



中国人工智能系列白皮书

——AI+Art

中国人工智能学会

二〇二三年九月

《中国人工智能系列白皮书》编委会

主任：戴琼海

执行主任：王国胤

副主任：陈杰 何友 刘成林 刘宏 孙富春 王恩东
王文博 赵春江 周志华

委员：班晓娟 曹鹏 陈纯 陈松灿 邓伟文 董振江
杜军平 付宜利 古天龙 桂卫华 何清 胡国平
黄河燕 季向阳 贾英民 焦李成 李斌 刘民
刘庆峰 刘增良 鲁华祥 马华东 苗夺谦 潘纲
朴松昊 钱锋 乔俊飞 孙长银 孙茂松 陶建华
王卫宁 王熙照 王轩 王蕴红 吾守尔·斯拉木
吴晓蓓 杨放春 于剑 岳东 张小川 张学工
张毅 章毅 周国栋 周鸿祎 周建设 周杰
祝烈煌 庄越挺

《中国人工智能系列白皮书——AI+Art》编写组

邱志杰 陈抱阳 巴瑞云 张晓琳 宗钰淇 张冰纨 王梦瑶
何晓冬 汪军 王乃岩 黄木琪 杨星煜 陈杨 刘大庆
徐海 张奥宇 张濮 宋洪涛 代岩 陈海云

目 录

前 言	1
第 1 章 AI 与艺术：历史	1
1.1 数字世界中的前 AI 时代的相关艺术创作	1
1.2 AIGC 之前的 AI 艺术创作及技术机理	8
1.2.1 变化来自标准化问题	8
1.2.2 规则、指令、算法	12
1.2.3 信号、序列、分辨率	26
1.3 科技与社会构成的平台	39
1.4 从数据的可视化问题展开	41
第 2 章 AI 与艺术创作（生成艺术）	43
2.1 生成艺术	43
2.2 生成艺术的发展综述	45
2.3 AIGC 定义与内涵	46
2.4 AIGC 发展综述	48
2.4.1 判别式技术发展综述	49
2.4.2 生成式技术发展综述	51
2.4.3 Stable Diffusion	52
2.4.4 LoRA	57
2.4.5 ControlNet	60
2.4.6 NeRF 作为新型三维重构	63
2.4.7 AIGC 技术发展展望	68
2.5 生成式 AI 的工程化	75
2.5.1 生成式 AI 工程化面临的主要问题	75
2.5.2 AI 生图应用平台	76
2.5.3 利用云计算资源应对算力需求挑战	79

2.5.4	扩散模型的控制	82
2.5.5	通过解决方案加速生成式 AI 技术落地	84
2.5.6	展望：从工程化到工业化	90
2.6	人工智能作为创作的课题	93
2.7	人工智能艺术案例研究	94
2.7.1	AI 生成艺术案例研究	94
2.7.2	AI 交互装置案例研究	135
2.7.3	多智能体系统和 AI 控制的机器人艺术	148
2.7.4	总结	169
第 3 章	AI 对艺术生态的影响	171
3.1	AI 艺术品推荐与未来趋势预测	171
3.2	艺术品鉴定工具	174
3.3	艺术治疗	178
3.4	艺术品解读	185
3.5	艺术教育	190
3.5.1	AI 在艺术教育的应用	192
3.6	艺术品保存与修复	194
3.6.1	艺术品修复	195
3.6.2	艺术品保存	202
3.7	数字人在 AI 艺术领域的意义	208
3.7.1	数字人	208
3.7.2	数字人在 AI 艺术中的意义	209
3.7.3	数字人在 AI 艺术中的作用	210
3.8	AI 艺术面临的问题与挑战	227
3.8.1	人工智能创作的发展趋势	227
3.8.2	版权与道德	230
3.8.3	艺术创作模式的转变	237

3.8.4 技术与资源的挑战	239
结语与展望	241
参考文献	246

前言

自 1956 年提出人工智能（AI）以来，AI 产业与技术不断发展，AI 大模型成为一训多能的人工智能算法基础设施，2019 年以来，大模型泛化求解能力大幅提升，成为产业主流技术路线。AI 大模型是“人工智能预训练大模型”的简称，包含了“预训练”和“大模型”两层含义。“大模型+小模型”逐步成为产业主流技术路线，驱动全球 AI 产业的全面加速。

数字艺术是全球公认的具有独立审美价值的艺术形式，近年来迅速发展，AI 模型技术的成熟无疑将为数字艺术带来更广阔的发展空间，其中特别涵盖于国家文化数字化的战略构想，尤其是数字艺术产业化发展理念。党的二十大报告明确部署建设网络强国、数字中国，实施国家文化数字化战略。中共中央办公厅、国务院办公厅印发《关于推进实施国家文化数字化战略的意见》和《数字中国建设整体布局规划》，说明文化数字化已成为建设社会主义文化强国、实现文化高质量发展的战略选择，对我国文化发展、文化产业国际竞争力和文化安全等方面发挥着不可替代的重要作用。

此白皮书梳理研究 AI 与视觉艺术结合的应用案例和艺术作品，以及国家数字艺术相关战略，为全面更新数字艺术发展模式，实现数字艺术产业新格局提供参考。本研究希望助力推动 AI 艺术创新和产业应用的成果转化，以支撑社会美育建设。以 AI 艺术的发展过程向世界讲好中国故事，传播中国文化价值，建构中国美术教育的话语体系与评价体系，提升中国新时代新形象和文化软实力。通过产生一套面向艺术创作者的 AI 大模型的创作方法论，从而完善大模型数字艺术生成的评价标准，为建立数据库、资源交流平台、满足国家战略提供多维度的建议。同时积极探索运用 AI 大模型将数字艺术延伸到屏幕之外的可能性，为数字艺术形态的迭代提供建议。

将 AI 应用于艺术创作，能推动对 AI 算法的改进研究。通过在不

同应用场景下对算法的表现进行研究和分析，可以使人们不断反思 AI 算法的极限、原理以及它的未来发展。AI 大模型的应用将促进艺术家和科技研究人员之间的交流与合作，进一步拓宽数字化创作的可能性。艺术创作不仅能够拓展 AI 应用场景，还能为 AI 算法的改进提供实验数据和实践基础。跨界合作能同时推动技术的发展和文化与艺术的数字化转型，进而推动中国式现代化进程。因此，艺术与 AI 的深度融合将成为文化数字化建设的重要方向，这也是研究 AI 在国家文化数字化建设中的重要意义所在。

此白皮书旨在强调跨学科和文理贯通的精神，将文化发展和 AI 统一在人文精神框架下，反思科技发展对人的精神和社会心理的影响，并持批判性思考。同时强调艺术思维对科技创新的激发，将横向、发散、逆向等创造性思维融入学科建设，以艺术探索的能量激发双向创新。在艺术学科重视 AI 技术发展史，通过了解技术史、科学思维和实验方法，探索未知领域。

AI 使数字艺术创作来到了新的阶段，国内外大模型已经具有极高的工具属性，这项技术变革对文化与艺术的数字化产生方式带来极大的改变。这种改变渗透在创作生产、内容传播和文化消费，国际竞争的新赛道已俨然从技术路线和成果的比拼转向赋能文化数字化的应用场景之争。技术领域通过国家政策调控正逐渐发力，对于如何将技术成果服务于国家文化数字化亟待研究。

AI 是庞大的系统工程，需要技术和文化数字化应用场景的联动。国际现有 AI 艺术应用广度较好，但在垂直方向有欠缺，无法匹配中国文化艺术数字化生产的逻辑和需求。需要加强技术、设备、内容、内容生产方法、管理机构、政策导向、产业的多维度结合，为广大文化和艺术从业者提供多层次支持，打通技术研究者与文化艺术从业者之间的隔阂。AI 如何成为数字文化艺术创作的工具和合作者，AI 艺术创作工具产品化后如何更好与数字艺术产业结合，艺术工作者如何

更高效地利用大模型生成工具，更好地满足人民群众的美学体验，这些问题亟待解决。

AI 技术在全球化的影响下西方文化逐渐进入中国。目前 AI 艺术创作的研究成果较多来自西方，这显示出一定的研究难度，又揭示了该研究的必要性。扎根中国，学习西方的观点强调了对中国传统文化的保护和传承，同时也注重吸收和融合西方文化中有价值的元素，以此来促进中国与世界的文化交流同时提升国际竞争力、文化软实力。

最后，我们衷心地感谢中央美术学院、京东人工智能研究院、亚马逊云科技、以及匿名的多位师友、单位给予的无微不至的支持，感谢人工智能学会各方的信任和耐心。本白皮书最需要感谢的是，那些在 2023 年酷暑之下，犹如精心雕琢每一件艺术品的雕刻师，倾心打磨、细心塑造的编写团队成员，你们在这白皮书中不可或缺的灵魂。正是因为你们无私的奉献，这份白皮书才得以绽放光彩。考虑到这项工作来的突然，毅然担下这份责任的同时，编写组全体成员在没有额外经费资助的背景下，凭着对 AI 与艺术结合的热情，以及为社会与学术界贡献的初衷，仅仅利用自己的业余时间 and 精力，完善了这份白皮书。在编写过程中，我们参考了众多的资料、艺术家网站和以前的研究成果，对于这些资料的原作者和一直在实践的艺术工作者们，表示最深的敬意和感谢。他们的作品和研究成果为我们提供了宝贵的灵感和支持，帮助我们完成了这个项目。本白皮书由中央美术学院副院长邱志杰主持编写，陈抱阳主笔。由于时间、资源等实际因素的限制，文中难免存在疏漏和不足之处。我们深感歉意，并热切期待广大读者的反馈和建议，以帮助我们不断完善和进步。

第 1 章 AI 与艺术：历史

1.1 数字世界中的前 AI 时代的相关艺术创作

二战后的几年是数字媒体发展的重要时期，其特点是重大的理论和技术发展。1945 年 7 月，《大西洋月刊》（The Atlantic）发表了美国工程师和科学家万尼瓦·布什（Vannevar Bush）的文章《诚如所思》（As We May Think），他在战争期间担任联邦科学研究与发展办公室的负责人，负责监督曼哈顿计划。在展望未来中，布什描述了一种他称之为“MEMEX”的设备，其所代表的科学和技术将是服务于人类，而不是毁灭人类。布什设想内容书籍、期刊、图像可以在缩微胶卷上购买，用户也可以直接输入数据。布什的设备虽然从未被制造出来，但他的概念对塑造计算历史产生了深远的影响。Memex 现在被公认为是数据互联的概念先驱，最终作为一个巨大的、链接的、全球可访问的数据库成为如今熟知的互联网。1946 年，宾夕法尼亚大学展示了占据了整个房间的第一台被称为 ENIAC（电子数值积分器和计算机）的数字计算机；五年后，第一台能够处理数字和文本数据的商用数字计算机 UNIVAC 获得了专利。那时，美国数学家诺伯特·维纳（Norbert Wiener）从希腊语 *kybernetes*（“舵手”）中创造了“控制论”（英文 *cybernetics*）一词，用于描述不同通信和控制系统（例如计算机）与人类的大脑进行比较研究的新兴科学领域。艺术家们立即看到了电子控制系统的创造潜力，并开始尝试控制论相关的艺术创作，例如，在响应式雕塑中使用的反馈回路。

1960 年代对数字技术的历史尤为重要，这一时期为当今的大部分技术及其艺术探索奠定了基础。布什的基本思想由西奥多·纳尔逊（Theodor Nelson）进一步发展，他在 1961 年创造了超文本和超媒体这两个词来描述写作和阅读的空间，其中文本、图像和声音可以通过电子方式相互连接，并可以由任何为该网络文档作出贡献的人进行

链接。纳尔逊的超链接环境是具备不同分支和非线性的，允许读者选择自己的路径。他的概念显然预见了互联网上文件和消息的网络传输。苏联于 1957 年发射人造卫星，促使美国在国防部内设立高级研究计划署 (ARPA)，以保持在技术创新方面的领先地位。1964 年，冷战时代最重要的智囊团兰德公司提出了没有中央权威的通信网络。到 1969 年，由当时的四台“超级计算机”组成的婴儿网络——以五角大楼赞助商的名字命名为 ARPANET：他们的节点分别在加州大学洛杉矶分校、加州大学圣巴巴拉分校、斯坦福研究所和犹他大学。或许犹他大学对于中国读者来说没有其他三所大学有知名度，但是这所大学的研究者对数字技术特别是计算机图形学做出不可磨灭的贡献。

1960 年代末，计算机技术和文化中的另一个重要概念诞生了：当今信息空间和图形用户界面的先驱。1968 年底，斯坦福研究所的道格拉斯·恩格尔巴特 (Douglas Engelbart) 提出了位图 (Bitmap)、窗口和通过鼠标直接操作的思想。他的位图概念是开创性的，因为它把流经计算机处理器的电子信号和计算机屏幕上的图像之间建立了联系。计算机处理以“开”或“关”状态显示的电脉冲，通常称为二进制“一”和“零”。在位图中，计算机屏幕的每个像素都被分配了计算机内存的小单元 (位)，它们也可以表现为开或关，并被描述为一或零。因此，计算机屏幕可以被想象成一个像素网格，要么亮要么暗，以此我们第一次有了一个二维空间。通过恩格尔巴特发明的鼠标，用户可以用手直接操纵这个空间。由恩格尔巴特和伊万·萨瑟兰提出概念，艾伦·凯 (Alan Kay) 和加利福尼亚帕洛阿尔托施乐 PARC 的一组研究人员进一步开发的交互式显示图形程序 Sketchpad 对于开创计算机图形的发展至关重要。他们的工作引导了图形用户界面和后来被称为“桌面”隐喻的建立。这种在屏幕上具有分层的“窗口”，最终随着苹果公司推出其 Macintosh 计算机而普及。虽然计算机和数字技术在 1960 年代和 70 年代绝非无处不在，但人们感觉它们将改变社会。

艺术家们总是采用并反思他们那个时代的技术，他们很快就对探索数字计算进步背后的理论和概念产生了兴趣。1950年代和60年代见证大量艺术家所创作的参与性艺术或技术艺术。约翰·凯奇、艾伦·卡普罗和激浪派；爱德华多·保罗齐（Eduardo Paolozzi）、理查德·汉密尔顿（Richard Hamilton）、威廉·特恩鲍尔（William Turnbull）等人所代表的“独立团队”（Independent Group / IG）；奥托·皮恩（Otto Piene）、海因茨·麦克（Heinz Mack）等人为代表的零（ZERO）；弗朗索瓦·莫雷莱（François Morellet）、胡里奥·勒帕克（Julio Le Parc）等人为代表的 GRAV 视觉艺术研究组（Groupe de Recherche d'Art Visuel）；和杰弗里·斯蒂尔（Jeffrey Steele）、彼得·洛（Peter Lowe）所代表的系统组合（Systems Group）。当时，艺术和计算机技术之间的关系通常是概念性的，这在很大程度上是由于技术并不普及。一些艺术家能够使用废弃的军用计算机，而另一些艺术家则通过他们工作的大学获得了使用计算机技术的机会。1966年，电气工程师比利·克鲁弗（Billy Klüver）想要在工程师和艺术家之间建立起他所描述的有效合作关系，他在贝尔实验室创立了艺术与技术实验（EAT）。克鲁弗与安迪·沃霍尔、罗伯特·劳森伯格、让·廷格利、约翰·凯奇和贾斯珀·约翰斯等艺术家共同开发了项目。这些项目大多首次出现在纽约展出，最终出现在日本大阪70年世博会的百事可乐馆。EAT是艺术家、工程师、程序员、研究人员和科学家之间复杂合作的早期案例，这终将成为数字艺术的标志。当时在贝尔实验室的赞助下创作开创性艺术的其他艺术家包括肯尼斯·C·诺尔顿（Kenneth C. Knowlton）、迈克尔·诺尔（Michael Noll）、马克斯·马修斯（Max Mathews）和莉莲·施瓦茨（Lillian Schwartz）。

1960年代还举办了以艺术与新兴技术之间关系为中心的重要展览。1961年至1973年间在为克罗地亚首都的萨格勒布举办的一系列以“Nove tentencije（新趋势）”为名的五场国际展览。展览旨在推

