



|  |  |
| --- | --- |
| **课程名称：** | **药学研究与生命健康** |
| **报告题目：** | **疗愈未来的全新可能：人工智能在药物发现领域运用的现状与展望** |
| **专业：** | **应用生物科学** |
| **学号：** | **3230105105** |
| **姓名：** | **张鸣韬** |

**疗愈未来的全新可能:人工智能在药物发现领域运用的现状与展望**

张鸣韬

**摘要：**文章主要探讨了人工智能与制药技术的结合，强调了这一融合对新药研发的重要性。随着新世纪科技的发展，人工智能在药物发现、筛选、设计和优化等方面展现出强大的潜力。文章简要介绍了人工智能在药物领域的应用方面，包括药物-靶标相互作用预测、药物敏感性及应答探测、药物-药物相互作用分析以及药物-药物相似性对比等。通过前述深度学习技术，人工智能能够高效处理大量数据，提高药物研发效率，并为新型药物研发提供有效支持。在技术基础上，文章报道了一些近年来人工智能制药领域取得的重要成果。文章还总结了人工智能在制药技术领域取得的突破与进展，强调了隐私保护、偏见消除、法规合规与监管履行等方面的挑战和未来方向，并且呼吁建立更严格的监管框架，推动人工智能在制药领域持续取得进展。最后，作为报告的总结，文章强调了人工智能技术与制药技术的融合在生命健康领域的潜能，并且乐观展望了发展方向与光明前景。

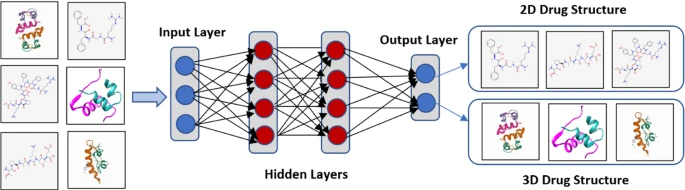
**关键词：**人工智能；深度学习模型；药物发现与制药技术；人工智能制药现实成果；人工智能制药融合发展前景

1. 前言

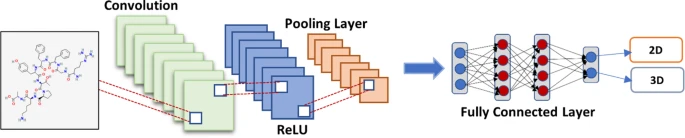
随着新世纪以来现代科学技术的不断发展和生活水平的不断提升，人类的生命健康日益受到重视，减少疾病疼痛、提升预期寿命以确保更加优质的生活水平成为人们所不懈追求的重要目标。而作为疾病防治以及健康保障的核心关键之一，各类药物在其中也起到了不可或缺的重要作用。同时，近年来，计算机科学技术的变革式创新促使了人工智能（Artificial Intelligence）的诞生与发展，为科学研究各个领域提供了强大的算法模型（Algorithm Models）与数据分析（Data Analysis）基础。可以确定的是，作为两项重要的高新科学技术，人工智能与现代制药技术（Modern Pharmaceutic Technology）的强强结合，能够为的药物研发创新带来全新的视角，有效加速药物发现、筛选、设计和优化效率，同时降低研发成本，提高成功概率。总的来说，这两种重要技术的时代性融合能够积极加速新药研发，从而更有效地应对疾病挑战，并推动精准化个性化医疗，为患者提供更安全有效的治疗选择，为未来人类文明发展进步展现更坚韧有力的持续支持。

1. 人工智能制药的技术基础——深度学习模型

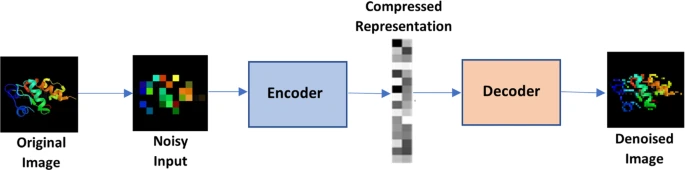
目前而言，主流的大部分人工智能技术采用的工作方式是运用一种被称为深度学习（Deep Learning）的计算模型。而深度学习的基础是神经网络的集成方式性学习，具体学习模式主要包括经典神经网络（Classical Neural Network）（图1）、卷积神经网络（Convolutional Neural Network）（图2）以及自编码器（Autoencoder）（图3）等[1]，人工智能制药技术也不例外。因而在生物化学层面，这些模型能够在大数据的支持下，通过对目标分子的特征提取、深层分析以及降噪、去重等处理，完成对这些生物分子三维结构的预测，并在此基础之上能够为新分子结构、性质的预测以及其生物活性的探查提供有效的数据算法支持。



