而，合成数据的特点可能会对人工智能系统的训练产生有利影响。特别是，合成数据可以生成高度多样化甚至新颖的数据集，对数据属性进行精细控制，在必要时自动注释或标注数据，并具有成本效益。本文探讨了合成数据的主要特点如何使军方和防卫组织受益，使其能够在防御和进攻型自主系统中集成更强大、更可靠的人工智能系统。

虽然合成数据有利于训练人工智能系统，并有助于缓解军方和防卫组织面临的一些数据问题，但它并非灵丹妙药，也伴随着风险和挑战。使用合成数据所带来的益处将取决于各组织驾驭这些风险的能力，取决于其是否能以负责任和安全的方式，并按照法律要求和道德价值观使用通过合成数据训练的人工智能系统。

导言

人工智能（AI）以及支持其使用的机器学习模型的进步，使其广泛应用于优化性能以应对日益复杂的任务和工作环境。在国际安全领域，这一点尤为重要，因为人工智能的整合带来了前所未有的法律、伦理、安全和安全挑战。在国际安全领域，[1]人们正在探索将人工智能用作决策支持、行动规划和情报分析的工具，人工智能可集成到进攻和防御型自主系统中，如目标识别系统、集群机器人技术和网络行动。事实上，有人认为，人工智能的使用在某些任务中的表现要优于传统方法——例如，在提高防御性网络基础设施的稳健性或加强情报分析方面[2]——这意味着各国除了提高行动效率外，还能更好地履行其国际法律义务，特别是在国际人道主义法中的义务。

与此同时，为人工智能设想的任务所产生的下游效应意味着，机器学习模型需要日益多样化和高速的优质数据，通常是优质的标记数据。如果没有所需的多样化的大量优质数据来训练复杂的人工智能系统，这些系统可能会出现更多故障，包括意外伤害。标记数据会明确告知机器学习模型数据的含义，而不是让模型自己去理解数据的含义，这样可能会出错。然而，优质的真实世界数据非常稀缺，再加上敏感数据相关的隐私、法律、监管和成本挑战，使其通常不适合用于训练日益复杂的人工智能系统，[3]尤其是在国际安全领域。正是由于优质的真实世界数据的稀缺，合成数据逐渐成为开发、改进和训练日益复杂的人工智能系统的重要工具，特别是在没有数据的领域提供数据，抵消各种形式的偏差，以及在必要时自动标记数据等等。[4]

然而，在联合国相关安全进程中，如致命自主武器系统领域新兴技术问题政府专家组（GGE on LAWS）或信息和通信技术安全和使用安全不限成员名额工作组（OEWG on ICT），仍未探讨在自主系统中使用合成数据的影响。事实上，与合成数据相关的附加值和风险与这些讨论以及围绕在国际安全领域使用人工智能的其他辩论息息相关。例如，参与致命自主武器系统领域新兴技术问题政府专家组辩论的一些方面担心，由于缺乏对武器系统进行适当培训的培训数据，此类系统自主性的提高可能会导致意外伤害的增加。[5]此外，信息和通信技术安全和使用安全不限成员名额工作组的与会者还讨论了这样一种可能性，即人工智能支持的网络攻击可以自主适应防御性网络

[1] 联合国第一委员会对“国际安全”的定义：影响国际社会的全球性挑战和对和平的威胁。见联合国大会，“裁军与国际安全（第一委员会）”，https://www.un.org/en/ga/first/。

[2] 见 Alex Wilner，“人工智能与威慑的未来：承诺与陷阱”，国际治理创新中心，2022 年 11 月 28 日，https://www.cigionline.org/articles/ai-and-the-future-of-deterrence-promises-and-pitfalls/。见国防创新委员会，“人工智能原则”，国防创新委员会，2019 年，https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF。

[3] 见 Jie Yan 等，“用于人体检测的合成数据集生成与适配”，DEVCOM 陆军研究实验室，2020 年 11 月，https://apps.dtic.mil/sti/pdfs/AD1115446.pdf。 见 Arthur Holland，“已知的未知：数据问题与军事自主系统”，裁研所，2021 年 5 月 17 日，https://unidir.org/known-unknowns。见政府商业理事会，“推进边缘的情报、监视和侦察：数字作战空间网络和处理技术调查”，2020 年 7 月，http://cdn.govexec.com/media/advancing-isr-at-the-edge-isr.pdf。

[4] 见 Jie Yan 等，“用于人体检测的合成数据集生成与适配”，1。见 Holland，“已知的未知”，27。

[5] 见巴基斯坦，“关于致命自主武器系统（LAWS）国际文书的提案”，CCW/GGE.1/2023/WP.2/Rev.1，2023 年 3 月 8 日，https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.3_REv.1_0.pdf。见巴勒斯坦国，“巴勒斯坦国关于自主武器系统规范和行动框架的建议”，CCW/GGE.1/2023/WP.2/Rev.1，2023 年 3 月 3 日，https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.2_Rev.1.pdf。

措施，使其更难被发现和防范。[6]可以肯定的是，可以通过生成和使用合成数据来实现和增强人工智能支持的网络攻击。因此，至关重要的是，在自主系统中使用合成数据不会减损对国际法或负责任的人工智能的任何承诺，[7]即与数据质量、安全性、公正性以及人类监督、判断或控制有关的承诺。8

因此，本入门指南旨在为参与国际安全讨论的政策制定者和外交官提供有关合成数据的介绍性概述，包括其主要特点、附加值、风险以及在国际安全领域的重要性，特别是作为自主性的促进因素。本入门指南进一步试图说明合成数据日益增长的重要性，以及在国际安全领域中数据使用和治理范式的演变。为此，本入门指南阐明和描绘了合成数据的特殊性，然后将其与现有的数据挑战重新联系起来。

1. 了解合成数据

要点

- 真实世界数据指来自真实世界的数据和输入，与其不同的是，合成数据是在数字世界中人为创建的数据，通常旨在再现现有数据集的特征和属性，或根据现有知识生成数据。
- 合成数据的目的是提高训练数据集的质量和实用性。至关重要的是，训练自主系统所用的数据要有足够的质量和多样性，以避免意外伤害，特别是在国际安全领域。
- 防卫组织目前在数据管理方面面临着无数挑战，从而限制了用于训练日益复杂的人工智能和自主系统的真实世界数据的质量和实用性。
- 虽然合成数据可能不是解决防卫组织内现有数据挑战的灵丹妙药，但可以提供一种提高训练数据集质量和实用性的方法。

1.1 什么是合成数据？

合成数据是在数字世界中人工生成的数据，其属性通常来自于“原始”数据集。这与真实世界数据截然不同，顾名思义，后者是从真实世界的事件和输入中收集的数据。“原始”数据集通常指真实世界的数据和信息，但也可以是人工数据本身。虽然生成合成数据集的方法多种多样（本文第2.3节详述），但目标往往是重现原始数据集的特征和结构，而大多数方法都依赖于从原始数据中提取和复制属性。[9]这意味着，在进行相同的统计分析时，合成生成的数据和原始数据的结果即

[6] 见 Hoda Alkhzaimi 教授，“纽约大学/纽约大学阿布扎比新兴研究与安全中心对第五届实质性会议的供稿”，非政府组织工作文件，2023 年 7 月 28 日，https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/Stakeholder_Recommendation_for_Open-ended_workinggroup_on_security_APR.pdf。

[7] “负责任的人工智能”指的是一种广泛的方法，旨在确保人工智能系统在开发和使用过程中是合乎法律和道德、安全、可靠且负责任的。见 Alisha Anand 和 Harry Deng，“探索防卫工作中负责任的人工智能：各国人工智能原则的梳理与比较分析”，裁研所，2023 年 2 月 13 日，https://unidir.org/publication/towards-responsible-ai-defence-mapping-and-comparative-analysis-ai-principles-adopted。