进战后时代的新艺术概念。他们将“计算机与视觉研究”程序作为第四次展览（Tendencije 4 (1968-69)）的一部分，突出了计算机作为艺术创作的媒介。前两次计算机艺术展览于 1965 年举行，2 月在德国斯图加特的 Technische Hochschule（技术学院）举办了乔治·尼斯（Georg Nees）的《Generative Computergrafik（生成计算机图形）》展览，随后于 4 月举办了由贝拉·朱莱斯（Bela Julesz）和迈克尔·诺尔在纽约霍华德·怀斯画廊的展览。尽管他们的作品类似于抽象绘画，并且似乎复制了传统媒体非常熟悉的美学形式，但这些艺术家在勾勒出驱动任何“数字绘画”过程的基本数学函数时，也捕捉到了数字媒体的基本美学。1968 年，伦敦当代艺术学院举办的《Cyber netic Serendipity（控制论的偶然性）》展览展示了从绘图仪图形到声光环境和感知“机器人”的作品，这些作品在今天看来可能显得笨重且过于技术化，但仍然有着当今许多数字媒体的重要特征。一些作品侧重于机器和变形的美学，例如绘画机器和图案或诗歌生成器。其他的则是动态的和面向过程的，探索交互和“开放”系统的可能性。1970 年，美国艺术史学家和评论家杰克·伯纳姆（Jack Burnham）在纽约犹太博物馆（the Jewish Museum）组织了展览《软件——信息技术》（Software - Information Technology）。除了展示艾格尼丝·丹尼斯（Agnes Denes）、约瑟夫·科苏斯（Joseph Kosuth）、白南准（Nam June Paik）、劳伦斯·韦纳（Lawrence Weiner）等艺术家的作品之外，该展览还展示了西奥多·纳尔逊（Theodor Nelson）的超文本系统《世外桃源（Xanadu）》的原型。

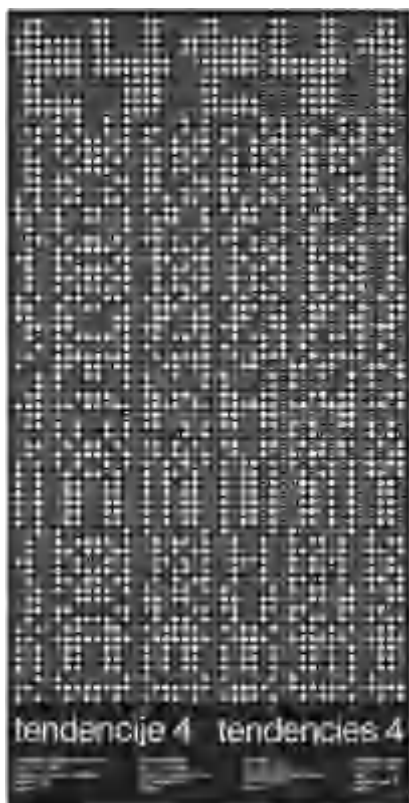


图 1-1 Tendencije 4 海报



图 1-2 Edward Ihnatowicz, Sound Activated Mobile (SAM) , 1968, “控制论的偶然性”现场

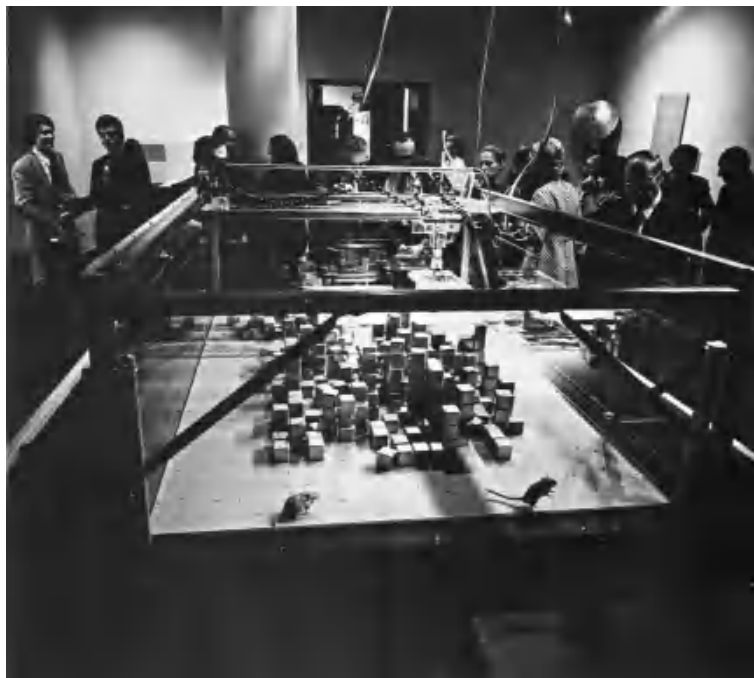


图 1-3 “软件—信息技术”展览现场，1970

1970 年代的艺术家用当时的新技术，例如视频和通讯卫星，也开始尝试现场表演和网络。这些创作有些类似于如今在互联网上以及通过在线视频和音频流媒体进行的互动现象。这些艺术家项目的重点从应用卫星进行电视广播的大规模传播，视频电话会议的审美潜力，到探索打破地理界限的实时虚拟空间。1977 年，在第 6 届卡塞尔文献展上，道格拉斯·戴维斯（Douglas Davis）组织了一次向超过 25 个国家和地区进行的卫星电视转播。其中包括戴维斯本人、白南准、激浪派艺术家和音乐家夏洛特·穆尔曼以及德国艺术家约瑟夫·博伊斯（Joseph Beuys）的表演。同年，纽约和旧金山艺术家合作建立了发送接收卫星网络，并在各城市之间十五小时的双向交互式卫星传输。同样在 1977 年，由基特·加洛韦（Kit Galloway）和雪莉·拉比诺维茨（Sherrie Rabinowitz）联合组织了被称为世界上第一个交互式卫星舞蹈表演——一场由美国东西海岸的表演者参与的三地点现场直播复合表演。与美国国家航空航天局和加利福尼亚州门洛帕克的教育

电视中心合作，该项目建立了一种被艺术家们称为“图像即场所”的观点。这是一种将异地的表演者沉浸在一种“虚拟”空间中的复合现实。



图 1-4 第 6 届卡塞尔文献展卫星电视转播，约瑟夫·博伊斯、道格拉斯·戴维斯和白南准，1977

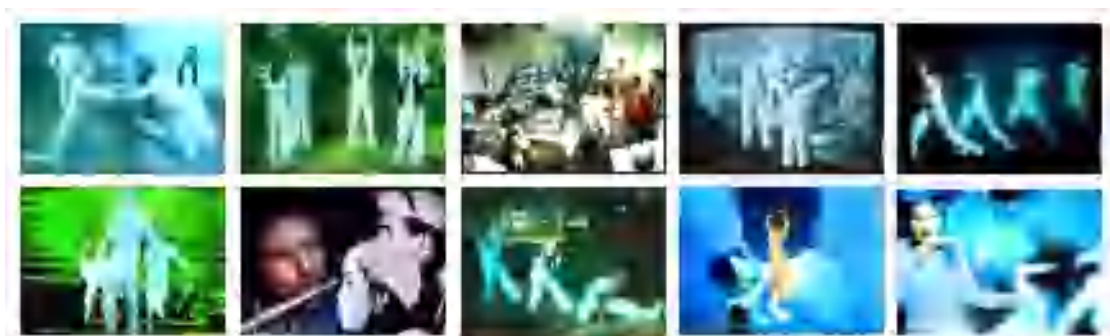


图 1-5 卫星艺术项目，基特·加洛韦、雪莉·拉比诺维茨，1977

在整个 20 世纪 70 年代和 80 年代，画家、雕塑家、建筑师、版画家、摄影师以及视频和表演艺术家越来越多地开始尝试新的计算机成像技术。在此期间，数字艺术演变成多种实践形式，在 1990 年代和 2000 年代将继续多样化，从更面向对象的作品到包含动态和交互方面的作品。随着 1990 年代中期万维网（WWW）的出现，数字艺术在网络艺术中找到了一种新的表现形式，成为互联网艺术探索的多种形式的总称。在 2000 年代初期，网络艺术进入了一个新阶段，艺术家们开始批判性地参与到 Web 2.0 和社交网站上创作作品。随着数

字技术成为我们周遭的一部分（物联网），以及人们对数字语言的熟悉程度不断提高，艺术家开始从事现在被称为后数字的实践，创作广泛依赖于物质形式的数字技术的一系列媒体艺术创作。

1.2 AIGC 之前的 AI 艺术创作及技术机理

1.2.1 变化来自标准化问题

1.2.1.1 算法和数据库

人工智能生长于数字世界，数字世界有两种基本的元素：算法与数据库。我们可以借电子游戏对算法产生形象的解读。当我们玩一款电子游戏时，逐渐熟悉游戏玩法的过程就是了解游戏背后算法的过程。我们虽然并不知道具体代码的样子，但是从玩游戏的过程中我们可以提炼出这款游戏的逻辑——而逻辑就是代码的宏观样貌。当掌握足够多的逻辑关系时，玩家就可以总结出对应的这款游戏讲了一个什么故事。我们可以将数字世界中的算法理解为一种逻辑叙事。

1.2.1.2 变化来自标准化

假设在一张纸上有随意分布的 4 个点，它们可以按照不同的方法连接起来。点与点之间的连线代表着连接的逻辑，也就是算法。同时，我们可以将每一个点看成数据，这 4 个点便是一个简单的数据库。那么连接这 4 个点有多少种不同的方法呢？我们可以利用排列组合方面的经验回答这个寻找规律的问题，但是对于更复杂的情况我们的经验就显得捉襟见肘了。

1.2.1.3 规律和新奇

在人工智能所研究的课题中，大多都离不开算法与数据。比如从一组天气数据中寻找阳光与温度关系，当然在这个例子中的答案似乎从生活经验可得，多阳光照射的地方往往会给当地带来更高的温度，那么南极洲的降雪和北极的温度呢？人工智能则擅长从数据中寻找不易察觉的规律并展现在人们的眼前。这样的规律可以被用在对未来

的预测从而为人们提供辅助和建议。在人工智能领域中，基于现有的数据推算新的结果被称为预测。

当人工智能击败知名棋手时，我们看到了人工智能不论是学习人类棋谱，或还是自我对弈的过程中都产生了人类未曾想到的棋路。棋路顾名思义是下棋的套路，是对下棋规律的总结。当人工智能向我们展示新的规律时，人工智能棋手顺理成章可以成为国家围棋队的训练软件。同理，人工智能也可以为艺术创造提供辅助，我们称之为创造力辅助。

对抗性的体育竞技，在策略层面的取胜之道大体在于出其不意。当对手无法预测我们下一步的行动时，我方便占据了上风。那么怎么才能让对方无法判断呢？如前文所言，便是使用对方未见过或是想到的策略，出奇制胜。这个“奇”可以来自那时那刻的小概率事件，如果是首创的招数自然更能确保新奇。在艺术创作中，也对新奇有着同样的追求，这在二战后的观念艺术中更为显著。彼时艺术创造从技法的较量逐渐过渡到点子的比拼。新奇成为创造力的一种表达。

当然并不是每一个新招都是好招，这就需要审时度势将新招放到对应的情境中检验。新招的产出可以是来自一次随机事件，也可以来自过往的经验。根据我们人类的经验，从历史中学习前车之鉴要比掷骰子来得靠谱，但这是为什么呢？学习下棋的路径包含熟读棋谱，以史为鉴；以及与同僚切磋检验学习成果。对于我们来说棋谱以及同僚对弈皆是过往经验，只不过前者来自前人，后者来自亲身经验。因为人类的一生所能下的棋是有限的，所以我们需要借助书本中前人留下的体悟来更快地提升水平。

像棋谱一样，人类的艺术创作也被著书编册成艺术史，将各家各派都一一地纳入。大体上艺术创作也是由亲身体验和阅读典籍结合而成的。艺术创作同样面临着检验点子的困境——产生点子的速度要远远快于检验点子。这时我们便可借助人工智能寻找规律的能力检验点

子的好坏。借助人工智能背后可扩展的计算能力，可以提高效率，在更短的时间内检验更多的点子。

1.2.1.4 新的图像与方案

人工智能还可以更直接地参与艺术创作。在艺术史中，图像的生产方式几经更迭。在摄影术诞生之前，人们产生图像的方式几乎离不开手工生产。在这个过程中，需要由人将所见所闻的世界转译成二维平面所能承载的信息。这之中包含着创造者对世界不同的理解、多样的技法以及各异的心境。绘画作为产生图像的方式从数量上看是低效的。摄影术的诞生则使得我们具备了利用光学仪器快速产生大量图像的能力。艺术家逐渐开始质疑生产更多图像的意义，并寻找新的生产方法——产生新的图像可以来源于挪用老的图像。但是，这种生产方法却在引入新奇性和多样性上略显不足。借助人工智能，我们可以融合上述两种图像生产的方式，依据给定的条件快速地产生大量全新的图像。

人工智能可以快速地为艺术创作引入多样性，且这种多样性是可以被检验的。在上述图像生产的例子中，为了学习生产人像，那么人工智能会先学习大量的人像照片然后寻找之中的潜在规律，所以其生产的结果是优于随机的，甚至接近前人经验的。在更现代的人工智能系统中，例如对抗性生成网络就同时具备生成器和辨别器。生成器从随机图像开始进行生成，通过辨别器的筛选逐步接近人像的特征。在这个自我进化的过程中，生成器并未学习外来的人像数据，只是根据辨别器所给出的好与坏中寻找规律，最终得出独特的人像生成方案。显而易见的是，这种方案的结果具备很强独特性，同时也更容易被人们所接受。

这些独特的图像既可以被当作创作的最终产物，也可以作为素材库再加以使用。我们在艺术史中所得的经验例如挪用，依然可以应用在人工智能所创造的数据上。曾经艺术家们共用几乎相同的数据库，

借助人工智能艺术家具备了构建独有数据库的能力，这将在根本上提升艺术创作的多样性和独特性。

人工智能还可以产生更复杂的逻辑关系。人工智能有一种研究课题是知识图谱（Knowledge-Graph），用数学可视化的方法表现一组信息，可以是客观事实也可以是一系列点子。人工智能在这之中起的作用是帮助我们发现隐藏的规律，并将这些信息归纳整理并表示出来。

从事音乐创作的艺术家大多有一个属于自己的音频库，其中包含着数以万计的音频素材。这些素材往往按照发出的声音来分类。例如一段鸟叫的音频会被命名为“鸟鸣.mp3”，雷声会被命名为“雷声.mp3”。随着素材量逐渐地增多这种归纳方法会逐渐失效，比方如何区分雷声和重物跌落地面的声音。这两种声音听起来是类似的，但是却有不同。随着这样模棱两可逐渐增多，我们的分类体系逐渐失效。这时候我们可以借用知识图谱系统来对我们收集的声音进行分类，最终的产物不是一个层层叠叠的树状分类，而是一个三维的点云图。每一个点都代表着一种声音，每个点之间的距离则代表声音的相似程度，也就是说人工智能帮助我们完成物以类聚，将类似的声音放在一起，同时我们又具备全局的视野，可以看到对应的声音在我们整个数据库中所处的位置。

现在还可以大胆假设这个知识图谱系统中储存的是从艺术史里摘录的信息，经过合适的人工智能系统的处理可以得到一个完整的关于艺术的点子库。通过不断加入新的其他门类的学科知识来不断完善这个系统。同时，有针对地加入创作者的特异化信息。这种加入可以是主动的，也可以是在潜移默化间的。在艺术创作者的使用中逐渐地适应和调整生成的内容。这种调整可以是纵向的只关于使用者的，也可以是横向的借用其他类似艺术家使用习惯的。这种高效的点子生成器，或许可以将艺术家从观念的比赛中解放出来，回归到艺术中更核

心的共情中来，让艺术创作更关注于情绪的沟通。

1.2.2 规则、指令、算法

当对系统有了相应的了解之后，再回到人类技术、科学和艺术历史发展的背景下，通过“规则、指令、算法”这三个关键词，来串联早期基于规则和指令的艺术形式（如概念艺术）、算法艺术以及尝试建立开放技术的系统（组织）的艺术实践。