*图1：经典神经网络的一般工作原理。经典的神经网络是指由算法神经元组成的计算系统，分为输入层（Input Layer）、隐藏层（Hidden Layers）和输出层（Output Layer）。输入的分子结构数据通过连接权重传递到隐藏层，经激活函数处理后传递到输出层，再根据与期望输出的误差调整权重，优化网络。这个过程不断迭代，使网络能够正确建立输入数据与输出的二维或三维结构之间的关系。*



*图2：卷积神经网络的一般工作原理。卷积神经网络通过卷积层（Convolution）、池化层（Pooling Layer）和全连接层（Fully Connected Layer）组成。卷积层通过滤波器检测输入的分子图像特征，池化层降低空间维度——这些层能够帮助网络识别分子图像中的模式和结构。最后，全连接层整合分子特征进行分类或回归，并且通过权重共享减少参数量以提高对分子结构图像等数据的学习效率达到输出标准并输出结果。*

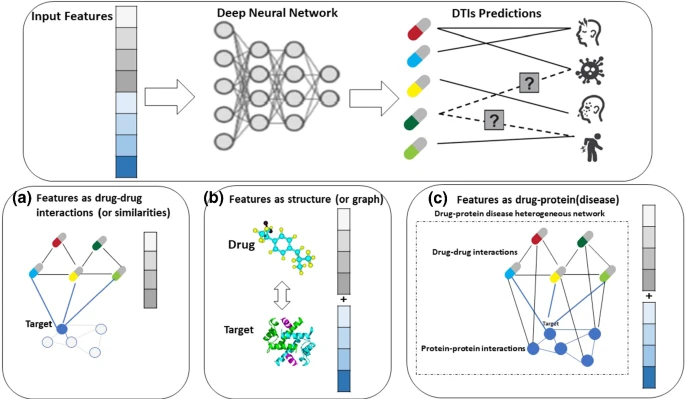


*图3：自编码器的一般工作原理。自编码器通过编码器（Encoder）将输入的蛋白质结构参数数据压缩为复合的潜在再表示，再通过解码器（Decoder）将其还原。其目标是使机器还原的蛋白质结构输出尽可能接近实际输入，强制网络学习数据的有效表示，有助于降维、去噪以生成更符合实际的样本。*

1. 人工智能在药物发现领域的具体应用

（一）药物-靶标相互作用预测（Predictions of Drug-target interactions, DTI）

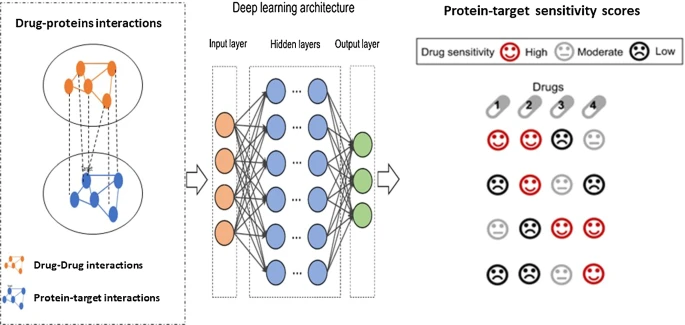
探索药物与靶点的相互作用是新药研发的首要步骤，也是药物筛选和合成制备中至关重要的基础。我们知道，针对这一问题，传统的方法采用了基于对接靶标分子结构和基于药物结构的两种基本策略。但是，分子对接方法在靶蛋白的三维结构不可获得时存在一定的局限性，而基于药物的方法在靶标仅有少数已知结合分子时通常会呈现低于预期的效果。相较于这些传统方法，人工智能技术通过采用非结构化方法，基于深度神经网络对输入的分子结构特征进行学习并建模，克服了药物和靶蛋白高维结构的限制，无需依赖具体的三维结构数据或药物-靶蛋白相互作用的预测。目前，最新的深度学习模型可根据它们的输入特征分为三类：基于药物的模型、基于结构的模型和基于药物-蛋白质（疾病）的模型（图4a, b, c）[2]。在大量数据以及过程的支撑上，通过对靶标分子以及药物分子结构的组合分析，这些模型可以有效比较药物与靶标的相关指数并进行筛选，从而预测药物-靶标的相互作用。