[8] 同上。

[9] ubarmaniam Kannan，“用于边缘分析的合成时间序列数据生成”，F1000 Research，2022 年 1 月 20 日，https://doi.org/10.12688/f1000research.72984.1。

使不完全相同，也应该非常相似。[10]简而言之，合成数据通常是人为生成的信息，用来代表其希望替代的原始数据，从而产生等效功能，或者用来补充原始数据，从而提高训练数据集的价值。不过，也可以通过生成合成数据来增强训练数据集，这种合成数据不会再现原始数据集的特征，但会夸大某些特征（本报告下文将对此进行更详细的讨论）。

坦克的真实世界图像：

合成的坦克图像：

图1.现实世界数据与合成数据[11]

然而，在某些情况下，合成数据也可以是不依赖原始数据集的人为生成数据。根据现有知识，亦可生成新数据。例如，可以根据现有的物体物理知识，合成不同重量的骰子的运动表现数据。在这些情况下，合成数据不是再现原始数据集的特征，而是会产生反映假设产生该数据的系统特征的数据。

1.2 现有数据挑战

防卫组织内部的数据挑战不仅是技术挑战，也是组织挑战。[12]这就意味着，防卫组织不能简单地用技术解决方案来找到克服自身不足的办法。相反，要应对防卫组织内部的数据挑战，除了要找到技术解决方案外，还应考虑组织文化、政策和程序的影响。归根结底，自主系统的任何使用，特别是在作战环境中使用的自主系统或打算攻击人类目标的自主系统，都必须承担起预测和应对数据问题的责任，以避免意外伤害。虽然合成数据可能不是缓解所有现有数据挑战的灵丹妙药，但它可以提供一种提高训练数据集质量和实用性的方法，特别是在数据问题可能不容易暴露的日益复杂和不透明的机器学习模型中。

1.2.1 专题1——数据管理

[10] Robert Riemann，“合成数据”，欧洲数据保护监督机构，https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en

[11] 见 Frank Longford，“合成数据实验”，《法证建筑》，2018 年 11 月 6 日，https://forensic-architecture.org/investigation/experiments-in-synthetic-data。

[12] 见政府商业理事会，“推进边缘的情报、监视和侦察”。

导致缺乏足够优质的标记数据的第一类问题是数据管理不善。数据管理过程包括以下几个阶段：

1. 收集
2. 处理（如数据标记）
3. 存储
4. 访问
5. 共享和传播
6. 提取和挖掘
7. 利用
8. 处置

美国国防部指出，数据的处理、挖掘和传播尤其具有挑战性。在一项研究中，只有 29%的现役军人和文职人员表示，75%以上的数据能够传递到适当的行为者手中。[13]在现役军人中，这一比例更不乐观，只有 11%的人表示至少有 75%的时间能将数据传递到适当的分析人员。[14]此外，65%的现役人员指出，作战人员花在寻找正确数据上的时间多于使用数据的时间[15]——这表明在建立适当流程以正确标记数据、将数据存储在适当的数据库中以及确保适当的访问途径和可用性方面存在不足。这也可能表明，既要保护敏感或机密数据，又要与那些可能从挖掘这些数据中获益的人共享这些数据，两者之间很难取得平衡。

事实上，国防部人员指出，孤立的数据、存在多个且往往相互排斥的安全域、有限的带宽和有限的数据标记，是影响其组织有效收集、传播和分析数据的能力的一些最普遍的挑战。[16]在数据标记方面，只有 32%的防卫文职人员和 13%的现役人员表示，其所在的防卫机构拥有有效标记数据的系统。[17]监测和管理传入数据所需的人员数量以及必要的流程和基础设施跟不上日益增长的数据量。缺乏事前或事后的数据质量控制意味着分析人员被淹没在数据中，而当获得正确的数据集时，这些数据往往已经过时且不可靠。
澳大利亚 2021 年国防数据战略[18]、联合王国 2021 年国防数据战略[19]、加拿大 2021 年国防部数据战略[20]以及印度尼西亚国防大学开展的信息网络研究也提到了类似的数据管理挑战。[21]例如，数据可见性方面的挑战、孤立的数据、组织内部和组织之间缺乏共同的数据标准，以及在能力开发的初始阶段未考虑数据要求的文化问题等，都是这些文件中发现的共同问题。

1.2.2 专题 2——数据质量

数据管理不善和数据素养偏低造成了第二类问题——数据质量差。常见的数据质量问题包括：不完整数据、未标记数据、中毒或欺骗数据、不准确数据、数据偏差和差异数据。虽然数据质量差

[13] 同上，4。
[14] 同上。
[15] 同上，8。
[16] 同上，15。
[17] 同上。
[18] 澳大利亚，“2021-2023 年国防数据战略”，澳大利亚国防部，https://www.defence.gov.au/about/strategic-planning/defence-data-strategy-2021-2023#:~:text=The%205%20pillars%20in%20the,capability%20within%20the%20Defence%20workforce。
[19] 联合王国，“国防数据战略：实施国防数据框架并充分利用数据的力量”，2021 年 9 月 27 日，国防部，https://www.gov.uk/government/publications/data-strategy-for-defence/data-strategy-for-defence。
[20] 加拿大，“国防部和加拿大武装部队数据战略”，国防部，2021 年 5 月 18 日，https://www.canada.ca/en/department-national-defence/corporate/reports-publications/data-strategy/data-strategy.html。
[21] Putu Aryawan Udayana 等人，“印度尼西亚国家军队综合土地信息系统网络安排战略”，2022 年，https://doi.org/10.33172/jspd.v8i1.1054。

可能是外部因素造成的，比如恶劣条件（如灰尘、烟雾、振动、污染物、伪装、传感器磨损等）和敌对行动（如信号干扰、数据投毒、攻击传感器、意外战术等），但适当的数据管理方法可以帮助过滤掉受损数据，避免在训练数据集中产生曲解或偏差，并确保正确的数据传递到适当的机构。

不同的自主系统会面临不同类型的数据质量差的问题。例如，用于防御性网络行动的自主系统不太可能面临恶劣条件（如灰尘、烟雾、污染物等）带来的问题，但很可能面临欺骗或数据投毒等敌对行动。另一方面，在“不可控”的多变量作战环境中，无人驾驶车辆可能会面临恶劣条件和敌对行动。

如果自主系统依赖于训练它们时使用的数据来在环境中进行导航、响应和操控，那么训练它们时使用的数据必须具有足够的质量和多样性。[22]然而，值得注意的是，并非所有的人工智能系统都依赖于数据的训练；也可以使用强化学习模型。[23]强化模型的工作原理是利用奖励函数来认识所采取行动的后果，而不是利用训练数据集。

然而，在任何大型真实世界数据集中，都应该想象到会存在一定数量的劣质数据，尤其是在国际安全领域，无论是数字空间还是实体空间，敌对环境都对收集完整的优质数据构成了广泛的挑战。[24]因此，有人提出，合成数据可以在减轻收集优质真实世界数据的某些压力方面发挥重要作用，例如，填补因传感器故障而缺失的数据。[25]

1.3 生成合成数据的方法

可以利用决策树或深度学习算法等各种技术来完成生成合成数据的过程。作为替代数据，合成数据可按原始数据的三种类型之一进行分类：

- 真实世界数据
- 开发人员的信息或知识
- 真实世界数据与开发人员信息的结合

如前所述，合成数据的生成依赖于提取和复制原始数据集的属性。提取和复制原始数据集属性的方法取决于原始数据的类型和结构。合成数据的生成方法主要有三种：基于规则的方法（包括预定义的数据结构）、基于代理的模型（模拟可能需要数据的环境）和深度学习算法（使用基于神经网络的方法）。下面的小节将简要解释每个系列的方法。