1.2.2.1 音乐演奏



图 1-6 音乐演奏

作曲家的乐谱以及指挥家的解读是指导乐手演奏的两个主要外部因素。乐谱所承载的音符构成乐曲的规则，而指挥家下达的则是指令。



图 1-7 《Berlin Mix》, Christian Marclay, 1993

1993 年，艺术家克里斯蒂安·马克莱（Christian Marclay）在柏林莫阿比特区前电车棚的大厅举办了一场由 180 多位音乐家同时举办的音乐会，包含弦乐乐团、放克乐队、管风琴演奏家、合唱团、手风琴管弦乐队、嘻哈 DJ、克莱兹默（Klezmer，一种犹太音乐）二重奏组、歌剧歌手、桑巴乐队、弦乐四重奏、管乐队、摇滚乐队，民族乐团，各种独奏家。马克莱作为指挥在讲台上用纸板发出信号，指示各色音乐家的成员应该何时开始和停止表演。他们都演奏了自己特有的曲目。马克莱将其建立的音乐流派称为“在线混音 Live-Mix”。在作品的演绎中，马克莱作为指挥所下达的指令部分是显而易见的。除此之外，这件作品还引入了对于规则的演绎。它被称为《柏林混音（Berlin Mix）》，也是后现代主义“重新混合（Remix）”的集中表现——把既有的信号重新混合产生新的信号，并广泛存在于各类 DJ 的创作中。重构及再解读也存在于图像一代关于挪用的探索中，但在此我们主要研究指令与规则的关系。

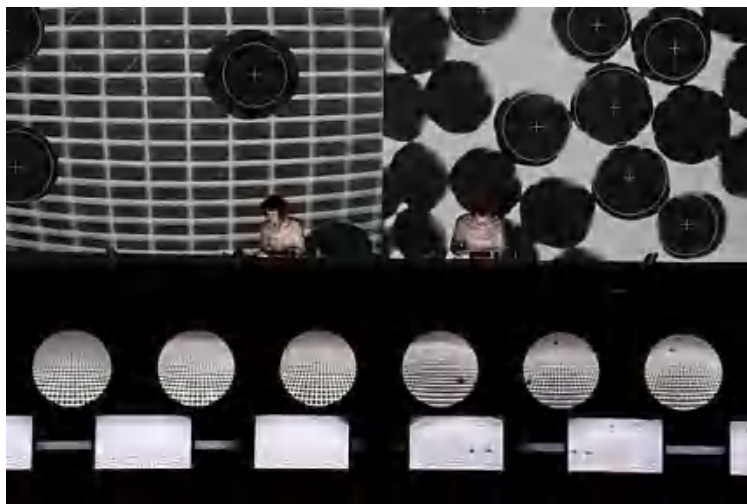


图 1-8 《叠加》，池田亮司

在池田亮司《叠加》的视听表演中，有一幕是台上两位表演者抛掷小球的场景。被抛掷的小球散落在网格图上，被电脑系统读取位置坐标，这些位置坐标则进一步作为输入信号进入池田的生成系统中。在这个环节中，池田作为这件作品的创作者，设计抛掷小球这个规则，并向两位表演者下达执行规定的指令。指令的主要特点是下达指令者与执行者不是同一个人，也就是说他们是两个不同的子系统。如这件作品的名字，叠加这一科学现象本身在作品中发挥着重要作用，最明显的是视觉图像的性质。现场通过闪现数学、方程式和公式、数据、原子结构、数字序列、宇宙的元素为观众构建了量子力学的基本原理的叠加式视觉呈现。现场有一个大型电影院式屏幕，一个较低的长条矩形屏幕和十台电脑显示器排列在舞台前面。池田借用了可视化的系统，但是他并不是为了向观众科普量子力学，而是为了挑战人们的思维。毕竟，现场大多数观众接受过量子理论和高等数学的教育能够完全理解作品的可能性接近于零。池田借用人难以投出相同落点的小球，即对规则重复后得出的多样性和不可能预知性，暗合了量子力学中的基础知识结构。

1.2.2.2 版画

2016年纽约现代艺术博物馆在埃德加·德加的展览中展出了他的

一系列版画。通常，我们看到的版画是由原版复制而来，除去不可避免的误差版画的每一版会被认为是相同的，是同一幅作品的不同版。而德加显然不这么认为，图 1-9 中的 4 幅版画显然来自同一块“板”印制而来，但粗浅浓淡迥然不同——从展厅中的多媒体资料可以得知，德加的这一行为不是“错误”而是有意为之。印制那块“板”的过程可以看作德加这件作品中的规则，而他追求不同效果的方式就是不断重复这个规则。这组版画从传统媒介的角度为我们提供了理解重复与多样性之间关系的例子。



图 1-9 2016 年纽约现代艺术博物馆“埃德加·德加：一种新奇的美”(Edgar Degas: A Strange New Beauty) 展览现场

1.2.2.3 Grasshopper

德里达指出即兴需在一定语境内方能被理解，而这个语境正是由一系列规则所构成的，于是我们得出了变化来自标准化，来自对规则的不断推演。Grasshopper 是一种三维参数化设计软件，这种软件大量地应用在曲面的设计中。借助三维设计软件 Rhino 中的 Grasshopper

插件，设计师可以通过输入实地空间参数，驱动电脑运算产生无穷尽的可能结果，并从中选择最优解。在这个例子中，建筑师具备设计建筑的初衷(为什么要设计)和对最终结果的决定权(具体要设计什么)，但把部分的“如何”也就是过程的一部分交由数字程序负责构建一个庞大的“结果池”供设计师选择。通常我们会因为数据库(结果池)的客观存在，而忽视数据库的建立是主观行为的结果。如果我们把数字程序和电脑想象成小精灵，设计师指挥着一群小精灵画出数以千计的图纸，那么小精灵是图纸的创造者吗？

在此，我们可以将小精灵创造图纸理解为一个图纸生成系统。

1.2.2.4 色彩规律为例

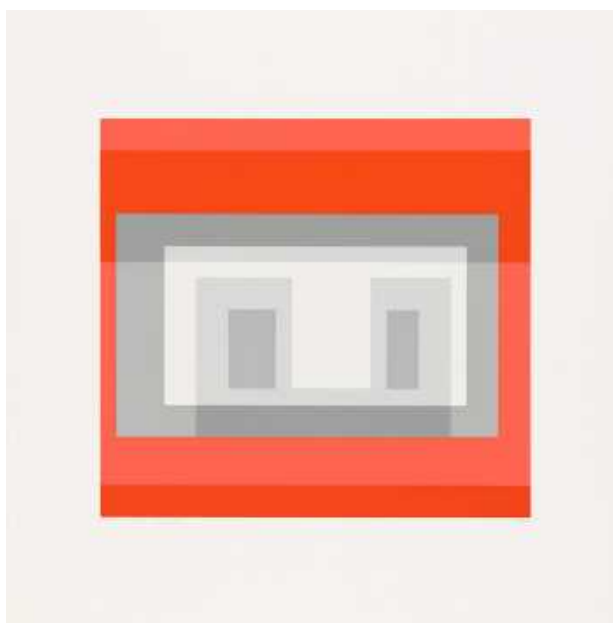


图 1-10 约瑟夫·阿尔伯斯的丝网版画

上图是约瑟夫·阿尔伯斯 (Josef Albers) 的丝网版画，该作品是在这位艺术家从德国移民到美国很久之后制作的，他曾在包豪斯 (Bauhaus) 担任讲师，这所艺术学校于 1933 年被纳粹关闭。从 1990 年代开始，他对于设计的创新激发着艺术和学术界对数字包豪斯的呼声——统一艺术、设计和技术。阿尔伯斯对色彩理论特别感兴趣，并

研究了将不同颜色彼此相邻放置所引起的色调感知变化。像“向正方形致敬”和“变体”系列的作品中，他制定了嵌套彩色正方形和矩形的规则，以强调我们对单一颜色的感知——它的色调、饱和度和透明度——如何根据其接近度而变化以及与相邻颜色的交互。除了简单的色彩混合理论外，人类认知色彩的规则也大量地被认知科学、神经科学界所研究。

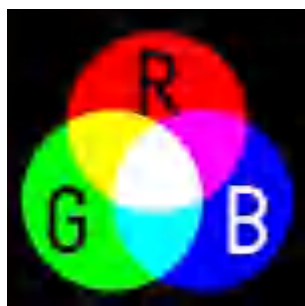


图 1-11 加色混合（例如在计算机中）



图 1-12 减色混合（例如在打印机中）

John F. Simon Jr. 的 *Color Panel v1.0* (1999) 是阿尔伯斯画作的数字变体，是一款基于包豪斯色彩实验的软件艺术作品，显示在由艺术家修改的笔记本电脑上。西蒙编写的软件将屏幕分成五个矩形，对透明度、颜色编码和混合的变化进行编码。其中一个矩形是阿尔伯斯向他的学生提出的“透明度问题”的程序版本，要求他们混合中间颜色，使一个形状看起来覆盖另一个形状。在 *Color Panel v1.0* 中，它是混合颜色以模拟透明度的算法。从西蒙的尝试中，我们可以感到他

将阿尔伯斯画作中色彩规则的探索进一步打包成了一种软件，而软件则是一些算法的集合。结合前面章节对于算法和叙事的比较中对于游戏规则的讨论，可以得出算法也是一组规则的集合，即规则在数学、数字空间中的表达。西蒙的这件作品实际上是艺术实践中较早建立开放性技术系统的尝试。

1.2.2.5 指令成为规则

基于指令和规则的实践，是数字艺术的发展历史的重要脉络之一，在诸如达达（1916 年至 1920 年达到顶峰）、激浪派（1962 年命名并松散组织）和观念艺术（1960-70 年代）等艺术运动中占有重要地位，每一个都关注概念、事件和观众参与，以及正式教学的变化。规则是创造艺术的过程的想法也与构成所有软件和每个计算机操作的基础的算法有明确的联系：一种在有限数量的步骤中完成结果的正式指令过程。就像达达诗歌或激浪派表演的组合和基于规则的过程一样，任何形式的计算机艺术的基础都使用指令作为概念元素。

艺术史中有大量作品构建了编程和概念艺术家之间的联系，这些艺术家将想法的概念视为自己艺术实践背后的驱动力。杰出的概念艺术家索尔·勒维特（Sol LeWitt）在绘画、照片、版画和他称之为“结构”的雕塑等媒介中创作了大量基于想法和指导的作品。他的壁画由以自然语言编写的指令组成，这些指令在特定的展览现场作为图纸执行。

将作品的执行留给艺术家以外的人是勒维特概念艺术概念的核心。在概念艺术中，想法或概念是作品最重要的方面。当一个艺术家使用一种概念性的艺术形式时，这意味着所有的计划和决策都是事先制定的，执行时却是像作品运输那样成为另外的一件事情。这个概念、想法变成了一套制造艺术的系统。这个系统可以向外下达指令，不断地生产出作品。在勒维特过世之后，我们依然可以看到新的壁画出现在各大美术馆。这些新的壁画就像是布展时安装一件作品，布展方按照勒维特留下的说明书（自然语言指令）将壁画重现在墙壁上。

卡西·瑞斯 (Casey Reas) 创造了一款类似于勒维特壁画的软件。瑞斯的软件会执行诸如“创建一个充满一百个中小圆圈的表面”，“每个圆圈都有不同的大小和方向”，“以相同的缓慢速度移动”这样的指令。瑞斯进一步让软件显示圆圈的各种交点。瑞斯这件《{软件} 结构》({Software} Structures) 明确引用勒维特的作品，并探索了观念艺术与软件艺术理念的相关性。艺术家的创作过程从描述视觉元素之间动态关系的文字出发，然后将它们转化为来软件实现。这是一个将指令、规则变成算法的过程。瑞斯还与本·弗莱 (Ben Fry) 一起创建了编程语言《Processing》。这是一个使用 Java 语言的开源开发环境，致力于在可视化环境中向非程序员传授计算机编程的基础知识。

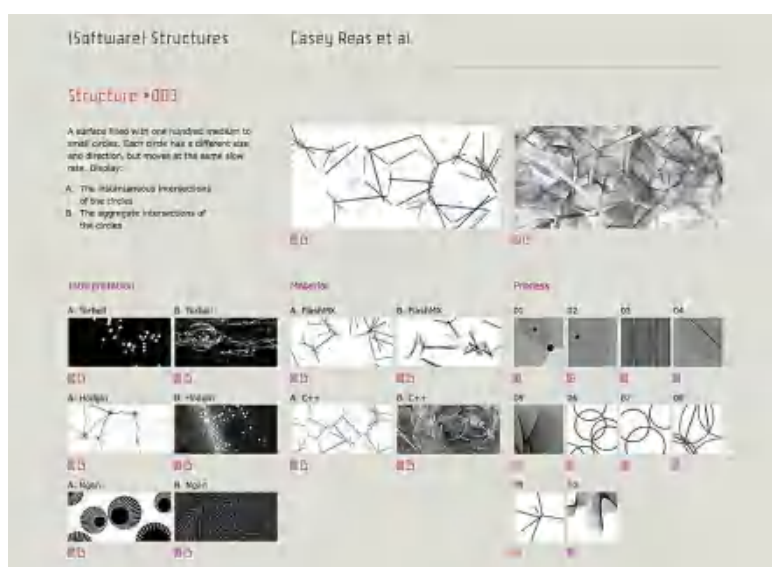


图 1-13 卡西·瑞斯. {软件} 结构.创作于 2004 年, 更新于 2016 年

1960 年代, 像观念艺术那样基于指令的实践几乎是与艺术家开始使用计算机创作早期算法艺术同时发生。早期算法艺术的前驱者编写代码并存储在打孔卡片上, 然后通过计算机运行这些打孔卡片, 以驱动笔式绘图仪, 从而在纸上创建“数字绘图”。

查克·库里 (Chuck Csuri) 在俄亥俄州立大学与程序员詹姆斯·谢

弗（James Shaffer）合作使用 IBM 7094（被认为是 1960 年代早期最强大的计算机之一）创建了《正弦曲线人（Sine Curve Man）》（1967 年）。这是在美国完成的第一幅具象计算机绘图。IBM 7094 被 NASA 用于双子星和阿波罗太空计划，也被用于早期的导弹防御系统。IBM 7094 的由 4×7 英寸的“打孔卡”组驱动。打孔卡其中存储信息被用来驱动鼓式绘图仪，例如绘制一条线需要包含的信息有何时拿起、移动和放下笔，以及何时结束。我们可以发现这种驱动形式是序列式的，同样的驱动方式也出现在 DMX 驱动的机械臂等等。这样驱动方式的特点是非响应式的，也就是说系统其实并不知道自己在画什么，只是忠实地复制和执行指令。机器也不知自己执行的效果，甚至不知道绘图仪上的笔是否还有油墨。如果中途出现了差错，系统会视而不见地继续执行。这些早期的实践者被称为算法学家，其还包括艺术家 Harold Cohen、Herbert Franke、Manfred Mohr、Vera Molnar、Frieder Nake 和 Roman Verostko、Frederick Hammersley 和 Joan Truckenbrod。



图 1-14 查克·库里、詹姆斯·谢弗.正弦曲线人.1967

从上述的例子我们已经可以粗略地理清一组指令可以构成规则，而一组规则可以构成系统，当这些规则被转译成数字世界里的语言时

就成为算法。

接下来我们来研究其他基于指令的实践。比如激浪派艺术家、音乐家和表演者团体所举办的“事件”（happenings），这些“事件”通常也基于精确指令的执行。激浪派将观众的参与和“事件”融合在一起，可以看作是当今许多计算机艺术作品的交互和基于社会事件的先驱。与随机性相关的现成元素（found element）和指令的概念也构成了美国先锋作曲家约翰·凯奇的音乐作品的基础，他在 1950 年代和 60 年代的在互动方面进行大量实验性作品与数字艺术的历史最为相关。凯奇将音乐中的结构描述为可分割成连续的部分，并且经常在他的作品中往预先定义的结构中填充找到的、现成的声音。

1.2.2.6 无为

先锋艺术大量受到了道家无为思想的影响。无为指向的是道家对世界运行规律的理解，强调顺势而为。凯奇于 1952 年首演的作品《433》可说独树一帜、最具代表性。这部作品首演为钢琴演出，它包含三个乐章：钢琴师打开琴盖后，不演奏直到时间到，盖上琴盖。重复三次，过程恰好 4 分 33 秒，演出结束。对于他而言，或是每一位听众，其实没有所谓绝对的无声。又如他以《易经》（Book of changes）为名的作品《Music of Changes》，将声音与乐谱的关系交予随机的卜卦。与其将凯奇的这些作品归于音乐的范畴，还不如归为观念艺术。借无为的思维，这对后结构主义式罗兰·巴特“作者已死”与福柯“人的消逝”宣言的重新思考，重新把作者的重要性通过无为非不为带了回来。



图 1-15 约翰·凯奇.4 分 33 秒原稿.1952

除了前面提到的作品,还有以教学为基础的尝试。《舞蹈(Dance)》(1979) 作品由舞者兼编舞露辛达·柴尔兹 (Lucinda Childs) 和索尔·勒维特 (Sol LeWitt) 合作, 由菲利普·格拉斯 (Philip Glass) 配乐。在图中我们可以看到, 投影在舞蹈下方的画廊地板上的是柴尔兹的表示舞者动作的图表。这些舞者的动作与凯奇的音乐相似, 都使用了重复的结构。这应和了前文提到的对于规则不间断的重复, 将引出变化。但是, 在这里重复有着很强的教学属性。这一点与中国画传统中的临摹非常相近。