*图4：a)基于药物-药物相互作用（或相似性）（drug-drug interactions or similarities）的模型；b)基于结构（或图像）（structure or graph）的模型；c)基于药物-蛋白质（疾病）（drug-protein or disease）的模型。这些模型通过对不同结构的组合分析进行相互作用性质及程度等的筛选预测。*

（二）药物敏感性及应答探测（Detection of Drug sensitivity and response）

药物的敏感性及其对特定症状的应答程度是药物治疗及其临床效果的重要体现之一。人工智能主导的深度学习方法以药物和蛋白质相互作用的异质网络为输入，通过其预置的阀门层式结构进行多层面的筛选与解析，并在此基础上进行自主评分，并综合各层不同处理模型所得出的结构输出一个比较性的加权评分（图5）。这样，通过直观的分数比较，科学家能够高效地比较并确定各候选药物的敏感特性以及应答程度，极大提升了药物筛选效率和准确精度[3]。



*图5：通过深度学习模型对药物-靶标敏感性的机器评分。深度学习模型通过网络化计算结构对药物-靶标相互作用的相似性进行多轮评分并自动输出标准量化的得分结果以供比较和参考。*

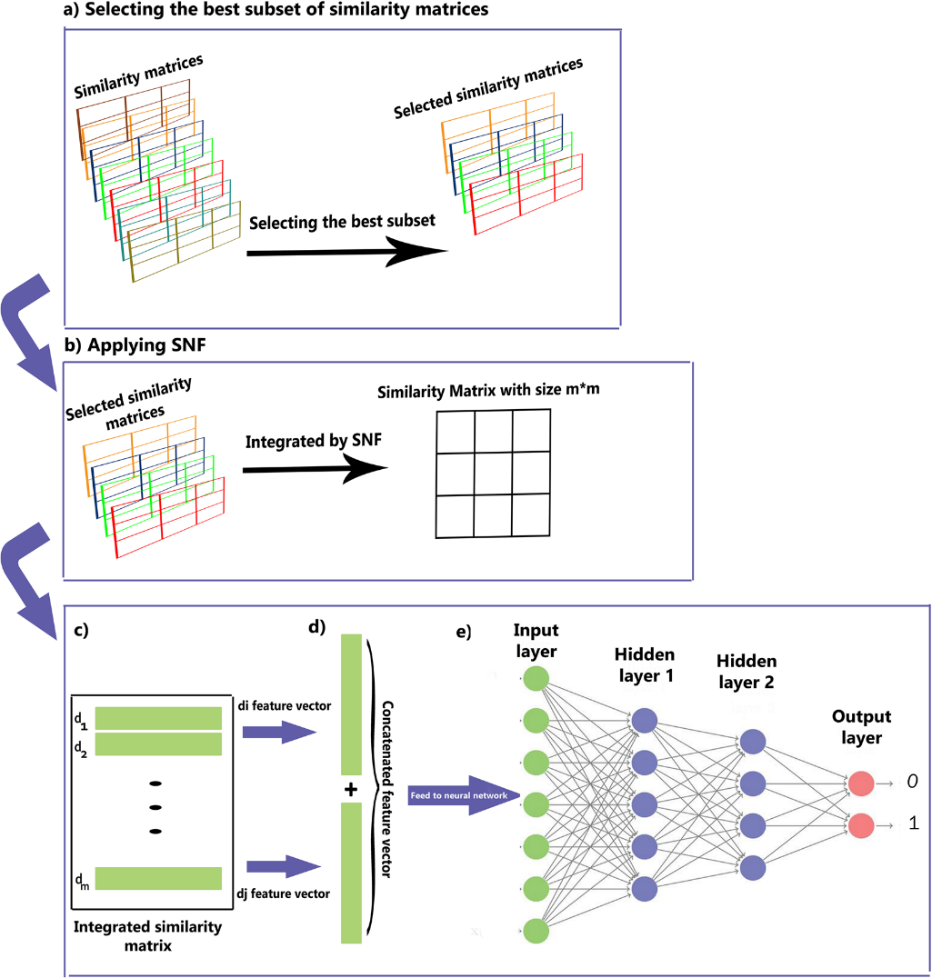
使用深度学习进行药物敏感性和应答预测的一个显著优势是其能够发现复杂的分子模式及其内在关联，这使得深度学习在揭示药物的作用机制、预测药物疗效以及识别潜在副作用方面表现出色。