1.3.1 基于规则的方法

[22] Arthur Holland，“已知的未知”，3。
[23] 与Tim Watson 教授的采访，2023 年 4 月 11 日。
[24] Holland，“已知的未知”，6。
[25] 与Tim Watson 教授的采访，2023 年 4 月 11 日。

[26]基于规则的方法相当于元数据和人机可读数据，这些数据由预定义的数据结构（如数组）组成，提供遵循人类定义的特定规则的有序列表和对象。这些规则的复杂程度各不相同，既有只考虑列中指定数据类型的简单规则，也有定义多个参数和变量之间关系的更复杂规则。常见的数据格式包括逗号分隔值（CSV）、JavaScript 对象表示法（JSON）和文档类型定义（DTD）。基于规则的方法具有模块化、成本效益高的特点，而且可以支持不同的统计分布，这对于训练网络行动中使用的人工智能系统尤为重要，因为在网络行动中，标准化协议需要预定义的数据结构。

基于规则的合成数据生成方法已被应用于国际安全以外的其他领域。例如，Kannan（2021 年）使用 JSON 结构生成了一个由空气质量指数（AQI）数据集衍生的合成数据集。在这项特殊研究中，Kannan 制作了一个合成的 AQI 数据集，在训练机器学习模型预测数据集中指定的四项 AQI 参数时，该数据集的表现优于原始的 AQI 数据集。[27]Kannan 得出结论，合成数据集表现更好，可能是因为其“填充”了原始数据集中的不完整数据。[28]

然而，使用基于规则的方法生成合成数据集也有局限性。最显著的挑战是可扩展性、漂移和偏差。[29]首先，在可扩展性方面，合成数据集越复杂，例如，如果合成数据集需要成千上万条相互依存、相互交织的规则，那么该合成数据集的生成就越复杂、越深奥。因此，对于更为错综复杂的关系网络而言，使用基于规则的方法的实用性受到了限制。与此相关的是，数据漂移，即数据分布随着时间的推移而变化，可能会限制基于规则的方法的实用性，尤其是在没有完善的变更管理来规范如何改变规则以适应其应用的情况下。最后，由于规则是由人类定义的，开发人员的偏见会反映在生成的数据中，无论是有意识的（如商业逻辑）还是无意识的（如性别偏见[30]）。

1.3.2 基于代理的建模

基于代理的建模是一种行之有效的模拟技术，在现实世界中有着广泛的应用，从解决商业问题到公共政策评估。基于代理的建模本质上是一个描述代理和代理之间关系的系统，目的是得出结果。这些代理能够根据它们之间的相互作用、行为模式和输入的参数不断进化，从而出现意想不到的行为。[31]这使得基于代理的建模尤其适用于捕获涌现现象——即使是简单的基于代理的模型也能展现出复杂的模式，并提供有关真实世界动态的有价值信息。[32]深度学习还可被纳入基于代理的模型，以实现代理之间更加动态的相互作用，以及更加真实、复杂和适应性更强的结果。

1.3.3 深度学习算法

---

[26] 在编程中，数组是一种数据结构，由值和/或变量（如数字、单词、对象等）的集合组成，并根据其类型进行格式化和分类。数组的作用是将多个相同类型的数据存储在一起。
[27] Kannan，“合成时间序列数据生成”，7。
[28] Kannan 指出，原始空气质量指数数据集中的数据不完整，是因为其中一个站点的传感器故障导致只记录下一部分数据。见，同上。
[29] Manuel Pasieka，“合成数据生成方法和合成数据类型比较”，2022 年 9 月 1 日，https://mostly.ai/blog/comparison-of-synthetic-data-types/。
[30]见 Katherine Chandler，“军事人工智能是否有性别之分？理解偏见并在人工智能的军事应用中推广伦理方法”，裁研所，2021 年 12 月 7 日，https://doi.org/10.37559/GEN/2021/04。
[31] Eric Bonabeau，“基于代理的建模：模拟人类系统的方法和技术”，《美国国家科学院院刊》，2002 年 5 月 14 日，https://doi.org/10.1073/pnas.082080899。
[32] 同上。

深度学习算法是一类基于“表征学习”的方法，是指自动学习训练数据的特征和统计分布，并能根据这些学习到的特征和统计分布生成新数据的机器学习技术。与生成合成数据的基于规则的方法或基于代理的模型不同，深度学习算法中人工指导和监督可能是最少的，甚至不存在，这取决于所使用的深度学习模型。此外，深度学习模型不受其可学习数据复杂程度的限制，理论上，应用深度学习生成合成数据集是“无限制的”。[33]与基于规则的方法相比，深度学习算法可以管理更加错综复杂的数据分布，并能合成非结构化的数据，如可视化数据。

深度学习技术主要有三个系列：

1. 生成式对抗网络（GAN）

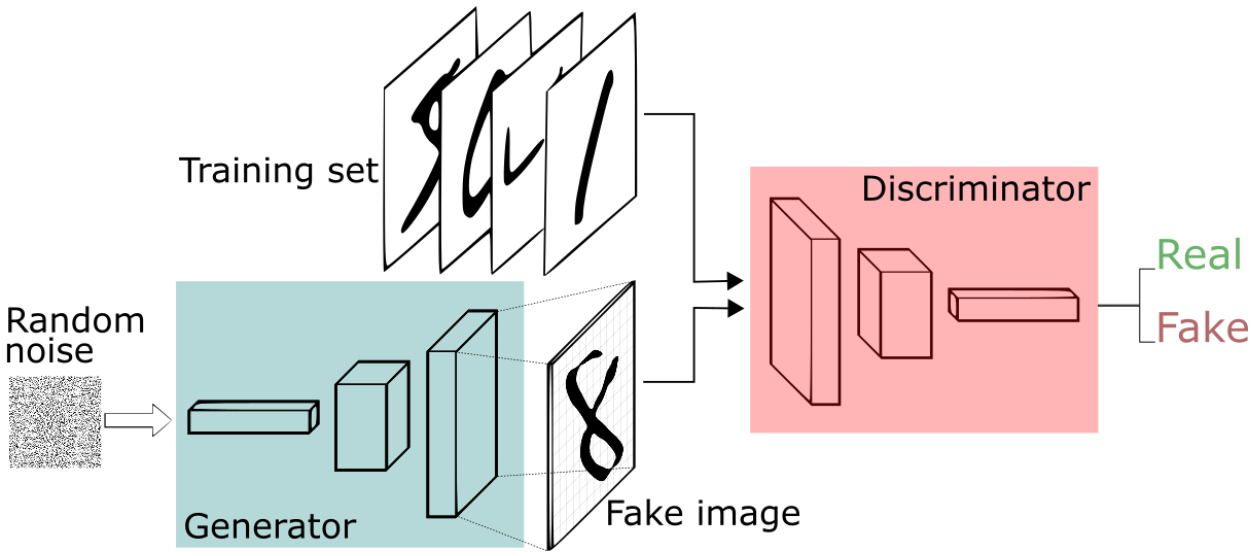


图2.生成式对抗网络[34]

GAN 通常用于图像识别和图像生成。[35]GAN 通常由两个神经网络[36]（生成器网络和判别器网络）组成，这两个网络在迭代的基础上相互训练。生成器网络将生成一个与训练数据具有相同特征的合成数据点（如图像）作为输入，然后，包含成批训练数据和合成数据的判别器网络将尝试把观测结果分为真实结果或生成结果。生成器网络会根据从判别器网络获得的反馈，不断改进自身表现。当判别器无法再区分“真实”数据和合成数据时，这两个网络就达到收敛状态。[37]