1.2.2.7 临摹

临摹是指按照原作仿制书法和绘画作品的过程。临, 是照着原作写或画; 摹, 是用薄纸 (绢) 蒙在原作上面写或画。广义的临摹, 所仿制的不一定是字画, 也可能是碑、帖等。这是学习书画的一个重要过程, 为谢赫六法之一。在临摹过程中所积累的细微变化, 最终会引起质变, 进而表现在创作上。

1.2.2.8 指令与形式

在前文中, 我们梳理了观念艺术中指令的概念, 并与规则、算法

建立了联系。在其他一些作品中，我们会发现指令与形式进一步融为一体。劳伦斯·韦纳（Lawrence Weiner）和约瑟夫·科苏斯（Joseph Kosuth）的文字类作品都揭示了这一点。韦纳的《HERE THERE & EVERYWHERE》（1989）由四个文本片段组成，通过直白的文本让观众展开关于地点的想象。科苏斯的《绿色霓虹灯中的五个字》（1965年）则直接用绿色霓虹灯拼出作品的标题，使用语言作为材料和媒介。在2000年左右流行的数字艺术中，很多艺术家选择直接将构成作品的代码直接展览。



图 1-16 约瑟夫·科苏斯.绿色霓虹灯中的五个字.1965

正如科苏斯的作品一样，观众在看比如布拉德福德·佩利（Bradford W. Paley）的《代码配置文件（Code Profiles）》（2002）直白的展示作品的代码一样，观众直接看到的正是创作作品的语言。佩利还展示了阅读器、编写器和计算机在作品构建中的作用：琥珀色线表示观众如何逐行阅读代码，白线表示佩利是如何编写代码的，绿线表示计算机如何执行代码。借由不同的实体用不同的方式阅读或是理解代码，我们感觉到其讨论的是背后不同的规则。



图 1-17 布拉德福德·佩利.代码配置文件.2022

1.2.2.9 算法

曼诺维奇在 1999 年《作为先锋的软件(Avant-Garde as Software)》中归纳了所有数字媒体的新科技，根据其所支持的功能排成四组：获取（access），生成（generation），操控（manipulation），还有分析（analysis）。随着 1999 年以来软件的进化，出现了各类功能的逐渐一体化，他在 2013 年的《由软件统帅指挥》中进一步指出：“实际上当用户获取进入（access）某软件时，它自动地提供了一些修改功能（manipulation）”——而这种对用户的操控有时候被我们所忽视。一些看似对“传统媒介”的数字化模仿夹带着一些独有的数字逻辑。比如，一个我们每天都在自己的智能手机上使用的技术——触摸屏幕，是对我们在真实空间中“拖拽”“拿放”“开关”等动作的数字化模仿。我们对此习以为常，认为这不过是“媒介的另一种数字化”而已，其实不然，例如，“双指并拢然后分开”这一每天都在使用的“放大”技能，请找一本“古老”的家庭相册试一试，看看能不能如法炮制把 4 寸小照片中的某一个令人怀念的细节放大？我们在使用软件的同时，软件也在改变我们的思维——教会我们一些数字逻辑。

在 Adobe Photoshop 中有一组神奇的功能叫做“Content Aware（图

像识别)”。顾名思义, Photoshop 依据给定的图像, 列入一双鞋子的照片, 通过分析 (Analysis), 推演生成 (Generation) 出不同的结果, 达到对原始照片进行操控 (Manipulation) 的目的。例如, 图像识别填充就是依据周边像素信息, 推算出给定区域的像素值。图像艺术家卢卡斯·布莱洛克 (Lucas Blalock) 有一系列作品利用图像识别功能, 通过图像识别笔刷在画面上涂抹, 产生一组组形似但其实不同的照片。在这些图像产生的过程中, Photoshop 依据自己的内在算法逻辑推演出结果, 最终被平实地展现在艺术家的作品中。

编织与算法之间的关系非常密切。Joseph Marie Jacquard 于 1804 发明了一种提花编织机, 可以将所需编织的方法 (路径) 储存在打孔卡上, 通过打孔卡实现自动化编织。实际上之后的早期计算机采用打孔卡正是受到此的影响, 只不过打孔卡所承载的不仅仅是编织的方法, 可以是更广阔的指令、规则、算法。

更年轻一代的艺术家的作品则直接揭示了算法的潜力。其中一种思路正是前文重点介绍过的生成艺术, 借由一个系统进行艺术创作的过程。这个系统可以是一组人类的语言, 可以有机的生物过程, 也自然可以是数学的、计算机、机械的过程。这之中所强调的是一定程度的自主性, 勒维特的壁画作品显然具有规则和指令的特点, 但是缺乏了一定的自主性。我们无法仅仅依靠勒维特的说明来产生作品, 我们需要有人根据他的说明去执行。同时, 瑞斯的《软件结构》则表现出了极强的自主性, 程序可以自主地决定并且执行生成动作。

算法还能驱动交互类的作品。在这种作品中, 强调的是作品与观众之间的互动。这种作品往往与游戏有一定相似之处, 都涉及观众参与的驱动力, 交互的目的, 交互中的付出及收获的回报——强调的是算法构成的系统的反馈性质。

算法所直接构成的系统还能直接作为对象参与到艺术创作中。早期的例子是以建立开放技术的系统为目标的艺术实践。随着人工智能

的发展，艺术家开始探索算法所代表的规则本身的含义。一些艺术家开始在生成任务和交互任务的基础上，从更宏观的角度研究算法所组成系统。例如，艺术家开始让人工智能（算法）在给定的范围内执行预设任务，例如走迷宫或者捉迷藏，目的是看该算法执行的效果。更进一步的实践是用复杂的算法构建一种生态系统即不同的算法生物之间的互动及自我进化的过程。

1.2.3 信号、序列、分辨率

除了从观念艺术中派生出的指令、规则、算法脉络，还有其他的线索。虽然在整体上这些创作也采用了规则和指令，但它们追溯了与观念艺术所不同的历史源头。它们较多地关注光、光影和动态图像。它们的源头可以上溯至早期动力学的轨迹艺术，并延续到新的数字影像和电视的互动概念。嵌入在这一轨迹中的是不同技术产生“光”的演变。这个脉络同样起源自 1960 年代，那时的艺术家正处于前数字时代，他们所使用的还是电子技术，利用的是电子（模拟电子）信号。电子信号实际上包含两种，模拟与数字。从时间上来看，1960 年代所使用的电子技术都属于模拟电子信号的范畴。数字电路是模拟电路的基础上发展起来的，数字电路是以模拟电路为基础的，虽然它们的基础都离不开电流和电压，但它们有着本质的区别。数字电路是处理逻辑电平信号的电路，它是用数字信号完成对数字量进行算术运算和逻辑运算的电路。模拟电路和数字电路它们同样是信号变化的载体，模拟电路在电路中对信号的放大和削减是通过元器件的放大特性来实现操作的，而数字电路是对信号的传输是通过开关特性来实现操作的。在模拟电路中，电压、电流、频率，周期的变化是互相制约的，而数字电路中电压、电流、频率、周期的变化是离散的。

在模拟电路中，由于信号几乎完全将真实信号按比例表现为电压或电流的形式，造成模拟电路对于噪声的影响比数字电路更加敏感，

信号的微小偏差都会表现为相当显著，造成信息损失。作为对比，数字电路只取决于高低电平，如果要造成信息传递的错误，那么信号的偏差必须至少达到高电平的一半（具体的大小根据不同的电路规格有所不同）。因此，对信息进行量化的数字电路对于噪声的抵御能力比模拟电路更强，只要偏差不大于某一规定值，信息就不会损失。在数字电路中，噪声在各个逻辑门的地方都可以得到消减。

这里我们可以把模拟与数字的区别对应到故障艺术中产生故障的方法。艺术家更容易对模拟信号产生改变，改变与结果的对应更加直接，或者说用数学用语来讲就是更连续。在模拟时代的电视信号、录像带等各式磁带会因为划痕或者干扰产生雪花噪点或者噪声，但是我们总还是可以看到一些原本的画面或者是听到声音。在数字信号时代，比如一条质量欠佳的 **HDMI** 视频线，会让我们完全看不到信号，数字时代“0”或“1”的比喻在此更多代表的是离散的是或非，有或无，并没有中间地带。因为自然界的大多数实际信号是模拟的，所以现代的系统都有将模拟信号与数字信号相互转换的器件。

1.2.3.1 电子信号作为创作媒介

1960 年代的前数字化作品大多将注意力集中在将电子信号作为指令和视觉信息载体的功能上。在这个阶段所提到的电子信号如无特殊注明都是指电子模拟信号，电子数字信号将被简写为数字信号。例如本书多次提到，影像艺术先驱白南准 (**Nam June Paik**) 在他的作品《磁铁电视 (**Magnet TV**)》(1965) 中，将一块工业大小的磁铁放在电视上，这样电视内部的电子信号受到磁干扰就会将画面扭曲成抽象形式。在《推力 (**Thrust**)》(1969) 中，厄尔·雷巴克 (**Earl Reiback**) 同样将注意力放在通过电子信号创建图像上。他将电视的阴极射线管显示器替换成雕塑元素。



图 1-18 《推力 (Thrust)》，厄尔·雷巴克 (Earl Reiback)，1969

在吉姆·希赖特 (Jim Seawright) 的《搜索者 (Searcher)》(1966) 中展现了与光和信号的不同结合。这是一个早期的动力雕塑，其探照灯既能产生光又能对其做出反应。动能是身体通过运动而拥有的能量，而动能艺术在 1960 年代中期至 1970 年代中期达到顶峰。这时期的作品经常受到控制论控制系统思想的启发，并由观众激活机器。



图 1-19 《搜索者 (Searcher)》，吉姆·希赖特 (Jim Seawright)，1966

数字动态图像和数字电影的概念是由多种媒体艺术的历史和实践形成的，包含动画、真人电影到沉浸式环境，以及图像的空间化。

现在主要将电影与真人演绎联系在一起，但这只是动态图像历史的众多轨迹之一。另一个轨迹起源于 19 世纪早期的动态图像，这些图像基于手绘图像并通过前电影设备，如西洋镜和活动电影放映机（Kinetoscope）进行观看。电影将这条线索将发展成动画，它也通过数字媒体的新可能性获得了新的发展动力。1960 年代，老约翰·惠特尼（John Whitney Sr.）和查克·克苏里（Chuck Csuri）等人物继续在计算机图形和电影制作方面进行着开创性工作，贝尔实验室常驻艺术家莉莲·施瓦茨（Lillian Schwartz）则创作了三部开创性的电影。在她的电影《英格玛（Enigma）》（1972）中，施瓦茨使用编程宏语言将屏幕划分为像素网格，并用点图案生成图像。通过矩形之间的快速转换，她创造了频闪色彩的感知。在《牛顿式 I & II（Newtonian I & II）》（均为 1978 年）中，施瓦茨利用数学系统来创造 3D 图像的错觉。

1.2.3.2 图像至空间

白南准的《世纪末 II》（1989 年）首次展出便一石激起千层浪。这件作品由两百多个视频监视器组成，白南准编排音乐和艺术视频的序列，并将影像在网格状的电视机网络多通道的混合在一起。为了创作具有纪念碑意义的《世纪末 II》，白南准从之前播放的电视节目的视频中重新编辑和排序，在看似完全不相关的图像中描绘出一种新的规律。在这个过程中，他将动态图像从其原始背景中解放出来。“重新混合信号”这一概念已在前文通过马克雷的《柏林混音》中谈到过，在这里将重新定义使它从时间维度扩展到空间中。当然，高级的混音系统也能构建出声音的方位，但是在《世纪末 II》中对空间的重构则表现得更为剧烈。白南准使用电视节目作为他的媒介，也对电视节目重新进行了编排（编程），将图像序列重新排列成我们看到的构图。

《世纪末 II》反映了这种编程思维通过媒体内容本身，以及传输

此类内容的底层技术机制来塑造我们的世界。这件作品一共有七个视频通道。其中六个视频频道展示了白南准的各种主题的无声图像处理视频拼贴画。第七个也就是白南准的“主频道”，也是唯一一个有声音的通道，驱动着显示在 64 台电视机，形成了装置的核心。这个主要频道共有三段较长的视频：丽贝卡·艾伦（Rebecca Allen）受发电站乐队（Kraftwerk）委托为其音乐视频《音乐永不停（Musique Non-Stop）》（1986）制作的电脑动画；第二段是大卫·鲍伊（David Bowie）演唱《在愤怒中回看（Look Back In Anger）》，并由露易丝·勒卡瓦利尔（Louise Lecavalier）和她啦！啦！啦！人类步伐舞团（La La La Human Steps）伴舞（录音来自白南准 1988 年的卫星连接项目 Wrap Around The World）；最后一个是保罗·加林（Paul Garrin）的《裸体（Nude）》（1989 年），由菲利普·格拉斯（Philip Glass）创作，其中包含对白南准自己的表演作品《带弦的小提琴》（1961 年）的影射。



图 1-20 白南准和他的《世纪末 II》

白南准的《环游世界》是国际友谊的象征。在首尔奥运会开幕前一周，全球十多个国家参加了这一场估计吸引了 5000 万观众的广播。与 1984 年 1 月 1 日的“早安奥威尔先生”卫星作品不同，传输的材料

主要由参与的广播公司制作，白南准这次仅提供了作品的概念。“早上好，奥威尔先生”则是白南准的第一个国际卫星“装置”——将纽约的电视台与法国巴黎的蓬皮杜美术馆通过卫星连接起来。演出者通过卫星系统从德国和韩国参与到这场向公众直播的事件。与《世纪末 II》相比，这两部作品也是通过新技术打破图像传播的界限，共同着眼于媒介超越一般时空的传播属性而不是图像的空间性。



图 1-21、图 1-22 1989 年展览现场

回到《世纪末 II》这件作品所处的年代，再来从编排中指令、规则的角度审视一番。1989 年，白南准在完成惠特尼首展的布展工作后，给官方留下了作品的图纸——图纸对于作品的维护是很有必要的。对于这样的大型装置而言，图纸还为拆装指明了道路。这件作品在 1989 年的展出后被私人藏家收藏，并转运至夏威夷。在 2018 年，

这件作品重新回到惠特尼展出时，除了必要的硬件更换以外，美术馆发现按照白南准的图纸无法将作品复原——图纸中的连线和 1989 年的照片对不上。

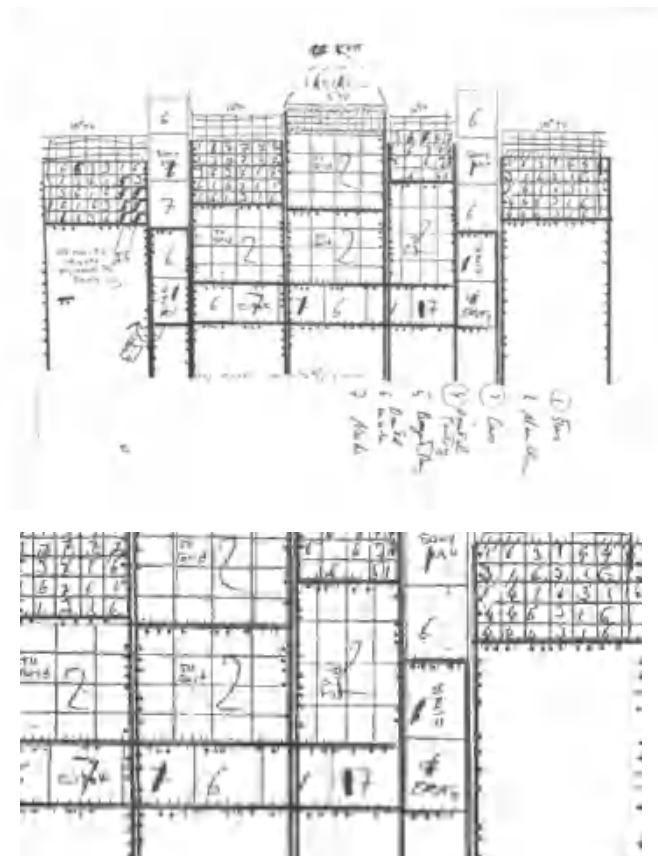


图 1-23、图 1-24 白南准的图纸



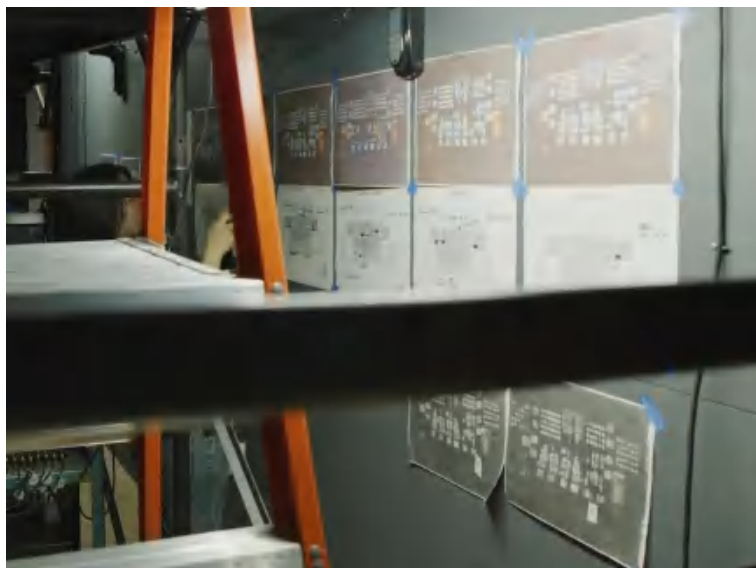


图 1-25、图 1-26 2018 年复原的现场

这样的情景让我们看到了白南准对于规则和指令的思考，特别是与勒维特对待规则与指令的不同。图纸对于白南准也是一种规则，但是在执行规定的时候他还在尝试变化。这是一种骑墙左右摇摆的感觉，既不完全顺应规则也不对规则全盘否定。这也可以用来一窥前面一节从观念艺术出发的脉络和本节所讨论的脉络之间的不同之处。当然，从项目的角度考虑，一份切合现实的图纸将更有益于作品的传承。