另一个优势是深度学习模型具有强大的泛化能力[4]。一旦模型完成广泛训练，它可以通过前述预测药物-靶标相互作用的类似方法预测未曾进行临床实验的药物敏感性与应答，这使得深度学习成为预测新药物敏感性和应答的有力工具，有助于积极加速药物研发过程。

总的来说，尽管深度神经网络最近才开始在药物应答预测中发挥作用，但其出色的准确度以及效率使之在当下迅速成为了人工智能制药行业所关注的首要热点之一。

（三）药物-药物相互作用分析（Analyses of Drug-Drug Interaction, DDI）

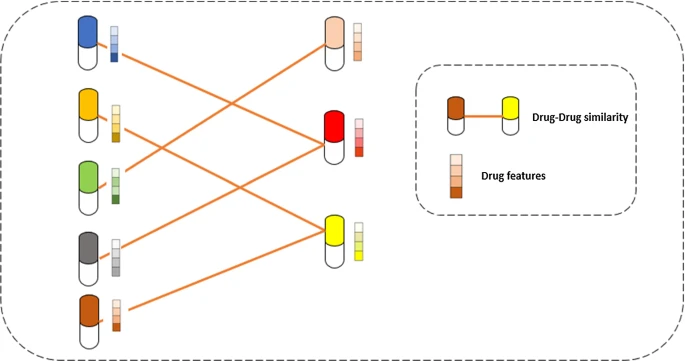
尽管药物与某个靶标的结合可能具有强大的亲和力，但在多种药物联用时，其与多种蛋白结合并具有不同亲和力可能导致毒副作用等不良后果。因此，在药物发现的进程之中，预测药物间相互作用十分重要，有助于减少不良反应的风险，同时，以此为基础来优化药物开发和简化上市后的监测过程，也有助于提高研发效率并节省成本。目前，主要的一种药物相互作用分析方法是基于相似性融合-半监督学习（Integrated Similarity and Semi-supervised Learning, DDI-IS-SL）的计算模型[5]。这个模型主要通过整合药物的化学、生物和表型数据，采用基于数学分析的相似度判断方法计算药物之间的特征相似性，从而提升药物相互作用的预测性能。具体来说，这一计算模型主要使用了正则化的最小二乘分类器进行多重筛选，同时进行二分选择，计算分析药物对之间相互作用可能性的布尔型得分。在5次折叠交叉验证、10次折叠交叉验证等方面，这一模型相较于其他方法表现出更好的DDI预测性能（图6）。此外，相较其他传统方法，这一模型的平均计算时间更为短暂，表现出更高的筛选效率。



*图6：通过相似性融合-半监督学习的计算模型来分析药物-药物相互作用的原理。通过多层的相似性度量，该模型可以揭示不同药物之间的相似性。而且，该模型不仅能考虑有标签的相互作用数据，还能整合无标签数据（无有效临床数据）的信息。这允许模型更全面地理解药物特征，从而更准确地预测未知药物相互作用，输出更简明可靠的对比结果。同时，通过优化半监督学习框架，模型能够更充分有效地利用有限的有标签数据，同时学习分析大量未知的（或潜在的）无标签数据，为药物相互作用的深度解析与广度拓展提供更可靠的支持。*

（四）药物-药物相似性对比（Comparison of Drug-Drug Similarities）

药物的药理学相似性在药物研发的各种研究中都具有至关重要的意义，主要涉及对不同药物的靶标、副作用以及药物间相互作用相似性的预测，以及药物定位前景的问题。人工智能模型在这一领域采用多视角的观点建立算法，包括化学结构、蛋白质靶标、副作用特征和基因表达特征等等，以更有效地预测药物的相似性并进行智能筛选（图7）。这种综合性的方法既可以纠正不同数据源中的数据缺口，也能够为药物重新定位和其他用途提供新的洞察性视角。而对这一问题的深入研究运用，一方面能够全位提升药物发现的“深度”以及“广度”，对药物完成临床及市场定位，从而让有限的研发资源得到最优分配；另一方面，这一技术有望使得人工智能模型指导的药物发现真正从“模仿”转向“创造”，以完成从“仿制（MeToo）”到“进步（MeBetter）”的跨越[6]，从而增强制药企业乃至新药研发技术的持续发展能力。