[33] Pasieka，“合成数据生成比较”。
[34] Thalles Silva，“生成式对抗网络简介”，Thalles 的博客，2017 年 6 月 7 日，https://sthalles.github.io/intro-to-gans/。
[35] Riemann，“合成数据”。
[36] 神经网络又称人工神经网络，是一个由相互连接的节点层组成的网络，这些节点将信息从一层传递到另一层，每一层对其输入执行不同的功能。神经网络依靠训练数据来学习，其性能会随着时间的推移而提高。见 IBM，“什么是神经网络？” https://www.ibm.com/topics/neural-networks#:~:text=Neural%20networks%2C%20also%20known%20as,neurons%20signal%20to%20one%20another
[37] Jiri Hradec 等人，“多用途合成人口或政策应用”，欧盟委员会联合研究中心，2022 年 4 月 13 日，14，https://dx.doi.org/10.2760/50072。

2. 变分自动编码器（VAE）

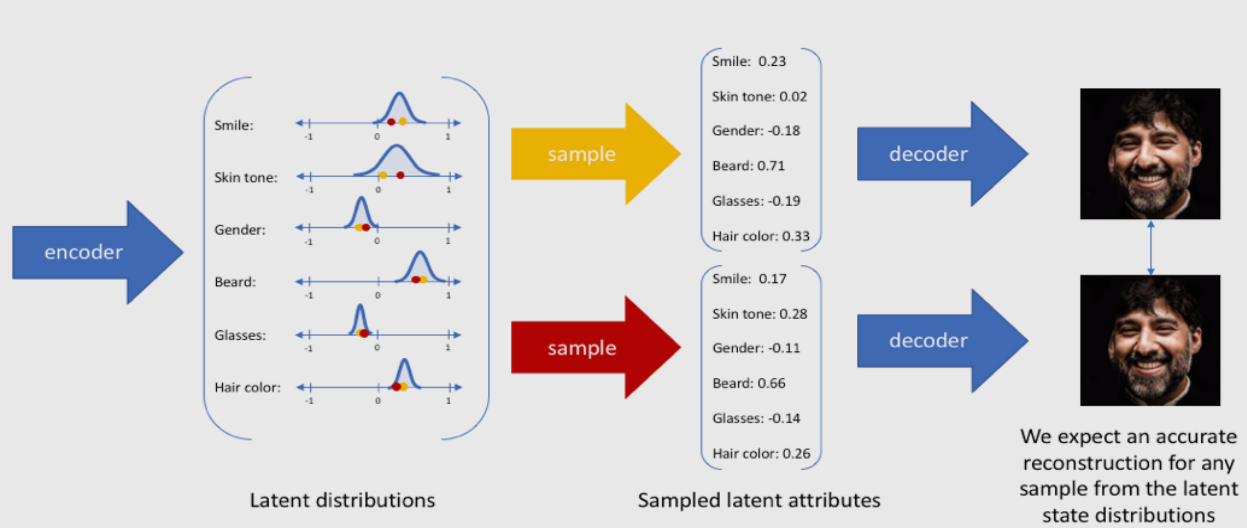
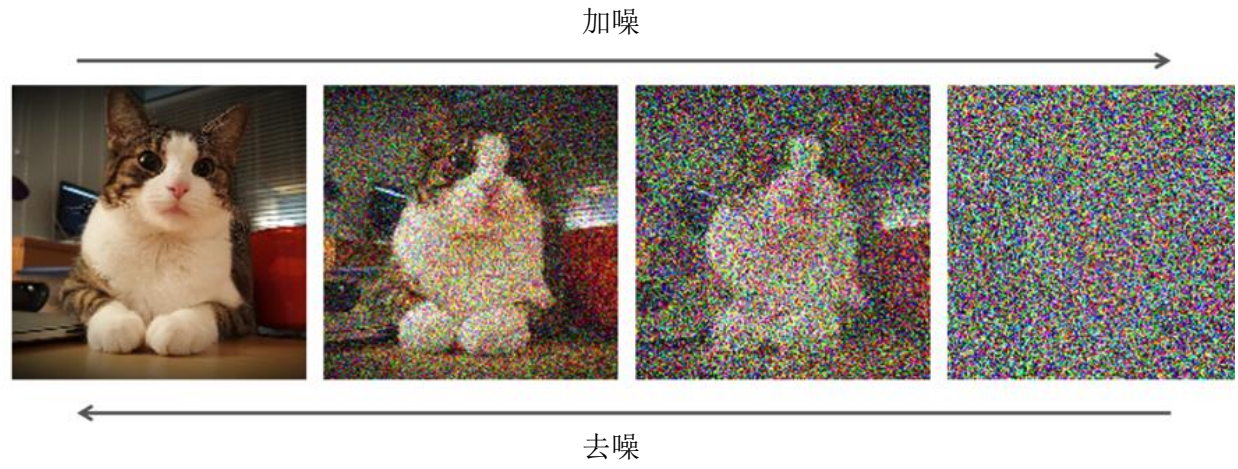


图3.变分自动编码器[38]

VAE 是一种基于似然法的生成模型。VAE 由编码器和解码器组成，编码器摄取数据并对其进行简化（即潜在表示），以表示数据的关键特征；解码器接收潜在表示并返回对其的重构。与 GAN 一样，VAE 也在迭代的基础上运行。每次迭代时，VAE 都会摄取数据，然后将其与编码器-解码器输出进行比较。因此，VAE 的基本功能是学习最佳的编码-解码方案，以迭代优化流程。所以，更复杂的 VAE 架构可以支持更高的降维（即学习关键特征），同时保持较低的重构误差。 [39]

3. 扩散模型



[38] Jeremy Jordan，“变分自动编码器”，2018 年 3 月 19 日，https://www.jeremyjordan.me/variational-autoencoders/。
[39] Diederick P. Kingma 和 Max Welling，“变分自动编码器简介”，《机器学习的基础和趋势》，2019 年，https://arxiv.org/pdf/1906.02691.pdf。

图4.扩散模型[40]

扩散模型是一类新兴的深度学习模型，通过迭代去噪过程从训练分布中生成数据（如图像）[41]。换句话说，扩散模型的工作原理是破坏图像，例如增加噪声，然后模型学习如何去除噪声（或去噪），以生成连贯的图像。然后，扩散模型可以通过在原本连贯的图像中加入不同的噪声，使图像产生变化。 虽然 GAN 是一项突破性技术，能够大规模生成高保真图像，但近年来扩散模型在很大程度上取代了 GAN。[42]有人认为，由于扩散模型能够合成表面上与其训练数据不同的新的高保真图像，而且易于使用，因此是生成大规模图像的事实方法。[43]流行的扩散模型包括 DALL-E 和稳定扩散。

2. 合成数据与国际安全

要点

- 军方和防卫组织可从人工智能和自主系统的不断进步中获益。在国际安全领域，至关重要的是确保人工智能和自主系统在部署和使用前得到适当的培训。
- 合成数据的优势包括：高度多样化的数据集、更短的训练周期、更精细的控制和灵活性、生成假设数据的能力，以及识别和处理偏斜数据集等。
- 合成数据还可以消除收集、存储、传播和处置敏感数据的法律挑战，从而有可能使盟友之间共享更多敏感数据。
- 使用合成数据也有一系列的风险，包括难以完全复制现实世界的复杂物理现象、数据投毒、意外偏差以及与某些合成数据生成技术相关的较低隐私水平。
- 虽然这些风险可能也适用于真实世界的数据集，但合成数据可能会扩大出现其中一些风险的可能性。因此，至关重要的是建立确保合成数据集可靠性和质量的程序。

人工智能和机器学习的不断进步让人们对增强自主系统的功能性和可靠性寄予厚望。事实上，有人认为，通过赋予自主系统更多的自主权，军方和防卫组织都可以获得更强的能力和更高的效率。[44]无论自主程度如何，投入实战的自主系统至少应可靠、可预测、安全，并能按照国际人道主义法运行。