回到本节的标题《图像至空间》，从白南准的例子中我们可以看到艺术家对于将图像重新在空间内编排做的尝试。冰岛艺术家斯提娜·瓦苏卡（Steina Vasulka）的 MYND 则通过视频编辑软件对冰岛风景图像进行不同类型的处理，围绕观众创建不同类型的图像空间，一种现在被广泛称为沉浸式空间的形式。斯提娜在作品的自述中表示，创作这件作品时她无法从大量优质的素材中做取舍，于是决定将视频做成 6 条独立的通道。三个水平通道和三个垂直通道，六台投影机将所有图像从地板一直覆盖到天花板。

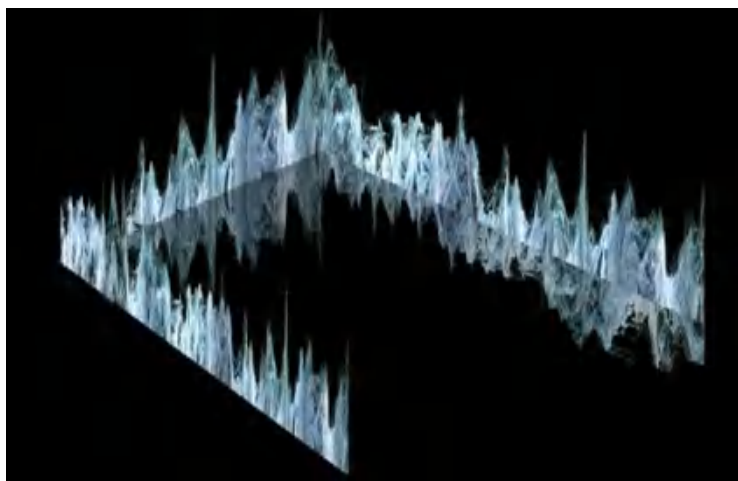


图 1-27、图 1-28 斯提娜 (Steina) (右) 和 伍迪 (Woody) 瓦苏卡在电子设备前，他们创办了纽约的厨房 (The Kitchen) 艺术空间。

吉姆·坎贝尔 (Jim Campbell) 的《倾斜的平面 (Tilted Plane)》(2011 年) 和他早期的其他作品相同是一个“像素”的装置。这件作品足足有一个房间那么大，由数百个悬挂 LED 灯组成，形成一个“像素”网格，就像一个低分辨率视频显示器。当观众进入房间时，在三维倾斜的灯光平面上将显示的鸟类起飞和降落的二维视频抽象成了一种闪烁光效。坎贝尔对分辨率和像素化的探索可以追溯到点画法，这是由乔治·修拉和保罗·西涅克于 1886 年开发的绘画技术，其中不同的微小颜色点 (大部分相同大小) 被应用于画布以形成图像。当然，坎贝

尔的灯泡约等于像素艺术中的一个像素，在前面章节所提及的像素艺术以及数据可视化和坎贝尔的创作有着异曲同工之处。



图 1-29 吉姆·坎贝尔 (Jim Campbell) 《倾斜的平面 (Tilted Plane)》(2011 年)



图 1-30 模棱两可的图标 #1 奔跑 跌倒 (2000)

坎贝尔更早期的创作，如上面的《模棱两可的图标 #1 奔跑 跌倒 (Ambiguous Icon #1 Running Falling)》(2000) 则更像像素艺术，更直接地关乎灯泡所显示的信号在平面上的排布，这种排布如果足够多就是显示器本身。《倾斜的平面 (Tilted Plane)》与《模棱两可的图标 #1 奔跑 跌倒 (Ambiguous Icon #1 Running Falling)》相比，我们可以看到他通过空间的排布，大尺度的在超越平面的第三轴上做出延展，灯泡所蕴含的信号被赋予了更强烈的新含义。

1.2.3.3 从序列出发的交互

在更传统的电影中，数字技术作为一种制作工具发挥着越来越重要的作用。在画廊或博物馆中，常常可以看到利用数字技术制作的作品，例如使用数码相机或后期制作软件拍摄的视频或计算机辅助设计的雕塑。尽管如此，这些作品不被归类为数字艺术，因为数字技术在这里更多地被视为制作工具，而非媒介。真正的数字艺术探索了数字技术作为媒介的关键特征，如实时性、交互性、参与性等，并深入研究其作为媒介的性质和影响。

即使那些不是以特效盛宴为卖点的电影，看起来完全逼真的场景通常也是通过数字处理构建而成。然而，使用数字技术作为制作电影的工具并没有从根本上挑战电影的语言。虽然数字技术可能以多种方式重新定义电影。除了改变电影的视觉外观和扩大创建动态图像的可能性，无论是真人还是动画，数字技术还通过其固有的交互潜力深刻地影响了叙事和非叙事电影。

电影和视频中的互动元素并不是数字时代所特有的。早期的艺术家和表演者通过光的投影，将观众的影子融入艺术作品中。在后来，闭路电视和现场视频捕捉的发展使观众可以成为投影图像的内容。拉杜兹·恰因切拉（Radúz Činčera）的《基诺自动机（Kinoautomat）》，又名《人和他的房子（Človek a jeho dum）》，被认为是世界上第一部互动电影，于1967年蒙特利尔世博会的捷克馆中首次放映。观众投票决定情节将如何展开。但讽刺的是，这个电影只有一个结局。但是，数字媒体无疑激发了互动的潜力。



图 1-31 《基诺自动机（Kinoautomat）》预告片中的截图



图 1-32 林恩·赫什曼·利森（Lynn Hershman Leeson）的《洛娜（Lorna）》

林恩·赫什曼·利森（Lynn Hershman Leeson）的《洛娜（Lorna）》（1979-84）是第一件在光盘上的交互式艺术品。在一个房间（洛娜的客厅）中有一台电视，故事在这台电视里展开，其分支叙事由观众通过遥控器进行导航。在故事中，洛娜是一个与世隔绝的人群恐惧症患者。电视既是作品的互动系统，也是视频主角洛娜唯一与外界联系手段。

我们可以从恰因切拉和利森的作品中看到，当技术允许的情况下，艺术家已不仅仅满足对于序列的编排，他们希望序列的形态是可变的。引入观众的互动，显然是给这个序列系统加上了最直接和天然

的输入。随着进步的发展，人们希望让这种变化的形态脱离人们手动的输入。在列夫·马诺维奇（Lev Manovich）的《软（件）电影（Software Cinema）》（2002）中，艺术家尝试将规则的自动化推演作为输入，引入到作品中。《软（件）电影》是一种动态的计算机驱动媒体装置。观众将看到由定制软件动态构建的无限系列叙事电影。使用马诺维奇定义的规则系统，软件决定屏幕上出现的内容、位置和顺序，同时还能自动选择音乐曲目。这些元素都是从媒体数据库中选择，该数据库包含 4 小时的视频和动画、3 小时的旁白和 5 小时的音乐。

艺术家的创作可以是批判技术带来的社会影响。更宽泛的数字技术早已影响了日常生活的方方面面，艺术家们不仅将数字技术应用于创造新的形式和内容空间，而且大量还用于批判性地参与这些技术的社会、文化和政治影响。

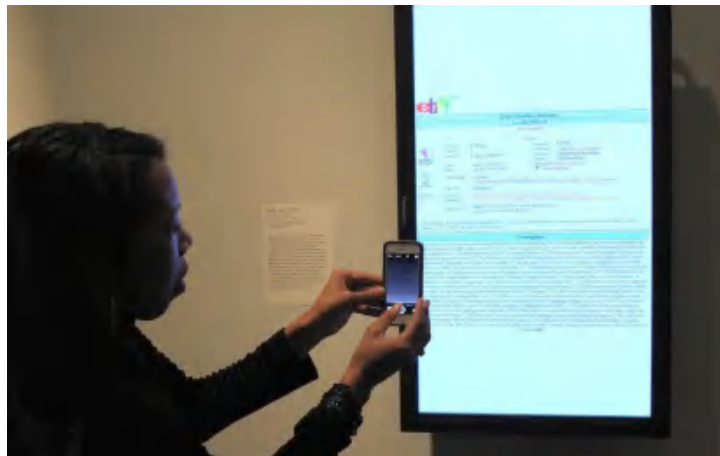


图 1-33 门迪·奥巴迪克（Mendi Obadike）和基思·奥巴迪克（Keith Obadike）的《有色的互动（The Interaction of Coloreds）》

艺术家对于技术偏见的讨论实际上远远早于人工智能时代。门迪·奥巴迪克（Mendi Obadike）和基思·奥巴迪克（Keith Obadike）在 2002 年创作的一件名为《有色的互动（The Interaction of Coloreds）》

中，探索了客户的肤色如何影响在线商务的，质疑了关于互联网被认为是“色盲”（不关乎肤色）的假设。

乔纳·布鲁克-科恩（Jonah Brucker-Cohen）和凯瑟琳·森胁（Katherine Moriwaki）合作的《美国没有天赋（America's Got No Talent）》（2012）创建了一个围绕流行的电视选秀节目的社交媒体数据的交互式可视化系统。这件作品将注意力放在了电视真人秀对社交媒体上人们的互动和表达情绪的影响。上面这些作品指出了技术在收集、处理和分类数据方面给社会带来的深刻变化，不仅改变了我们通信的框架，也改变了整个社会结构。

这些艺术家将人们的交互作为一个整体的系统来看待，我们将在后文继续展开，看艺术家是如何更深刻地探讨这个议题。

1.3 科技与社会构成的平台

技术在艺术创作中的“作者性”不仅体现在上文提到的引入可能性。由算法构成的数字化应用，以及由软件构成的平台，对我们的流行文化和创造思维带来的影响不仅仅是方便快捷。亨利·詹金斯（Henry Jenkins）在《文化融合：新旧媒体的碰撞之所在（Convergence Culture: Where Old and New Media Collide）》一书中，从另一个角度以 Photoshop 为例阐述了技术的可参与性，以及我们在使用过程中产生的社会化影响。数字技术的应用则在多方面产生了变化，获取渠道、参与方式、交互形式，以及以“群体对群体”而不再是“单一对群体”为传播的方式。考虑到上述这些原则，我们应当期待数字分散性将会以去单一节点化的面貌慢慢地浮现出来，事实上一些最新的技术进步已经印证了前述的观点。在 2004 年的美国总统大选中，激进者、粉丝、形形色色的讽刺文章的作者们，都在使用图形处理软件 Photoshop 来挪用和调整图像以制造政治性的宣言。这些图像或许可以被视作是政治漫画的一种草根阶层的等同物——尝试将话题性的关注封装进强

有力的图像当中。**Photoshop** 的出现允许小规模团体以较低的花费拥有专业质量的图像，以及允许公众通过调整和重新流通强有力的图像来表达政治诉求。

这些观点在 2008 年大选中被再一次验证。奥巴马的肖像被艺术家谢泼德·费尔雷 (Shepard Fairey) 通过 **Photoshop** 再编辑而大规模地在互联网上传播。虽然在 2008 年社交网络的用户量只有今日的十分之一都不到，许多时下当红的社交网络还在萌芽中，但是彼时的社交网络已经逐渐形成了各自的平台。当时艺术家与奥巴马竞选团队之间的版权风波，让我们已经可以感受到艺术家和社交网络之间复杂的关系。而由奥巴马海报衍生出来的题为“如何制作你自己的奥巴马海报”的 DIY 网站和教程也如雨后春笋般涌现。这一切可以为接下来的两个例子提供一些基础。

艾玛利亚·乌尔曼 (Amalia Ulman) 的社交网络 (Instagram) 自拍项目展现了她在社交媒体上创建的一个虚拟角色形象 (persona)。这组作品被泰特现代美术馆 (Tate Modern) 纳入题为《为相机进行表演 (Performing for the Camera)》的展览，与辛迪·舍曼 (Cindy Sherman) 著名的自拍共同展出。但是，乌尔曼的作品不单单是一张社交网络照片，更不是单纯地探讨自拍与相机之间的关系。在她的这组作品中，她尝试探索的是社交网络这个图片分享平台的内在逻辑，包括社交网络和人们互动过程中产生的社会性。乌尔曼使用的正是由社交网络而流行的自拍，通过与平台上粉丝的互动、Hashtag 的使用，以及对既有社交网络用户分享照片模式的复制，她成功地塑造出一个虚拟的、在大洛杉矶地区渴望成名的演员的形象。在探讨社交网络的同时，乌尔曼的作品也在社交网络的平台逻辑框架内运行，直接影响着她作品的走向。再如，尼克·布里斯 (Nick Briz) 因为 Facebook 隐私条款的变更，在 Facebook 上表演了一次自己退出 Facebook 的过程。乌尔曼和布里斯的作品体现的是技术以一种平台的形式介入艺术创作的过

程。在这个过程中，引入的是对应平台的逻辑。这种逻辑直接影响和改变着艺术创作的过程和结果，假设 Instagram 没有设计“评论照片”这个功能，恐怕我们就见不到乌尔曼的这组作品，至少不是以这样的形态出现。

1.4 从数据的可视化问题展开

今天走进任何一个画廊或博物馆，我们大概都会遇到在制作过程中使用数字技术作为工具的作品，无论是使用数码相机和后期制作软件拍摄和编辑的视频，还是使用计算机设计的雕塑——辅助设计或照片作为数字印刷品。然而，这些作品本身通常并不被理解为数字艺术，因为它们使用数字技术作为生产工具而不是媒介。相反，数字艺术可以被定义为通过利用其媒体的关键特征或通过反思其性质和影响来探索作为一种媒介的数字技术。其关键特征包括实时性、交互性、参与性、生成性和可变性。

自艺术形式出现以来，数字艺术的定义不断演变。在 1960 年代和 70 年代，数字艺术主要包括算法绘图，其中艺术家编写代码的结果由笔式绘图仪绘制在纸上，以及计算机生成的电影也涉及编程语言的艺术使用。虽然这些作品的形式似乎更接近传统的绘画或电影，但它们的创作者并没有简单地将数字技术用作制作工具，而是深入研究了数字媒体及其潜在代码的潜力。从 1980 年代到 2000 年代初，数字艺术主要被理解为数字技术的艺术。也就是说，通过数字技术创造、存储和分发艺术。当今数字艺术作品的形式，包括实时、互动、面向过程和表演的软件和装置或互联网艺术。虽然并非所有这些特征都是数字艺术独有的，并且也可以在不同类型的表演活动或视频艺术中出现，但它们并不是数字照片或印刷品等物体的内在元素。