*图7：药物-药物相似性的智能对比原理。通过对各药物的性质特点按照对其成药特性的贡献独特程度进行分级分类，在此基础上，人工智能技术会对这些药物特征点进行加权的对比分析，进而分析出药物间可能的相似性质以及相似程度。*

1. 突破与进展：人工智能制药技术的现实成果

进入21世纪以来，随着计算机科学技术的飞速发展与进步，一系列由人工智能制药技术主导的药物创新重要成果陆续面世。这些成果在更好地造福人类的同时，很大程度上也成为了这一道路上一面面鲜活的旗帜，指引着后来者不断攻坚克难，攀登创造。

在这些重要成果的报道之中，我们观察到其中的一些典型可以作为不同时期的突破式进展，在人工智能制药发展的历史进程之中有着关键意义[7]：

1. 2006年，基于人工智能的药物结构相似性（Drug Structural Similarity Assessment）评估取得突破，这一成就由科创企业ChemAxon等主导推动，在药物设计中得到了广泛应用。通过计算分子结构相似性，能够有效加速新药发现的进程，使药物设计更为精准，同时提高了候选化合物的筛选效率，为制药工业带来了显著的技术进步。
2. 2014年，深度学习技术在药物定量结构-活性关系分析（Quantitative Structure-Activity Relationship Analysis, QSARA）领域迈出关键一步。在一批优势企业中，加拿大企业Atomwise研发的算法成为亮点：该技术能够准确预测药物与靶点的相互作用，为药物设计提供新思路，缩短研发周期，降低成本，为新药开发注入创新动力。
3. 2022年，基于深度学习的确定治疗指数分析（DL-Based Therapeutic Index Determination Analysis, DL-Based TIDA）在医疗领域也取得了重大突破。这一成果由以IBM Watson Health和Google Health为代表的多家企业和科研机构共同完成。这一技术通过深度学习算法，精准分析患者的病情、基因型等信息，制定个体化的医药治疗方案，有望提高治疗效果，降低治疗风险，促使临床医学治疗迈向个性化、精准化。
4. 2023年，Google旗下的DeepMind公司通过先进的深度学习技术实现了较为准确的蛋白质结构预测（Protein Structure Prediction）以及人体基因的致病错义突变和不正常未知致病基因的自主识别（self-identification of pathogenic missense mutations and unknown pathogenic genes）[8]，并成功开发了AlphaFold、Isomorphic以及最新的Alphamissense等数据模型。这些模型以残基-残基相互作用图网络（Residue-Residue Interaction Graph Network）为核心，为药物设计、靶标解析和疾病防治提供了关键信息与过程支撑[9]。这一突破性成果为新药研发、精准用药、靶向药物传递以及药物再利用方面提供了很大支持。
5. 总结与展望：人工智能深度参与制药技术的挑战与趋势

（一）人工智能技术主导的制药技术研究（AI-driven Pharmaceutics Research）所面临的挑战与未来方向

正如马克思在《资本论》中所指出的那样，在科学上没有平坦的大道，只有不畏劳苦沿着陡峭山路攀登的人，才有希望达到光辉的顶点[10]。当下，人工智能制药技术取得了一定突破，也必然存在诸多挑战。首先，因为不同人工智能模型所采用不同的算法，它们之间的互通兼容性本身即可能存在一些问题。此外，各人工智能模型针对药物、分子模型的结构认识与分析也可能存在较大差异，而这会导致基于人工智能的药物识别和分析的结果质量出现不同的结果[11]。最后，基于隐私和伦理道德方面的考虑，广泛的数据参与所带来的潜在泄露和歧视风险也应当得到足够的认识重视以及法律监管的积极跟进。

面对这些挑战的基本要求是在未来针对人工智能模型的迭代能力和自我学习能力的升级与它们之间算法框架的普及性、有效性同步的（至少是带动性的）发展进步。除此之外，为更好地促进人工智能模型在制药乃至多方面融合科技的创新发展，以下几个方面将成为未来社会所需要关注的重点[12]：