在国际安全领域，至关重要的是确保自主系统在部署和使用前经过适当的培训；自主系统有时会在“行动区内”内做出决策、推断和行动，以减少分析和采取的任何行动之间的延迟。换句话说，

[40] Arash Vahdat 和 Karsten Kreis，“改进扩散模型作为 GAN 的替代方案，第一部分”，英伟达，2022 年 4 月 26 日，https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/。
[41] 所谓去噪，是指去除视听数据中的瑕疵和缺陷，以还原实际特征和特性的过程。见 Linwei Fan 等人，“图像去噪技术简评”，《工医艺的可视计算》，https://doi.org/10.1186/s42492-019-0016-7。
[42] Nicholas Carlini 等人，“从扩散模型中提取训练数据”，2023 年 1 月 30 日，1，https://arxiv.org/abs/2301.13188。
[43] 同上。
[44] Paul Scharre，“战场上的机器人技术第二部分：即将到来的蜂群”，新美国安全中心，2014 年 10 月 15 日，https://www.cnas.org/publications/reports/robotics-on-the-battlefield-part-ii-the-coming-swarm。

收集数据的自主系统与执行分析和提供输出的系统是同一个系统。[45]这一过程被称为“边缘分析”。将自主系统置于边缘的能力已成为各种军事应用技术解决方案中日益重要的组成部分。然而，由于硬件的限制，旨在部署到边缘的自主系统，如用于军事行动的无人驾驶车辆，其开发与用于网络行动等其他情况下的自主系统不同。[46]可以肯定的是，问题不一定是缺乏数据，而是由于缺乏数据收集硬件而缺乏高质量的标记数据。还有一个问题是，自主系统收集的数据缺乏多样性，因为这些系统是为特定的业务功能而投入使用的，并不是为了收集数据。例如，在高空作业的无人驾驶飞行器（UAV）只能从高角度收集图像，因此生成的数据集可能与在低空和低视角作业的另一个 UAV 基本无关。

据推测，自主系统对军事行动具有巨大的价值，包括从在时间紧迫的任务（如防空或防御性网络行动）中比任何人类或人类操作的系统更快地执行任务，到所谓的 3D 任务（即枯燥、肮脏和危险的任务），在此类任务中人类的表现随着时间的推移容易下降。[47]然而，自主系统的数据问题仍然困扰着防卫组织。设计用于机载或机外数据处理的自主系统是一项需要权衡的任务，因为不同的利益攸关方需要满足独特的要求。[48]防卫组织正在努力解决这一权衡问题，并在获取真实世界数据和相关注释方面面临挑战，这些数据和注释可用于训练机载数据处理算法。[49]目前的数据管理架构只允许自主系统在受控环境中以有限的自主程度运行。[50]例如，以色列的 Guardium 无人驾驶地面车辆仅在以色列-加沙边界自主使用，而这一地点的地图绘制良好且相对固定。[51]因此，使用合成数据来训练自主系统可能是缓解与当前数据收集和处理架构相关的数据挑战的一种手段，从而为防卫组织提供了进一步利用自主系统的机会，将其置于高度动态和多变量的环境中，同时降低相关风险。

然而，在网络领域，人工智能的引入可能是大规模开展防御性网络行动和在威胁出现之前识别它们的关键要素。[52]换句话说，人工智能可以提高防御性网络基础设施的可靠性，特别是在应对人工智能支持的进攻性网络行动时。[53]人工智能可能在应对网络领域日益增加的规模和复杂性所带来的挑战方面至关重要。

规模： 随着社会和城市环境在数字互联和异质化方面的不断发展，为防御性网络行动的监管带来了更多的压力点和漏洞。数字系统中的漏洞不仅	**复杂性：** 通过人工智能（如合成图像、对抗性数据操作和其他欺骗性技术）增强和扩大的进攻性网络行动可能会对政府、私营企业或个人的正常运

[45] Martin Hagström，“机器学习和自主系统的军事应用”，斯德哥尔摩国际和平研究所，2019 年 5 月，https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf。
[46] Kannan，“合成时间序列数据生成”，3。
[47] Vincent Boulanin，“人工智能：入门指南”，斯德哥尔摩国际和平研究所，2019 年 5 月，https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf。
[48] 国防科学委员会，“特别工作组报告：自主性在国防部系统中的作用”，国防部，2012 年 7 月，20，https://irp.fas.org/agency/dod/dsb/autonomy.pdf。
[49] Yan 等人，“合成数据集生成与适配”，1。
[50] Jonathan Kwik 和 Tom Van Engers，“战争的算法迷雾：如果缺乏透明度违反武装冲突法”，《未来机器人生活杂志》，2021 年，7，https://doi.org/10.3233/FRL-200019。
[51] Rebecca Crootof，“杀手机器人已来临：法律与政策影响”，《卡多索法律评论》，2015 年，1869，https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2534567。
[52] 与 Tim Watson 教授的采访，2023 年 4 月 11 日。
[53] Wilner，“人工智能与威慑的未来”。

是攻击手段日益复杂的结果，也是因为攻击面扩大而造成的。换句话说，虽然攻击的类型不一定在变化，但风险的规模却在变化。因此，人工智能可以发挥力量倍增器的作用，通过提供足够的“监控视角”来覆盖数字空间中的足够多的细分领域，从而提高效率。[54]	作构成威胁。因此，在防御方面，可能有必要使用由人工智能增强的系统和做法，通过精细控制来检测和应对异常情况。所以，与规模问题不同，利用人工智能对付由人工智能增强的攻击不仅仅是解决社会或组织缺陷的问题，而更是弥补技术短板的问题。

显然，人工智能在国际安全领域有广泛的使用案例和潜在的使用案例。然而，文化和社会制约以及与人工智能技术的部署相关的技术壁垒继续引发人们对人工智能的安全性、可预测性和可靠性的担忧，尤其是在国际安全领域，从而造成了“实验工具和实战系统”之间的差距。[55]合成数据是建议的一种解决方案，通过提高训练数据的质量和可用性，有助于改善与将人工智能技术引入高风险环境相关的信任缺失问题。

2.1 合成数据的附加值

合成数据的附加值取决于其应用的地点、方式和人工智能系统。一般来说，合成数据可以生成高度多样化的数据集，对数据属性进行精细控制，自动标注或标记数据，并具有成本效益。目的是让合成数据的应用对人工智能系统的训练产生有益影响。

合成数据在多大程度上可以作为原始数据的适当替代，是衡量生成合成数据的方法以及使用合成数据的机器学习模型和人工智能系统是否有用的标准。[56]在某些情况下，甚至可以在没有真实世界等效数据的合成数据集上训练机器学习算法，特别是在无法适当收集真实世界数据的情况下，例如将物体放置在不常见或罕见的环境中。在这种情况下，合成数据的使用可能至关重要。这一点在军事领域尤为突出，因为用于复杂的作战环境中的自主系统的设计和建造是为了提高作战效率和效能，而不是为了进行高粒度的数据收集。因此，UAV 等收集的真实世界数据可能无法反映出所有可能的高粒度相关属性组合，例如相关物体在不同环境中的图像，在不同距离、视角和方向以及不同光照下捕捉到的图像。[57]所以，任何自主系统要想安全、可预测和可靠，尤其是在不受控制的环境中，就必须能够合成具有所有可能相关属性组合的多种场景，并能正确识别罕见情况。

使用合成数据训练自主系统还可以缩短训练周期。由于人工智能系统迫切需要数据中蕴含的经验，而不是数据本身，因此使用真实世界数据训练人工智能系统可能是不切实际的。收集足量的真实世界数据并确保数据集具有充分的多样性，是一个耗费大量资源和时间的过程。即便如此，也很难确保训练数据集中穷尽了所有可能的变化和多样性。此外，真实世界数据可能无法提供合成数据所赋予的高度精细的控制和灵活性，以训练人工智能系统满足不同的要求。另一方面，有时会出现[真实世界]数据过多的问题，数据集的特征可能会变得模糊或过于复杂，无法有效使用。在