随着数字语言变得越来越普遍，我们倾向于将“编程”一词与数字

技术的使用联系起来。然而在整个艺术史上，艺术家们使用程序——规则集和抽象概念——来创作他们的作品，使用数学原理来驱动形式和想法，或者建立规则来探索结构和颜色。前文我们集中追溯过去五十年中的一些实践，探索了基于数学和基于规则的艺术创作。我们看到这些作品涵盖了广泛的门类，包括绘画、编织、素描、舞蹈配乐和软件，以及1960年代的早期灯光和电视雕塑以及大型视频和沉浸式装置。虽然并非所有作品都是技术性的，但它们仍然受到艺术、科学和技术史的影响。在一个越来越被算法编码的世界——从我们与智能设备的对话到金融市场，人类需要审视这些代码的美学和社会影响并思考我们要创建什么样的程序来表达自身或主宰生活？

对于艺术创作而言，我们更需要关注如何利用这些先进的系统来产生我们需要的艺术的结果，也就是将这些由数字、函数、算法及数据库构成的世界转译成为艺术创作。数字艺术中创作的门类大致可以分为三种，数据可视化，数据创世纪，人工智能驱动的数字生态系统。前两种形态无论是从理论框架和实践尝试都较为丰富，第三种则是在2020年之后随着人工智能的发展逐渐得到发展的形式。

第 2 章 AI 与艺术创作（生成艺术）

2.1 生成艺术

生成艺术的定义是不断变化的，定义本身是从数字世界中来，但是它却又与早期的其他艺术门类有着相似性。从规则的角度来看，这种相似性部分来自数据可视化和计算机图形学中的那些模拟世界里的祖先们。前面章节提到的艺术门类中，艺术家着眼于建立规则，或是不满足于现成的规则，而想要另辟蹊径。虽然彼时艺术家与程序员经常涉猎对方的工作领域，但是他们更多的还是在意结果，关注画面是否达到了预期的效果，产生画面的路径及系统的探索则归于计算机图形学的范畴。生成艺术较少关注结果本身，而是将更多的研究放在生成系统本身，换言之生成艺术关注的是规则的建立，这也为数字艺术从数据可视化中衍生出数字生态系统做出了积极的尝试。当然从艺术史上看，现代艺术中的多个艺术思潮都为生成艺术构建了理论框架。

上面对生成艺术所作的简略定义是严谨的吗？生成艺术只能存在于数字系统中吗？本节，我们先尝试对生成艺术做出一定边界的划分。“生成艺术”在数字艺术蓬勃发展的这几十年间，通常是指算法艺术和合成媒介，前者即算法确定的计算机生成的艺术品，后者即任何算法生成媒体的总称，除计算机外艺术家也可以使用化学、生物学、力学和机器人技术、智能材料等进行艺术创作。

我们可以进一步明确：生成艺术是指全部或部分使用相应算法创作的艺术作品，艺术家负责训练算法系统，让系统可以独立生成相应特征的艺术作品。

我们在前文将生成艺术的范围拓展到了数字系统之外，那么生成艺术是指由具有某种程度的自主权的系统创作的任何作品，或者可以

在艺术家很少干预的情况下运作的作品。生成艺术家使用语言规则、机器、算法或基因序列来设计系统，以生成作为艺术品的最终结果。这些系统可以是数字的、化学的或手动的，并且在各种学科中得到实践，包括建筑、诗歌、文学、动画和视觉艺术。那么我们该如何定义使用人工智能系统绘制插画，这是生成艺术还是插画艺术呢？此时我们应该注意到，生成艺术更多的是对系统的探索，而不是内容的生产。所以，如果这位艺术家着眼于这个插画生成系统本身的研究，而不是只局限于对于某一张插画成果的追求，那么我们就可以将其纳入生成艺术的范畴。

我们在此可以回溯像素艺术和故障艺术相关作品，在这些艺术门类中虽然艺术家也在创设规则，或是不满足于现成的规则，而想要另辟蹊径。虽然彼时艺术家与程序员经常涉猎对方的工作领域，但是他们更多的还是在在意结果如画面是否达到了预期的效果等，像是产生画面的路径及系统的探索则归于计算机图形学的范畴。

虽然生成艺术直到 20 世纪后期才变得更普遍，但使用生成思想创作的艺术案例可以追溯到旧石器时代中期。1999 年，考古学家发现了一块石雕，其特点是刻在石头上的网格系统可以追溯到 70000 年前。这一发现是最独特的生成艺术元素的最早例子之一，并表明史前的生成艺术作品仍然依赖于当时最复杂的技术和系统，就像今天的生成艺术依赖于人工智能一样。纵观历史，生成艺术使用最先进的技术来创造以前人类无法实现的东西。在此，我们应该把这种“无法实现”理解的个体化一些，例如一位私营企业主并不具备绘制插图和设计 Logo 的能力，在使用人工智能生成系统之前，这位企业主只能求助于掌握相应技术的专业人士。当今人工智能这样的复杂系统的发展，就如同 2000 年之后 Photoshop 为代表的图像编辑软件的普及对数字生态造成的促进一样——更多的人具备了某种能力，虽然这种能力会变得习以为常，但是在刚发生时都是令人兴奋的，且我们应该意

识到这种技术的变革影响是深远的。如果没有计算机的帮助，就不可能如此便捷地实现以前只存在于艺术家脑海中的图像。这听起来可能有些自相矛盾，但是这台被认为是冷酷无情的机器却可以帮助我们了解人类最主观、最难以企及和最深刻的东西。

2.2 生成艺术的发展综述

随着时间的推移，人们逐渐在艺术讨论中使用“生成”一词。这样的“人工 DNA”的使用定义了一种生成艺术的方法。这个领域的研究专注于构建一个能够产生不可预知事件的系统，这些生成的产物都具有可识别的共同特征。这里提到的“人工 DNA”是借用了生物中基因序列影响蛋白质合成来类比生成艺术中艺术家构建规则与系统，借助人造物来产生艺术作品。我们还应该注意到，在像素艺术中创造规则的人与用该规则进行艺术生产的往往是同一个人，这种一体性在故障艺术中第一次被挑战。这也为我们理解生成艺术中系统的构造者与作品的生产者出现不一致时打下了基础。

从艺术史上看，生成艺术是在立体主义、达达主义和超现实主义等现代艺术流派之后出现的运动，和其前辈们有着一脉相承的混乱和偶然。虽然生成艺术直到 20 世纪末才流行起来，但早期的影响在 19 世纪末就开始出现。像保罗·塞尚这样的艺术家，他们的作品为立体主义原则奠定了基础。之后，未来主义和建构主义都引入了对技术和机器的迷恋。结合起来，这些将成为生成艺术的核心组成部分。

在生成艺术中，艺术家们利用了几乎没有受到自身介入的情况下能够生成艺术作品的系统。在 1960 年代后期，先驱艺术家哈罗德·科恩 (Harold Cohen) 使用被计算机控制的机器人生成绘画，他被认为是最早的生成艺术实践者之一。与摄影的命运类似，生成艺术因其被认为是借助人工系统进行创作的而遭受批评。批评者认为，由于艺

术家通过允许机器完成作品来放弃对作品的物理控制，因此成品与艺术相去甚远。近年来，人们则越来越接受一种生成性的艺术方法。今天，生成性素描、绘画、动画和诗歌在世界各地的博物馆和个人收藏中展出。例如，特纳奖获得者基思·泰森（Keith Tyson）构建了一个《艺术机器》（Art Machine），这是一个复杂的递归系统，可以为泰森制作的艺术品生成详细的提议。

当现代商用计算机在 1960 和 1970 年代出现时，生成艺术家开始在艺术和计算机科学的中间地带尝试使用房间大小的计算机进行艺术创作。在此期间，艺术家们也在生成艺术领域崭露头角。莉莲·施瓦茨（Lillian Schwartz）是第一位作品被现代艺术博物馆收购的生成艺术家，也是第一位将动画数字作品作为美术作品展示的艺术家。

随着计算机变得更加紧凑和易于使用，艺术家们开始使用算法来创作让评论家们兴奋或愤怒的作品。1990 年代后期程序员开发了能够为视觉艺术进行编码的计算机程序，例如 Design by Numbers 和 Processing。这些创新使得任何拥有计算机的人都可以创建生成艺术，从而激发艺术家对生成艺术的兴趣。

2014 年，生成艺术经历了它的下一个伟大里程碑。由于人工智能技术的突破，使得基于人工智能技术的生成艺术得到了进一步发展。生成对抗网络（GAN）被设计成像人脑一样思考，并成功地创造了与模拟艺术相抗衡的艺术品。2021 年扩散模型（Diffusion Model）的诞生则进一步使生成艺术向更加可控，更加聚焦的方向发展。

2.3 AIGC 定义与内涵

AIGC 是一个涵盖人工智能和生成性创造力领域的综合性概念。从字面上来看，AIGC 是继 PGC，UGC 之后的新型内容创作方式，可以在创意、表现力、迭代、传播、个性化等方面，充分发挥技术优

势，打造新的数字内容生成与交互生态¹。目前，AIGC 的概念尚无统一规范的定义。AIGC 从狭义上可被理解为是利用人工智能自动生成内容的生产方式。而从广义上理解，AIGC 可以看作是像人类一样具备生成创造能力的人工智能技术，可以让计算机具备学习、理解和模拟人类创造性思维的能力。在国际上，它与人工智能合成媒体（AI-generated Media 或 Synthetic Media）²术语相对应，被定义为通过人工智能算法对数据或媒体进行生产、操控和修改的技术的统称。综上所述，AIGC 是一种具有创造性的内容生产方式。

AIGC 作为一种新兴的内容生成方式，其概念可以从发展背景、技术能力和应用价值三个方面进行深入剖析。

首先，从发展背景来看，AIGC 的兴起源于深度学习技术的快速突破和数字内容供给需求的增长。随着人工智能技术的不断发展，特别是深度学习算法的应用，AIGC 技术的可用性不断增强。数据的快速积累、算力性能的提升以及算法效力的增强，为 AIGC 的发展提供了有力支撑。此外，数字经济与实体经济的融合以及平台型巨头的数字化场景转型，进一步推动了 AIGC 技术的应用和发展。

其次，AIGC 技术具备丰富的技术能力，包括智能数字内容孪生、智能数字内容编辑和智能数字内容创作等。智能数字内容孪生能力通过对现实世界内容的智能增强和转译，将现实世界的物理属性和社会属性高效地进行数字化，实现现实世界到数字世界的映射。智能数字内容编辑能力具备对现实世界内容进行语义理解和属性操控的能力，从而实现数字世界对现实世界内容的影响和反馈。智能数字内容创作能力则致力于让 AIGC 算法具备内容创作和自我演化的能力，形成具有创作能力的 AIGC 产品。

¹ 亚马逊云科 <https://aws.amazon.com/cn/campaigns/aigc/?refid=45d0b997-4d67-4dfa-96c0-1f1f25f8cd06>

² 维基百科：“人工智能合成媒体（AI-generated Media 或 Synthetic Media）”

https://en.wikipedia.org/wiki/Synthetic_media

最后，AIGC 的应用价值体现在数字内容创新和数字经济发展方面。通过 AIGC 技术，可以以低成本、高效率的方式满足海量个性化需求，推动内容生产向更具创造力的方向发展。AIGC 技术的应用还可以与其他产业进行融合，孕育新业态和新模式，为各行各业的发展提供新动能。例如，在传媒领域，AIGC 技术可以应用于写稿、采访、视频制作等环节，提高内容生产的效率和质量。在影视行业，AIGC 技术可以扩展创作空间，提升作品的质量。在娱乐领域，AIGC 技术可以拓展辐射边界，为娱乐产业注入新的发展动能。

2.4 AIGC 发展综述

在 AIGC 领域中，人工智能技术主要包括机器学习、深度学习和自然语言处理等，这些技术使计算机能够从大量数据中学习和理解规律，从而进行自主决策和创造。在 AIGC 的发展初期，生成性模型（Generative Models）起到了关键作用，如 GAN（Generative Adversarial Networks）和 VAE（Variational Autoencoders）等。这些模型通过学习真实样本数据的分布，能够生成新的、以假乱真的作品。如图 3-1 所示，每个子图的最后一列是 GAN 生成的图像在训练样本中的最近邻，其余列则是在训练期间随机选择的生成器生成的假图像，用来欺骗判别器。后来，大规模的预训练技术兴起，AIGC 作为人工智能技术的一部分，其能力也随预训练的深度学习模型的蓬勃发展而增强³。这些模型是基于大量数据进行预先训练的，通常被称为基础模型（Foundation Models）。机器学习的最新进展（特别是基于 transformer 的神经网络架构的发明）直接带来这一类模型的爆发式增长，这类模型通常包含数十亿个参数或变量。如今的基础模型，例如大型语言模型 GPT4 或 BLOOM，可以执行跨多个领域的多种任务，例如撰写博客文章、解决算术问题、对话聊天、基于文档回答问题等，由

³ <https://aws.amazon.com/cn/campaigns/aigc/?refid=45d0b997-4d67-4dfa-96c0-1f1f25f8cd06>

stability.ai 开发的文生图模型 Stable Diffusion⁴，可以生成创意图片，转换已有图像的风格。

在 AIGC 领域中，判别式技术和生成式技术是两个重要的发展方向，基于基础模型的技术也在最近发展迅速。下面，我们将分别介绍判别式技术和生成式技术以及基于基础模型的 AIGC 技术的发展。

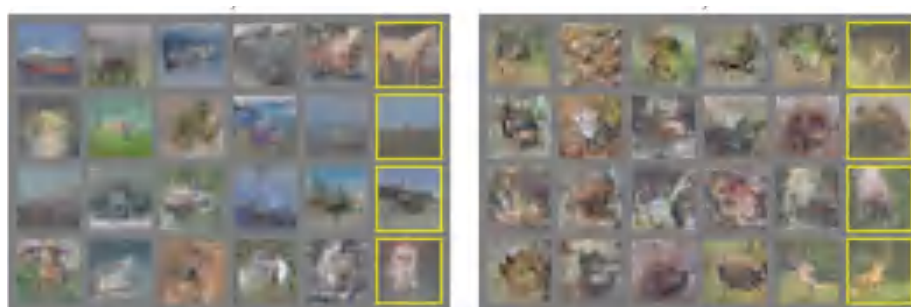


图 2-1 GAN 的生成结果图（来源：GAN⁵）

2.4.1 判别式技术发展综述

判别式技术在 AIGC 领域中具有重要的作用，可以实现分类和预测、特征提取和表示学习、异常检测和异常行为识别，以及决策支持和优化等多种任务。它为我们理解和应用数据中的模式和信息提供了有效的工具和方法。判别式技术可以基于已有数据对样本进行分类或预测。判别式模型关注数据与标签之间的关系，通过学习特征和建立决策边界来进行分类。早期的判别式模型包括支持向量机、逻辑回归和深度神经网络。例如，Alex Krizhevsky 等人在 2012 年提出的

⁴ Stable AI <https://stability.ai/>

⁵ Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.

AlexNet 模型⁶，通过深度卷积神经网络实现了在 ImageNet 图像分类竞赛中的突破。后来，Vaswani 等人在 2017 年提出了 Transformers⁷，并在自然语言处理和计算机视觉领域取得了巨大的成功。Transformers 基于自注意力机制，通过对输入序列中不同位置的关系进行建模，实现了更好的特征表示和序列建模能力。现在，Transformer 模型已成为视觉以及自然语言处理领域的重要基石，如 Google 的 BERT⁸就是基于这一模型。Vision Transformer (ViT)⁹将自注意力机制引入计算机视觉领域。它通过将图像划分为一系列的图像块，并利用 Transformer 模型进行特征提取和分类，取得了与传统卷积神经网络相媲美的性能。

⁶ Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

⁷ Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30

⁸ Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.

⁹ Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. *arXiv preprint arXiv:2010.11929*, 2020.