1. 隐私和数据保护：由于人工智能模型的基础建立在大量个人数据之上，涉及隐私问题复杂且多样，因此基于人工智能的制药应用对患者数据的收集、存储和分析必须遵守严格的隐私和数据保护法规，切实保障数据贡献者隐私安全。
2. 偏见消除和公平保障：人工智能模型是基于数据训练的，如果训练数据（如某些基因的表达情况或者药物耐受率等）存在主观意识或带有明显地域性、种群性，可能会导致医疗决策中的偏见性结果。因此，消除偏见与人为因素的干扰是确保其健康发展的关键。
3. 法规合规与监管履行：在制药领域使用人工智能必须符合现有的法规框架和指南，并积极履行相关监管手续。这包括遵守药物开发、临床试验和监管批准流程的标准。

总的来说，随着人工智能在制药领域的普及，我们需要制定更加严格完善的监管框架与指南标准，以确保人工智能在医疗保健和生产研究中负责任和道德的使用，并解决生物安全、数据隐私和责任确定等问题。此外，还应当鼓励研究人员与企业加强国际合作与沟通交流，共建共享多样化的数据集获取（Various Data Set Access）以多维验证人工智能模型，并加速将基于人工智能的药物发现技术转化为临床实践，进而持续推动人工智能制药领域的深刻进展。

（二）人工智能制药技术的融合发展前景

结合前述，我们可以做出总结是，人工智能在制药领域的合作融入无疑将产生极为深远的影响。首先，通过充分利用人工智能的能力，我们可以加速药物研发过程，优化配方设计与流程实施，进行更高精高效的临床试验，并为个性化治疗方法的发展铺平道路。其次，它也将重塑制药行业的格局，为药物研发、科学合作、临床试验以及精准医学等领域带来巨大机遇。总的来说，这一融合不仅会在制药领域内掀起新一轮产业革命，更将为全球医疗保健注入强大的内驱动力，为其自身乃至全人类的未来创造出光明的发展前景。

在最后，我想用制药业界的先驱者，也是默沙东公司联合创始人之一的乔治·默克（George W. Merck）先生的一句话作为这篇报告的总结：“仅仅发明一种新药，并非已经大功告成，我们还要继续探索更有效的途径，在我们这些最佳科技成果能够造福于全人类之前，我们永远不能停下。[13]”所以，希望我们都能够主动拥抱这一革命性的新兴技术，推进人工智能制药的跨学科合作、国际间交流，努力打造一个人工智能驱动的创新制药领域，以切实保障和改善世界人民的生活质量和健康水平，积极助力打造更有活力、更富生机的人类命运共同体，疗愈彼此，共赴未来。

**参考文献**

[1] Bhavesh Bhatt. A Tutorial on Deep Learning Part 2: Autoencoders, Convolutional Neural Networks and Recurrent Neural Networks. *Data Driven Investor,* 1-2 (2018).

[2] Xiaomin Luo，Yunlong Liu，Yuan Yao，*et al.* MINN-DTI: A Deep Learning Approach for Predicting Drug-Target Interactions Based on Molecular Graph and Distance Map. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* 2-5 (2020).

[3][4] Askr H., Elgeldawi E., Aboul Ella H., *et al.* Deep learning in drug discovery: an integrative review and future challenges. *Artif Intell Rev* **56**, 5975–6037 (2023).

[5] Cheng Y., Guihua D., *et al*. Predicting Drug-Drug Interactions Based on Integrated Similarity and Semi-Supervised Learning, *Trans Comput Biol Bioinform*, 19 (1): 168-179 (2022).

[6] 国家药品监督管理局药品评审中心.新药获益-风险评估技术指导原则（征求意见稿）(2022).

[7] Pun F.W., Ozerov I.V., Zhavoronkov A., AI-powered therapeutic target discovery. *Trends Pharmacol Sci* (2023).

[8] h*ttps://endpts.com/exclusive-deepmind-ceo-demis-hassabis-talks-secretive-new-ai-biotech-isomorphic-labs/* (2023)*.*

[9] Jun C., Guido N. *et al*. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* (2023).

[10] 马克思（著），中共中央马克思恩格斯列宁斯大林著作编译局（译）.资本论.人民出版社（2018）.

[11] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, *et al*. Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60 (2023).

[12] Göller A. H., Kuhnke L., Montanari F., *et al*. Bayer’s in silico ADMET platform: A journey of machine learning over the past two decades[J]. *Drug Discovery Today*, 25(9): 1702-1709 (2023).

[13] 梁贵柏.新药的故事.译林出版社（2019）.乔治·W.默克这段话的英文原文是“We cannot step aside and say that we have achieved our goal by inventing a new drug. We can never rest until a way has been found to bring our finest achievements to everyone.”