[54] 与 Tim Watson 教授的采访，2023 年 4 月 11 日。
[55] Hagström，“机器学习的军事应用”，37。
[56] Riemann，“合成数据”。
[57] Yan 等人，“合成数据集生成”，1。

某些情况下，仅仅几秒钟的时间就能获得数十亿字节的数据（如数据包捕获）。因此，在某些情况下，保持基础数据集的特征和统计分布的简单合成数据集就足够了。

此外，在收集经过解析和适当索引的数据可能不成问题的情况下，合成数据可用于生成和学习假设情况，例如防御性网络行动。例如，开发人员可以利用基于代理的建模，这是一种模拟多个变量（如人、物联网系统、时间等）之间相互作用的技术，来创建合成数据集，以反映人们在特定时间内在某些物联网或企业系统上工作的情况。[58]这里的附加值是，即使在物联网系统上工作的组织能够捕获大量完整的数据，但组织可能无法对其收集的数据进行精细控制，以发现或预测所有异常情况，或将异常情况与常规模式区分开来。通过使用基于代理的建模来生成合成现实，组织或许能够训练物联网系统来模拟、识别和分类不同复杂程度的恶意和非恶意活动。

这些技术的应用并不局限于国际安全领域，也不仅仅是理论上的。事实上，基于代理的建模等技术已被应用于其他领域。例如，基于代理的建模已被用于模拟和预测公共政策的影响，如城市规划、未来交通、政策评估或模拟疾病爆发和干预措施。[59]事实上，反映当地人口特征的开源合成人口已经用于联合王国[60]和美国[61]以及更具体的地理区域，如法兰西岛地区[62]（法国）和蒙特利尔岛[63]（加拿大）。

因此，这意味着基于代理的建模可以帮助军方为意外情况做好准备或规划行动。通过基于代理的建模模拟生成合成数据，军方可以为一系列潜在情况做好准备，并制定应对策略。这可能有助于改善军事行动的准备状态和有效性，使其更好地应对意外事件，并为罕见事件或不常见的环境创建数据点。

实例[64]：
基于代理的模型可以模拟人们站在房顶上躲避洪水的场景，但这种情况在现实生活中可能很少发生，因此可以用来训练自主系统识别这种场景的数据极少。不过，这种自主系统将有助于人道主义援助和灾后恢复（HADR）行动。因此，从基于代理的模型中生成的合成数据，通过生成高保真和高度多样化的合成数据，用罕见的数据点来增强训练数据集，可能有助于训练置于此类场景中的自主系统。

[58] 与 Tim Watson 教授的采访，2023 年 4 月 11 日。
[59] Manon Prédhumeau 和 Ed Manley，“用于加拿大基于代理建模的合成人口”，《科学数据》，2023 年 3 月 21 日，https://doi.org/10.1038/s41597-023-02030-4。
[60] Andrew Smith 等人，“预测小地区人口和土地使用变化的开源模型”，《地理分析》，2022 年 2 月 7 日，https://doi.org/10.1111/gean.12320。
[61]William Wheaton 等人，“合成人口数据库：用于基于代理的模型的美国地理空间数据库”，方法报告 RTI 出版社，2009 年 5 月，https://doi.org/10.3768%2Frtipress.2009.mr.0010.0905。
[62] Sebastian Hörl 和 Milos Balać，“基于开放和公开数据的巴黎和法兰西区合成人口与出行需求”，《交通研究 C 部分：新兴技术》，2021 年 12 月，https://doi.org/10.1016/j.trc.2021.103291。
[63] Liliana Perez 等人，“基于地理空间代理的移民人口空间城市动态模型：加拿大蒙特利尔岛研究”，PLOS ONE，2019 年 7 月 24 日，https://doi.org/10.1371/journal.pone.0219188。
[64] Yan 等人，“合成数据集生成”，3。

合成数据集的精细控制使开发人员能够对合成数据集的特性和特征进行微调，并测试机器学习算法的表现和局限性。[65]事实上，可以用相同的基础数据创建多个合成数据集，以用于不同的功能。[66]

合成体系还可以测试由相同基础数据衍生出的合成数据集的变化如何影响人工智能系统最终对环境做出反应的方式。这对于识别和处理偏斜数据集也特别有用，偏斜数据集中某一特性或某类特性的代表性过高（即数据或算法偏差）。合成少数类过采样技术（SMOTE）[67]等技术，可以通过平衡数据集中少数类和多数类的频率，为少数类和多数类获得数量大致相同的样本。[68]条件生成式对抗网络（CGAN）也可以通过对抗训练来减少数据集中的偏斜，从而提高判别器网络的能力，更准确地预测代表性不足的类别，消除整个类别的偏差。[69]

这些“基准”特征意味着在军事领域的适用性。例如，在美国陆军研究实验室进行的一项实验中，研究人员发现计算机视觉系统（如无人驾驶车辆中使用的系统）的性能与训练系统使用的图像的角度之间有相互关系。[70]研究人员注意到，分类器模型对在被摄体（如人、建筑物、坦克等）正上方采集的图像表现出偏差，而且随着摄像头移动的距离越远，视角减小，表现也就越好。研究人员得出结论，一个可能的原因是，由于分类器模型的训练中使用了地面图像，当实验输入看起来更像地面图像时，表现就会更好，因此需要使用更多更高角度的航拍图像重新训练系统。因此，研究人员指出，合成数据可用于比较模型复杂度和架构都不同的不同分类器模型，从而为特定任务选择最佳分类器。

最后，有人认为，创建代表真实世界数据的合成数据还可以消除收集、存储、传播和处置敏感数据的法律挑战。[71]目前，如果组织环境的敏感详细信息（如 IP 地址、网络类型等）被暴露，组织可能不愿意共享与其数字基础设施相关的数据，因为这可能会给其企业数字基础设施的安全带来风险。[72]这一点在国际安全领域可能更为重要，因为即使在盟友之间，敏感的真实世界数据也不容易共享[73]——例如，澳大利亚国防部就指出了与“五眼联盟”伙伴的数据标准不一致所带来的挑战。[74]隐私保护还意味着，合成数据能够防范数据隐私法规的变化，这些变化可能会扰乱组织和组织间共享敏感数据的常规做法，从而增加风险。

2.2 风险

[65] 与 Leslie Sikos 博士的采访，2023 年 4 月 25 日。
[66] Andreas Alfons 等人，“SILC 数据的合成数据生成”，欧盟委员会，2011 年，6，https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP6-D6.2-240611.pdf。
[67] 合成少数类过采样技术（SMOTE）是一种基于规则的合成数据生成方法。
[68] 国际电联，“数据和人工智能评估方法（DAISAM）参考”，ITU-T 人工智能促进健康焦点小组，2020 年 5 月，13。
[69] 同上，12。
[70] Yan 等人，“合成数据集生成”。
[71] Allan Tucker 等人，“生成用于评估机器学习医疗软件的高保真合成患者数据”，《NPJ 数字医学》，2020 年 11 月 9 日，https://www.nature.com/articles/s41746-020-00353-9#citeas。
[72] 与 Leslie Sikos 博士的采访，2023 年 4 月 25 日。
[73] Wilhelm Öhman，“在深度学习物体探测应用中使用军事模拟器进行数据增强”，皇家理工学院，2019 年 9 月 10 日，2，https://www.diva-portal.org/smash/get/diva2:1375838/FULLTEXT01.pdf。
[74] 澳大利亚，“国防数据战略”，20。