图 2-2 StyleGAN 生成结果图(来源: StyleGAN¹⁰)

2.4.2 生成式技术发展综述

生成式技术在 AIGC 领域中也扮演着重要角色,它在数据生成和增强、无监督学习、视觉和语言生成、强化学习和策略生成,以及创意和艺术生成等方面为我们开启了创造性和多样性的视角,推动了智能系统在创造、理解和交互方面的进步。生成式技术通过学习数据分布模型来生成新的数据样本。这些模型通常基于概率模型,如生成对抗网络和变分自编码器。其中,生成对抗网络在图像生成领域取得了重大突破,其用于生成逼真的图像样本。后来, Tero Karras 等人提出 StyleGAN¹¹, 用于生成逼真的人脸图像。它通过在生成网络中引入风

¹⁰ Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

¹¹ Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

格向量，并采用渐变叠加的训练方法，生成高分辨率、多样化和具有艺术风格的人脸图像。人脸中的“Style”通常是指头部的姿态、面部的表情、人物的发型等。如图 3-2 所示，StyleGAN 生成的人脸可以捕获到这些细节并生成高质量的图像，而且在不同分辨率的条件下有一致的表现。OpenAI 团队提出的 GPT 基于深度自回归 Transformer 模型。它在自然语言处理任务中取得了突破性的成果，具有强大的语言生成能力和广泛的应用前景。最近，Stable Diffusion¹²作为一种生成式技术，提供了一种有效的方法来生成高质量的图像样本。该方法通过对噪声进行多步扩散来生成图像，每一步都会逐渐减小噪声的规模，使生成的图像逐渐变得清晰。Stable Diffusion 在生成图像的质量和多样性方面取得了显著的进展，并被广泛应用于图像生成任务。

2.4.3 Stable Diffusion

2.4.3.1. Stable Diffusion 的背景介绍

2015 年，一篇研究论文将统计物理学中的扩散模型引入到机器学习领域，为生成高质量图像样本提供了新思路¹³。它通过正向和逆向的扩散过程生成图像，其中正向过程逐步破坏数据分布的结构，逆向过程恢复图像的细节和结构。这也正是最近大火的基于扩散模型的方法（如 Stable Diffusion¹⁴和 ControlNet¹⁵）所使用的基本思想。

初始阶段生成的图像质量较差，而在五年后，另一篇由加州伯克

¹² Rombach R., et al. High-resolution image synthesis with latent diffusion models[C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684-10695, 2022.

¹³ Sohl-Dickstein J., et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]. International conference on machine learning. PMLR, pages 2256-2265, 2015

¹⁴ Ramesh A., et al. Hierarchical text-conditional image generation with clip latents[J]. arXiv preprint arXiv:2204.06125, 2022.

¹⁵Rombach R., et al. High-resolution image synthesis with latent diffusion models[C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684-10695, 2022

利分校发表的开创性的研究论文提出了去噪扩散概率模型 (DDPMs)¹⁶。DDPM 是一种扩散生成模型，由两个参数化的马尔可夫链组成，它利用变分推断¹⁷在有限时间内生成与原始数据分布一致的样本。前向链逐渐将高斯噪声引入数据，使其分布趋近于标准高斯分布，而逆向链则通过参数化的高斯转换核逐步恢复原始数据分布。扩散模型在图像合成、计算机视觉、自然语言处理等领域展现出优异性能。

尽管扩散模型已经在图像数据上取得了最先进的合成结果，扩散模型高昂的计算成本仍让众多研究人员望而却步。为了解决这个问题，潜在扩散模型 (Latent Diffusion Model, LDM)¹⁸ 的开发者 CompVis 和 Runway ML 将扩散模型应用于强大的预训练自编码器的潜在空间中，兼顾计算复杂度和图像细节的平衡。LDM 在图像修复、条件图像合成等任务中取得了新的最优结果，并在多模态训练中拥有很好的表现。后续 Stability AI 也一同联合开源了 LDM 的预训练模型，称为稳定扩散 (Stable Diffusion)¹⁹。为了实现稳定扩散，研究人员使用了低秩矩阵来估计参数更新，从而在保持高质量图像生成的同时大幅减少了参数的数量。这种参数高效性能够在相对较小的计算资源下生成高质量的图像样本，该特性使得稳定扩散方法迅速席卷图像生成领域。

¹⁶ Ho J., et al. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.

¹⁷ Blei D. M., et al. Variational inference: A review for statisticians[J]. Journal of the American statistical Association, 2017, 112(518): 859-877.

¹⁸ Ramesh A., et al. Hierarchical text-conditional image generation with clip latents[J]. arXiv preprint arXiv:2204.06125, 2022.

¹⁹ <https://stability.ai/blog/stable-diffusion-public-release>

2.4.3.2 Stable Diffusion 的基本原理

(1) 去噪扩散概率模型 (DDPM) 是一种基于扩散模型的生成模型, 用于对数据进行建模和生成。扩散模型的基本思想是通过一个前向的迭代扩散过程逐渐将高斯噪声引入数据, 使其分布逐渐趋近于标准高斯分布。然后, 通过学习一个逆向的扩散过程, 将模糊的图像恢复到清晰的状态, 从而重建数据中的结构。

为了训练 DDPM, 研究团队采用了去噪评分匹配框架。这种框架通过前向扩散过程将图像转化为噪声来定义图像的分布。通过训练去噪函数, 使其最小化去噪评分匹配损失, DDPM 可以从随机噪声中生成高质量的样本。

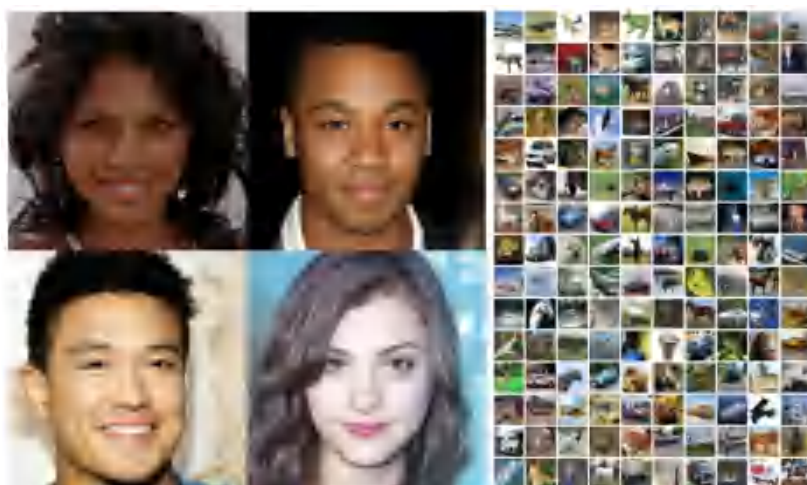


图 2-3 DDPM 生成结果图(来源: DDPM²⁰)

DDPM 在图像合成任务中表现出色, 并在计算机视觉、自然语言处理、波形信号处理、多模态建模、分子图建模和时间序列建模等领域展示了优异的性能。它在无条件图像合成方面已被证明胜过生成对

²⁰ Ho J., et al. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.

抗网络 (GAN) ²¹, 尽管 GAN 在图像质量方面表现出色, 但这些模型往往捕捉到的多样性较少, 难以训练并容易出现模式崩溃问题。在图 3-3 中, 左图是 DDPM 在 CelebA-HQ 数据集上通过去噪评分匹配训练后生成的高分辨率人脸图像, 右图展示了 DDPM 在 CIFAR10 上仅依靠纯噪声生成的图片, 即无条件合成的图片。

(2) DPM 包括前向过程 (forward process) 和反向过程 (reverse process), 其中前向过程又称为扩散过程 (diffusion process)。

扩散过程 (diffusion process), 指对数据逐渐增加高斯噪音直至数据变成随机噪音的过程。对于原始数据 $x_0 \sim q(x_0)$, 扩散过程的每一步都生成一个带噪音的数据 $x_t \sim q(x_t|x_0)$, 通过共计 T 步对上一步数据 x_{t-1} 逐渐增加高斯噪音。

与将数据噪声化的前向过程相反, 反向过程(reverse process)是一个去噪过程, 当反向过程每一步的真实分布 $q(x_{t-1}|x_t)$ 已知, 那么从第 T 步的随机噪声图像 $x_T \sim \mathcal{N}(0, I)$ 开始逐渐去噪便能生成真实图像。反向过程可定义为一个由一系列用神经网络参数化的高斯分布组成的马尔可夫链。

模型设计方面, DDPM 采用的是一个基于残差模块 (residual block) 和注意力模块 (attention block) 的 U-Net 模型。U-Net 属于编码器-解码器 (encoder-decoder) 架构, 其中编码器分成不同的 stages, 每个 stage 都包含下采样模块来降低特征的空间大小。与编码器相反, 解码器将压缩的特征逐渐恢复。U-Net 在解码器模块中还引入了跳连接 (skip connection), 合并了编码器中间得到的同维度特征, 这有利于网络的优化。

2.4.3.3 稳定扩散原理

稳定扩散 (Stable Diffusion) 是一种图像生成框架, 通过在潜在

²¹ Dhariwal P., et al. Diffusion models beat gans on image synthesis[J]. Advances in neural information processing systems, 2021, 34: 8780-8794.

空间上训练扩展的潜在扩散模型（Latent Diffusion Model, LDM）来生成高质量的图像样本。在稳定扩散的第一阶段，引入一个自编码器（AutoEncoder）来学习特征的潜在表示，接着用编码器 \mathcal{E} 对原始图像进行压缩编码，得到图像的低维表征 z_0 ，然后将 z_0 在潜在表示空间（latent space）中进行扩散操作，最后将反向过程后的低维表征通过解码器 \mathcal{D} 恢复到原始图像空间。由于 z_0 是经过压缩的潜在特征，其尺寸远小于原始图像，可显著降低扩散模型的计算成本。第二阶段的扩散过程与标准扩散模型类似，通过逐步引入噪声模糊潜在表示，然后逐步恢复图像的细节和结构。但稳定扩散中引入了条件机制（Conditioning Mechanism），将交叉注意力（cross-attention）作为通用条件实现多模态训练。通过这种扩散过程，模型能够生成多样化且与条件信号相关的高质量图像样本。

实验结果表明(图 3-4),相较于在像素空间上进行扩散建模(Pixel Baseline #1),将扩散建模应用于潜在空间(LDM #1,2,3)能够在降低复杂度和保留图像细节方面取得显著优势。此外,条件机制的设计令稳定扩散能够更好地控制生成图像的多样性和真实性,它可以根据不同的条件生成多种语义合理图像。

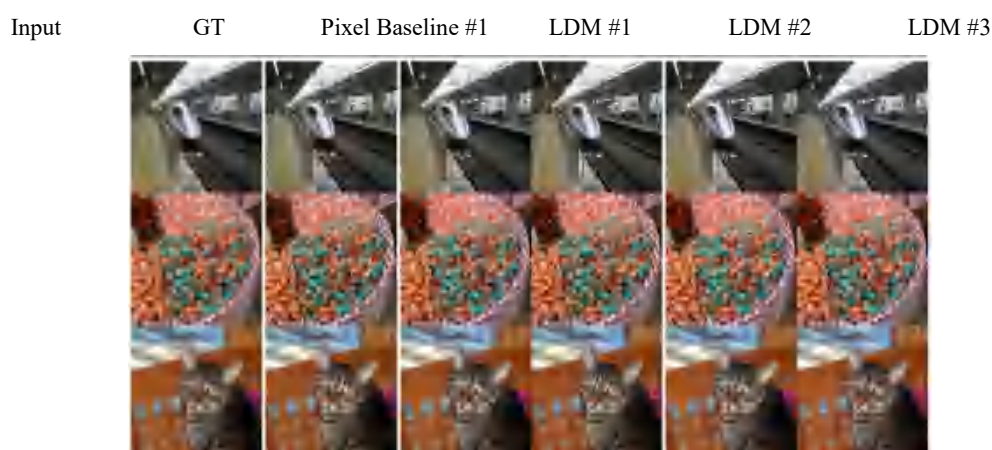


图 2-4 Stable Diffusion 的生成结果图(来源: Stable Diffusion²²)

²² Rombach R., et al. High-resolution image synthesis with latent diffusion models[C]. In Proceedings of the

2.4.4 LoRA

2.4.4.1 LoRA 技术简介

随着预训练模型规模的增大，通过全微调 (full fine-tuning) 重新训练所有模型参数变得愈渐困难。例如，使用 Adam 微调的 GPT-3 175B 模型²³部署独立实例的成本非常高昂，每个实例有 175B 个参数。为了解决这个问题，微软在 2021 年提出了一种名为低秩自适应 (LoRA) 的方法²⁴。

LoRA 的核心思想是冻结预训练模型的权重，并将可训练的秩分解矩阵注入 Transformer 架构的每一层中，这样可以显著减少下游任务需要训练的参数数量。与 GPT-3 175B 相比，LoRA 将可训练参数数量减少 10000 倍，并且减少了 GPU 内存需求。与 RoBERTa²⁵、DeBERTa²⁶、GPT-2²⁷和 GPT-3 等模型的微调结果相比，LoRA 在满足更少训练参数和更高吞吐量的同时，能取得与前者相当甚至更优的性能。此外，与适配器方法不同，LoRA 并没有引入额外的推理延迟。

IEEE/CVF conference on computer vision and pattern recognition, pages 10684-10695, 2022.

²³ Brown T., et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

²⁴ Hu J. E., et al. Lora: Low-rank adaptation of large language models[J]. arXiv preprint arXiv:2106.09685, 2021.

²⁵ Liu Y., et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.

²⁶ He P., et al. Deberta: Decoding-enhanced bert with disentangled attention[J]. arXiv preprint arXiv:2006.03654, 2020.

²⁷ Radford A., et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.

LoRA 从 Li 等人²⁸和 Aghajanyan 等人²⁹的工作中获得灵感，后者表明，学习的过度参数化模型实际上存在于低内在维度上，所以论文提出了一种低秩自适应 (LoRA) 的方法。LoRA 允许通过优化适应过程中密集层变化的秩分解矩阵来间接训练神经网络中的一些密集层，同时保持预训练的权重不变。

假设给定一个由 Φ 参数化的预训练自回归语言模型 $P_{\Phi}(y|x)$ ，LoRA 采用了一种有效的参数方法，令任务特定参数增量 $\Delta\Phi = \Delta\Phi(\theta)$ 由一组更小的参数 $|\theta| \ll |\Phi_0|$ 进一步编码。因此，求 $\Delta\Phi$ 的任务可转变为在 Θ 上进行优化：

$$\max_{\theta} \sum_{(x,y) \in \mathcal{Z}} \sum_{t=1}^{|y|} \log(p_{\Phi_0 + \Delta\Phi(\theta)}(y_t | x, y_{<t}))$$

2.4.4.2 LoRA 在 Stable Diffusion 上的应用

LoRA (Low-rank Adaptation) 技术提供了一种快速微调扩散模型的方法，使得在不同概念（例如角色或特定风格）上训练 Stable Diffusion 模型更加便捷。经过训练的这些模型可以导出并供他人在其自身生成任务中使用。

Stable Diffusion 模型因其出色的图像和文本生成能力而在机器学习领域受到广泛关注。然而，这些模型的主要缺点之一是它们的文件大小较大，给个人计算机的维护带来困难。LoRA 技术通过引入低秩适应方法，为稳定扩散 (Stable Diffusion) 模型带来了诸多优势，

²⁸ Li C., et al. Measuring the intrinsic dimension of objective landscapes[J]. arXiv preprint arXiv:1804.08838, 2018.

²⁹ Aghajanyan A., et al. Intrinsic dimensionality explains the effectiveness of language model fine-tuning[J]. arXiv preprint arXiv:2012.13255, 2020.