虽然合成数据可以帮助减轻防卫组织面临的一些数据挑战，但它并不非灵丹妙药。合成数据也会带来一系列风险和挑战。管理这些风险和挑战的能力尤为重要，能够确保以负责任和安全的方式并按照法律要求和道德价值观使用人工智能系统。

使用合成数据最突出的风险之一就是所谓的“现实差距”。这指的是合成数据与真实世界之间的细微差别。复杂的机器学习模型通常会学习如何利用微小的差异，这使得很难从模拟环境中学习。[75]换句话说，如果合成数据模拟不当，就会遇到无法完全复制真实世界复杂混乱的物理现象的问题，可能无法正确捕捉真实世界数据中出现的变化或一次性案例。

虽然合成数据可用于衡量数据质量、数据偏差和算法偏差，但合成数据本身也会形成（甚至放大）意外偏差。虽然预期偏差在某些应用中可能很有用，例如，过度呈现特定类别的罕见恶意网络流量模式，以便用于监控或事件响应的人工智能系统有更高的几率检测到这些恶意模式，但至关重要的是，这些预期偏差不会表现出意想不到的后果。在几乎所有人工智能系统中，都有一个最佳的合成数据点数量，这取决于训练人工智能系统使用的合成数据和真实世界数据的组成情况。过多的合成数据可能会“过度拟合”人工智能系统，从而降低系统的性能。[76]因此，确保正确的指定范围对于避免意外伤害或其他意外后果至关重要。范围界定不当不仅可能导致自主系统投入使用后产生意外后果，而且还可能导致数据质量偏低、抽样错误、性别或种族偏见、标记或聚合偏差，或生成不完整的合成数据集。[77]

数据偏差和算法偏差的问题除了是一个技术挑战外，也是一个社会和文化挑战。例如，如果一个合成数据集是根据原始的真实世界数据集的特性和特征生成的，而该真实世界数据集包含某些性别或种族规范假设，那么该合成数据集可能会进一步放大这些偏见。即使性别或种族没有“明确地体现在机器学习模型中，从制服或武器证据等中性特征中得出的模式仍可能隐含性别或种族规范”。[78]因此，基于性别和种族的方法凸显了使参与人工智能系统每个步骤（包括数据生成）的人员和专业知识范围多样化的重要性。[79]

此外，合成数据仍然很容易被高水平的恶意行为者进行数据投毒。对手可能在合成数据或数据集中埋下不希望出现的变化，以破坏学习程序，例如向训练数据集中输入一小部分恶意样本，或对合成图像进行微调。[80]不过，值得注意的是，虽然合成数据存在被数据投毒的风险，但与真实世界数据相比，合成数据不易被数据投毒，因为真实世界的数据通常是在遥远和/或不受控的环境中创建的。

最后，虽然某些合成数据生成技术可以保护隐私，但其他技术可能无法提供足够级别的隐私保护。具体来说，与 GAN 等其他技术相比，扩散模型是隐私性最差的图像生成方式。这直接关系到与 GAN 和 VAE 相比，扩散模型在生成更高质量图像方面的效用。[81]换句话说，在某些领域，合成数

[75] Öhman，“数据增强”，6。
[76] 同上。
[77] 与 Leslie Sikos 博士的采访，2023 年 4 月 25 日。
[78] Chandler，“军事人工智能是否有性别之分”，17。
[79] Chandler，“军事人工智能是否有性别之分”，9。
[80] 与 Tim Watson 教授的采访，2023 年 4 月 11 日。
[81] Carlini 等人，“提取训练数据”，1。

据可能会带来隐私与效用之间的权衡，因为越来越强大的生成式模型提出了关于扩散模型如何工作、如何以及在什么情况下应负责任地部署这些模型的问题。[82]

虽然这些风险可能也适用于真实世界数据集，但合成数据可能会扩大大多数风险的潜在风险面。合成数据本身不会带来新的离散风险，但这些风险可能更加普遍。简而言之，风险的类型可能相似，但载体在转变，规模在扩大。然而，有人认为，合成数据的使用可能会比真实世界数据引发更多问题，因为人们对它的信任度普遍较低，这可能会为建立验证合成数据的流程提供更多机会——比验证真实世界数据的机会更多。[83]

[82] 同上。
[83] 与 Tim Watson 教授的采访，2023 年 4 月 11 日。

结论

合成数据已被证明在多个领域都是一种有用的技术，从医疗保健到~~欺诈~~诈骗检测和公共政策规划等。虽然合成数据仍可被视为一种“新兴技术”，[84]但已经足够成熟，在各行各业和公共服务部门，包括与国际安全有关的行业和部门，都已具备了足够的专业知识来进行应用。

事实上，与合成数据相关的附加值和风险与联合国安全进程以及围绕在国际安全领域使用人工智能的其他讨论息息相关。合成数据的潜在优势不容忽视，尤其是精细控制、数据多样性和成本效益。合成数据可能为解决持续困扰防卫组织的一些数据挑战提供解决方案，如数据质量差和数据集多样性低。通过解决其中一些挑战，军方和防卫组织都可以提高作战能力，同时确保遵守国际人道主义法义务，特别是在三维行动中，因为在三维行动中人类表现容易随着时间的推移而下降。

同时，与合成数据相关的风险也不容低估。虽然合成数据不一定会产生有别于真实世界数据相关风险的新风险，但合成数据可能会扩大风险面。换句话说，风险可能相似，但产生相同风险的方式可能更多。例如，真实世界训练数据集缺乏多样性可能会产生意外偏差，就像人工智能系统与一个合成数据点过度拟合也可能会产生意外偏差一样。

虽然合成数据的特征使其成为在国际安全领域开发自主系统的一项大有可为的技术，但不应将其视为解决现有数据挑战的灵丹妙药或万能药。相反，应将其理解为数据管理工具箱中的一个工具。大量研究表明，合成数据和生成模型在过去几年中取得了长足进步。因此，下一步工作应包括但不限于找到具体的使用案例，更有针对性地研究如何在国际安全领域应用合成数据的现有方法和知识，同时考虑数据的性别和种族因素，以及如何将合成数据纳入现有的数据管理战略中。

参考文献

Alfons Andreas, Peter Filzmoser, Beat Hullinger, Jan-Philipp Kolb, Stefan Kraft, Ralf Münnich 和 Matthias Templ. “SILC 数据的合成数据生成”，欧盟委员会，2011 年，6，https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP6-D6.2-240611.pdf。

Alkhzaimi, Hoda. “纽约大学/纽约大学阿布扎比新兴研究与安全中心对第五届实质性会议的供稿”，非政府组织工作文件，2023 年 7 月 28 日，https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/Stakeholder_Recommendation_for_Open-ended_workinggroup_on_security_APR.pdf

Anand, Alisha 和 Harry Deng, “探索防卫工作中负责任的人工智能：各国人工智能原则的梳理与比较分析”，裁研所，2023 年 2 月 13 日，https://unidir.org/publication/towards-responsible-ai-defence-mapping-and-comparative-analysis-ai-principles-adopted

[84] 在本文中，“新兴技术”指的是为应对全球挑战创造了新机遇，同时也带来了新的监管挑战的技术。见经合组织，“经合组织 2012 年科学、技术和工业展望”，2012 年 9 月 13 日，222，https://doi.org/10.1787/sti_outlook-2012-en。

Aryawan Udayana. Putu, Tri Legionosukumo 和 Sri Sundari，“印度尼西亚国家军队综合土地信息系统网络安排战略”，2022 年，https://doi.org/10.33172/jspd.v8i1.1054。

澳大利亚，“2021-2023 年国防数据战略”，澳大利亚国防部，https://www.defence.gov.au/about/strategic-planning/defence-data-strategy-2021-2023#:~:text=The%205%20pillars%20in%20the,capability%20within%20the%20Defence%20workforce。

Bonabeau, Eric.“基于代理的建模：模拟人类系统的方法和技术”，《美国国家科学院院刊》，2002 年 5 月 14 日，https://doi.org/10.1073/pnas.082080899。

Boulanin, Vincent.“人工智能：入门指南”，斯德哥尔摩国际和平研究所，2019 年 5 月，https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf。

加拿大，“国防部和加拿大武装部队数据战略”，国防部，2021 年 5 月 18 日，https://www.canada.ca/en/department-national-defence/corporate/reports-publications/data-strategy/data-strategy.html。

Carlini, Nicholas, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito 和 Eric Wallace.“从扩散模型中提取训练数据”，2023 年 1 月 30 日，1，https://arxiv.org/abs/2301.13188。

Chandler, Katherine.“军事人工智能是否有性别之分？理解偏见并在人工智能的军事应用中推广伦理方法”，裁研所，2021 年 12 月 7 日，https://doi.org/10.37559/GEN/2021/04。

Crootof, Rebecca.“杀手机器人已来临：法律与政策影响”，《卡多索法律评论》，2015 年，1869，https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2534567。

国防创新委员会，“人工智能原则”，国防创新委员会，2019 年，https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF。

国防科学委员会，“特别工作组报告：自主性在国防部系统中的作用”，国防部，2012 年 7 月，20，https://irp.fas.org/agency/dod/dsb/autonomy.pdf。

Fan. Linwei, Fan Zhang, Hui Fan 和 Caiming Zhang.“图像去噪技术简评”，《工医艺的可视计算》，https://doi.org/10.1186/s42492-019-0016-7。

政府商业理事会，“推进边缘的情报、监视和侦察：数字作战空间网络和处理技术调查”，2020 年 7 月，4，http://cdn.govexec.com/media/advancing-isr-at-the-edge-isr.pdf

Hagström, Martin.“机器学习和自主系统的军事应用”，斯德哥尔摩国际和平研究所，2019 年 5 月，https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf。

Hörl, Sebastian 和 Milos Balać.“基于开放和公开数据的巴黎和法兰西区合成人口与出行需求”，《交通研究 C 部分：新兴技术》，2021 年 12 月，https://doi.org/10.1016/j.trc.2021.103291。

Hradec., Jiri, Massimo Craglia, Margherita Di Leo, Sarah De Nigris, Nicole Ostlaender 和 Nicholas Nicholson.“多用途合成人口或政策应用”，欧盟委员会联合研究中心，2022 年 4 月 13 日，14，https://dx.doi.org/10.2760/50072。

Holland, Arthur.“已知的未知：数据问题与军事自主系统”，裁研所，2021 年 5 月 17 日，https://unidir.org/known-unknowns。

国际电信联盟。“数据和人工智能评估方法（DAISAM）参考”，ITU-T 人工智能促进健康焦点小组，2020 年 5 月，13。

Jordan, Jeremy.“变分自动编码器”，2018 年 3 月 19 日，https://www.jeremyjordan.me/variational-autoencoders/。

Kannan, Subarmaniam.“用于边缘分析的合成时间序列数据生成”，F1000 Research，2022 年 1 月 20 日，https://doi.org/10.12688/f1000research.72984.1。

Kingma Diederick P.和 Max Welling.“变分自动编码器简介”，《机器学习的基础和趋势》，2019 年，https://arxiv.org/pdf/1906.02691.pdf。

Kwik, Jonathan 和 Tom Van Engers.“战争的算法迷雾：如果缺乏透明度违反武装冲突法”，《未来机器人生活杂志》，2021 年，7，https://doi.org/10.3233/FRL-200019。

Longford, Frank.“合成数据实验”，《法证建筑》，2018 年 11 月 6 日，https://forensic-architecture.org/investigation/experiments-in-synthetic-data。

Manon Prédhumeau 和 Ed Manley，“用于加拿大基于代理建模的合成人口”，《科学数据》，2023 年 3 月 21 日，https://doi.org/10.1038/s41597-023-02030-4。

Öhman, Wilhelm.“在深度学习物体探测应用中使用军事模拟器进行数据增强”，皇家理工学院，2019 年 9 月 10 日，2，https://www.diva-portal.org/smash/get/diva2:1375838/FULLTEXT01.pdf。

经济合作与发展组织。“经合组织 2012 年科学、技术和工业展望”，2012 年 9 月 13 日，222，https://doi.org/10.1787/sti_outlook-2012-en。

巴基斯坦，“关于致命自主武器系统（LAWS）国际文书的提案”，CCW/GGE.1/2023/WP.2/Rev.1，2023 年 3 月 8 日，https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.3_REv.1_0.pdf

Pasieka, Manuel. “合成数据生成方法和合成数据类型比较”，2022 年 9 月 1 日，https://mostly.ai/blog/comparison-of-synthetic-data-types/。

Perez, Liliana, Suzana Dragicevic 和 Jonathan Gaudreau. “基于地理空间代理的移民人口空间城市动态模型：加拿大蒙特利尔岛研究”，PLOS ONE，2019 年 7 月 24 日，https://doi.org/10.1371/journal.pone.0219188。

Scharre, Paul. “战场上的机器人技术第二部分：即将到来的蜂群”，新美国安全中心，2014 年 10 月 15 日，https://www.cnas.org/publications/reports/robotics-on-the-battlefield-part-ii-the-coming-swarm。

Silva, Thalles. “生成式对抗网络简介”，Thalles 的博客，2017 年 6 月 7 日，https://sthalles.github.io/intro-to-gans/。

Smith, Andrew, Luke Archer, Alistair Ford 和 James Virgo. “预测小地区人口和土地使用变化的开源模型”，《地理分析》，2022 年 2 月 7 日，https://doi.org/10.1111/gean.12320。

巴勒斯坦国，“巴勒斯坦国关于自主武器系统规范和行动框架的建议”，CCW/GGE.1/2023/WP.2/Rev.1，2023 年 3 月 3 日，https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.2_Rev.1.pdf。

Tucker, Allan, Zhenchen Wang, Ylenia Rotalinti, Paja Myles. “生成用于评估机器学习医疗软件的高保真合成患者数据”，《NPJ 数字医学》，2020 年 11 月 9 日，https://www.nature.com/articles/s41746-020-00353-9#citeas。

联合国大会，“裁军与国际安全（第一委员会）”，https://www.un.org/en/ga/first/。

Wheaton, William, James Cajka, Bernadette Chasteen, Diane Wagener, Phillip Cooley, Laxminarayana Ganapathi, Douglas Roberts, and Justine Allpress. “合成人口数据库：用于基于代理的模型的美国地理空间数据库”，方法报告 RTI 出版社，2009 年 5 月，https://doi.org/10.3768%2Frtipress.2009.mr.0010.0905。

Wilner, Alex. “人工智能与威慑的未来：承诺与陷阱”，国际治理创新中心，2022 年 11 月 28 日，https://www.cigionline.org/articles/ai-and-the-future-of-deterrence-promises-and-pitfalls/。

Riemann, Robert. “合成数据”，欧洲数据保护监督机构，https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en

联合王国，“国防数据战略：实施国防数据框架并充分利用数据的力量”，2021 年 9 月 27 日，国防部，https://www.gov.uk/government/publications/data-strategy-for-defence/data-strategy-for-defence。

Vahdat, Arash 和 Karsten Kreis. “改进扩散模型作为 GAN 的替代方案，第一部分”，英伟达，2022 年 4 月 26 日，https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/。

Yan, Jie, Eung Joo Lee, Damon Conover 和 Heesung Kwon. “用于人体检测的合成数据集生成与适配”，DEVCOM 陆军研究实验室，2020 年 11 月，1，https://apps.dtic.mil/sti/pdfs/AD1115446.pdf。