特别是在应用于图像与文本生成任务方面。

使用 LoRA 技术训练 Stable Diffusion 具有几个关键优势：

1. 高效的微调过程

LoRA 允许使用低秩适应技术对扩散模型进行快速微调，相较于传统的优化方法，其在微调过程中大幅降低了计算成本和硬件要求。由于 LoRA 不需要计算所有参数的梯度或维护优化器状态，仅需优化注入的远比原模型小的低秩矩阵，从而大幅简化了微调的计算复杂度。

2. 降低硬件门槛

由于稳定扩散模型本身在训练和推理过程中需要较大的计算资源和存储空间，这使得普通用户难以维护一个包含大量模型的收藏。LoRA 技术的引入有效缩减了模型的文件大小，将其从原来的几 GB 级别降低至 2-500 MB，使得用户能够更轻松地在个人计算机上运行和应用稳定扩散模型。

3. 保持模型质量

虽然 LoRA 将稳定扩散模型文件大小大幅减小，但在此过程中并没有牺牲模型的生成质量。LoRA 技术在文件大小和训练能力之间取得了良好的平衡，使其成为拥有大量模型的用户们的理想选择。用户可以获得高质量的图像与文本生成结果，而无需担忧过大的模型参数和计算成本。

4. 导出与分享

经过使用 LoRA 技术微调的稳定扩散模型仍然保持了较高的生成能力和多样性，这使得用户可以轻松导出和分享这些经过优化的模型供他人使用。导出的模型可以用于在不同概念（如角色或特定风格）上进行图像和文本的生成任务，为研究人员和开发者提供了更多实验和应用的可能性。

总的来说，LoRA 技术在稳定扩散模型中的应用为用户提供了更

高效的模型训练和应用体验。通过降低硬件门槛、保持模型质量和方便的模型导出，LoRA 为生成高质量图像和文本的研究和应用带来了更多的便利和可能性。

2.4.5 ControlNet

ControlNet³⁰是一种控制预训练大型扩散模型的方法，它可以支持额外的输入条件。通过端到端的学习方式，即使在训练数据集很小的情况下，ControlNet 也能够有效地学习特定任务的条件输入，同时保证训练速度与微调扩散模型相当。ControlNet 能够成功地增强诸如 Stable Diffusion 这类大型扩散模型，使其能够支持边缘图、分割图、关键点等条件输入，丰富了大型扩散模型的控制策略，促进其应用发展。

1. ControlNet 网络设计

Stable Diffusion 模型基本上是一个 U-net 结构，包含一个编码器、中间块和一个带有跳跃连接的解码器。如果不加 ControlNet，扩散模型原始的神经网络 \mathcal{F} 输入 x 后获得 y ，其参数用 θ 表示，则：

$$y = \mathcal{F}(x; \theta)$$

ControlNet 对神经网络模块的输入条件进行操作，进一步控制网络的整体行为。它将模型原始的网络参数 θ 固定，并复制了一个原始模型分支 θ_c ，在此基础上应用了外部条件向量 c 进行训练，并将施加控制条件的分支通过零卷积和原始模型分支逐层相加。这种设计的优势在于它能够避免在小数据集上过拟合，并保留对数十亿张图像学习的原始大模型能力。加了控制条件之后，将原始网络的输出修改为：

$$y_c = \mathcal{F}(x; \theta) + \mathcal{Z}(\mathcal{F}(x + \mathcal{Z}(c; \theta_{z1});); \theta_{z2})$$

³⁰ Rombach R., et al. High-resolution image synthesis with latent diffusion models[C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684-10695, 2022.

其中零卷积层 \mathcal{Z} (zero convolution) 的初始化权重和偏置都为 0，两层零卷积的参数为 $\{\theta_{z1}, \theta_{z2}\}$ 。

需要注意的是，未经训练的 ControlNet 分支输出为零，因此添加到原始网络的数值也为零。这对原始网络没有任何影响，确保了原网络性能的完整保留。

2. 基于稳定扩散的 ControlNet

(1) 训练过程如下：

原始的 Stable Diffusion 的优化目标如下：

$$\mathcal{L} := E_{\mathcal{E}(x), \epsilon \sim \mathcal{N}(0,1), t} [\|\epsilon - \epsilon_{\theta}(z_t, t)\|_2^2]$$

在采样 z_t 经过网络 ϵ_{θ} 进行去噪后，与原始特征经过网络 ϵ 后得到的潜变量计算 L2 损失，而 ControlNet 的训练目标改进为：

$$\mathcal{L} = E_{z_0, t, c_t, c_f, \epsilon \sim \mathcal{N}(0,1)} [\|\epsilon - \epsilon_{\theta}(z_t, t, c_t, c_f)\|_2^2]$$

其中网络 ϵ_{θ} 增量了两个控制条件，文字条件 c_t 和任务指定的条件 c_f ，例如 Canny 边缘图等。

(2) ControlNet 的应用如下：

ControlNet 的推理过程中使用 DDIM³¹(Denoising Diffusion Implicit Models) 的采样方式，默认使用 20 个时间步长。在这个推理过程中，用户可以选择不同的 prompt 模式来指导生成图像，其中 prompt 可以为空字符串、默认的“专业、详细、高质量图像”语句、利用 BLIP³²等模型自动生成的图像标注，或者用户自定义的词句。实验发现 ControlNet 在各种不同条件任务中都能取得较高质量的生成结果。

Canny 边缘检测图。当使用简单的 Canny 边缘检测来提取图像的纹理信息时，可以通过这些纹理信息生成各种不同风格的图像，这些

³¹ Song J., et al. Denoising diffusion implicit models[J]. arXiv preprint arXiv:2010.02502, 2020.

³² Li, J., et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation[J]. International Conference on Machine Learning. PMLR, 2022.

图像呈现出逼真而生动的效果。



来源：ControlNet^[4]

图 2-5 ControlNet 基于 Canny 边缘条件生成结果图

人体姿势图。基于学习的姿势估计方法令从人体姿势关键点到自然图像的转换也轻而易举，这类条件控制模型为各种人机交互、动画、虚拟现实等应用领域带来了更广阔的发展空间。



图 2-6 ControlNet 基于人体姿势条件生成结果图（来源：ControlNet^[33]）

³³ Ho J., et al. Denoising diffusion probabilistic models[J]. Advances in neural information processing

卡通线图。利用网络上的卡通插图提取线描并生成彩色图像，ControlNet 能够从卡通插图中捕捉线条轮廓，然后将其填充上合适的色彩，使得上色过程变得自然而轻松，同时呈现出生动的笔触效果。



图 2-7 ControlNet 基于卡通线图条件生成结果图（来源：ControlNet³⁴）

2.4.6 NeRF 作为新型三维重构

三维重建与新视角渲染作为计算机视觉中的一个核心议题，有着非常广泛的应用。然而基于传统多视角几何的方法对于稠密渲染任务性能欠佳，且计算量巨大。自 2020 年起，基于神经网络的神经渲染类方法取得了长足的进步，本文旨在简单回顾 NeRF 的原理以及潜在

systems, 2020, 33: 6840-6851.

³⁴ Ho J., et al. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.

的应用。

2.4.6.1 NeRF 基本原理回顾

传统的 3D 重建与新视角渲染方法使用 Mesh, Point Cloud, Voxel 等方式来显式表示 3D 场景，然而由于显式表示的分辨率低，想获得照片级的渲染结果对于内存与计算的消耗都是指数级上升。隐式表示从另外一个角度解决了这个问题：隐式表示通常使用一个(可学习的)函数来描述场景，在 NeRF 出现之前，最为成功的案例是使用神经网络来近似 Signed Distance Function 表示物体表面。

NeRF (Neural Radiance Field) 首次将隐式表示带入了场景表示中来。简单来说，NeRF 训练的核心是得到以下函数：输入 3D 空间中的 (x, y, z) 坐标，以及对这个位置的观察角 (θ, ϕ) ，输出该位置对应的颜色信息 (r, g, b) 以及对应的体密度。

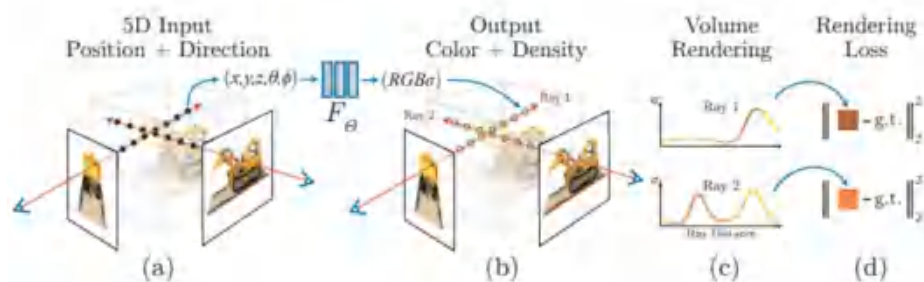


图 2-8

基于这样的隐式场景表示，我们可以通过体渲染的方式得到任意观察视角的场景图像。简单地说，我们首先离散化该观察视角对应的射线得到一系列的位置，对于每个位置，通过该隐式神经网络预测对应的体密度与颜色值，后使用每个位置的体密度归一化值对对应颜色加权平均即可得到对应该像素的 RGB 颜色值。

在训练阶段，我们输入同一场景的多张图像，以及对应的观察视角。对于每张图像的像素使用该隐式神经网络计算渲染结果，并计算损失函数，通过反向传播的方式优化该场景对应的隐式神经网络的参数，从而最终得到的隐式表示。对于新视角合成，我们则简单地使用此训练好的神经网络以及视角参数使用体渲染方法即可合成。

2.4.6.2 基于 NeRF 的新兴应用

NeRF 方法本身后续有一系列的加速工作，使得 NeRF 的应用场景更加广泛，这里就不再展开，下文主要介绍一下基于 NeRF 原理的一系列新兴应用。

1. 构建数字人体

IMAvatar 使用隐式神经网络从多帧人脸视频中分别学习几何、纹理、姿态和表情相关的形变场，并通过上述可微分渲染的方式进行端到端训练。在训练结束之后，可以无限制地操作虚拟形象外插至未见过少见的姿态和表情中。

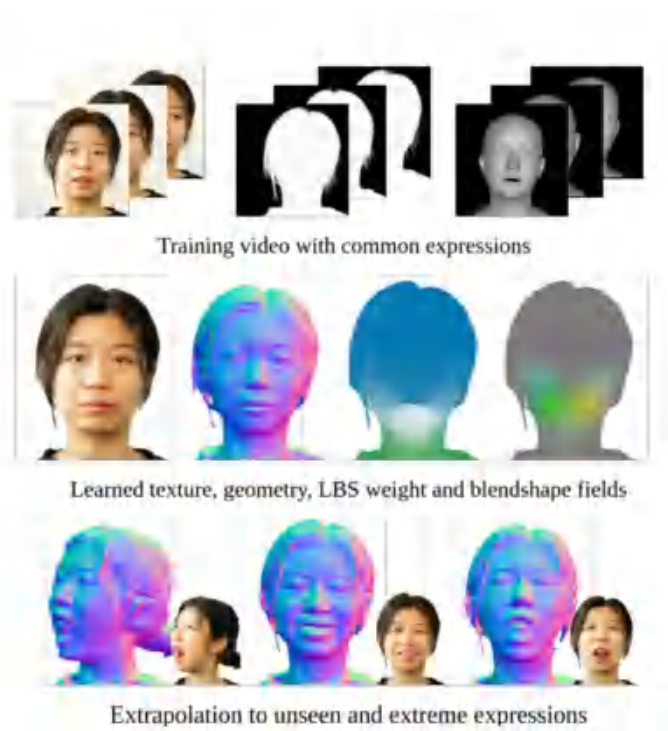


图 2-9

ICON 使用多帧无限制的分割完毕的人体全身图像作为输入，首先鲁棒地估计出每帧图片的法向图，并基于此使用隐式 3D 表示重建出人体形态。基于这样重建的结果，可以重新根据给定的姿态和需求渲染出虚拟形象，极大地拓展了数字世界中构建虚拟形象的成本与复杂度。



图 2-10

2.3D 场景的可控生成编辑

GIRAFFE 通过结合 AIGC 与 NeRF 的能力，实现了 3D 场景的解耦与可控编辑。具体而言，GIRAFFE 拓展了标准 NeRF 的隐式神经网络表示，对于场景中的每个物体在输入层面加入了代表形状与外观的隐编码，同时将输出中的 RGB 颜色变为了可供后续网络使用的隐式特征。随后 GIRAFFE 将场景中所有物体输出的隐式特征进行融合，并使用解码器进行体渲染从而得到最终的场景。

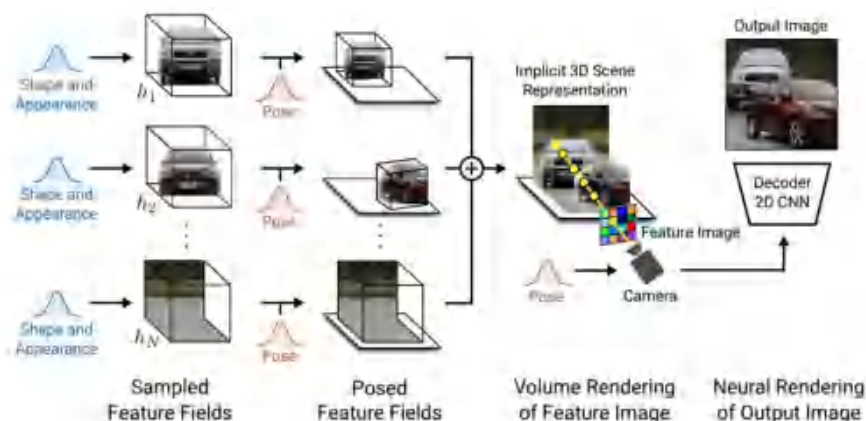


图 2-11

在上述流程中，可以通过改变每个物体的隐编码从而单独控制图像中每一部分的生成内容，从而达到符合用户需求的可控生成的目的。如下图所示，可以分别改变物体的方向、外观或观察的视角而不改变背景内容。



图 2-12

3. 真实感交通流仿真

NeRF 也可用于从真实的场景视频中构建虚拟世界。从技术方面而言，挑战在于需要能够预先解耦开静态场景与动态物体。在一些特定的场景中，进行动静态分割是一件比较容易的事情，例如驾驶场景。故可以解耦开两者分别使用 NeRF 进行建模渲染，而后再行融合，达到对虚拟世界可控编辑的目的。

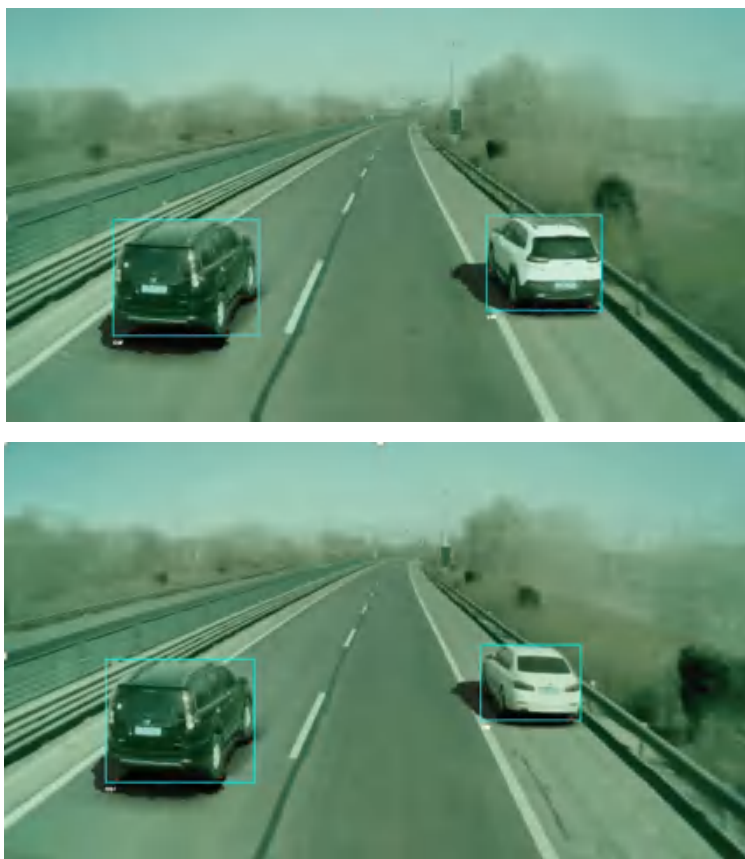


图 2-13、图 2-14

2.4.7 AIGC 技术发展展望

2.4.7.1 发展建议

1. 构建全民共享和多元共治发展格局

行业组织、媒体机构、职业群体等各类社会力量应及时做好正面宣传，可以通过 AIGC 技术在大型公共活动中的应用等方式，帮助广大群众及时了解 AIGC 应用水平、行业动态及发展趋势，推动 AIGC 技术较快较好融入各类场景。

鼓励公众参与和社会监督，促进 AIGC 技术的公正和透明。加强对 AIGC 技术的公众参与和社会监督，建立 AIGC 技术的公开透明机制，加强对 AIGC 技术的评估和监测，及时公布 AIGC 技术的使用情况和效果，推动 AIGC 技术的公正和透明。此外，还需要加强对 AIGC 技术的社会影响的研究和评估，推动 AIGC 技术的可持续发展。同时，

需要加强对 AIGC 技术的用户教育和培训，提高用户对 AIGC 技术的认知和理解，促进用户对 AIGC 技术的正确使用和评价。

2. 呼吁自律自治、共建产业发展生态

政产学研用各主体基于开源共享平台促成协同合作、加速应用创新。AIGC 作为一个新兴领域，需要各方的共同努力和合作来推动其健康发展。首先，政府可以提供政策支持和资源投入，鼓励产业界、学术界和研究机构参与 AIGC 领域的研究和创新。政府可以通过制定相关政策和法规，为 AIGC 的应用创新提供良好的环境和条件。其次，产业界可以加强合作，共同研发和推广 AIGC 技术和产品。通过共享资源和经验，产业界可以加速应用创新，推动 AIGC 技术在各个行业的广泛应用。同时，学术界和研究机构可以加强对 AIGC 技术的研究和探索，提供技术支持和创新思路。他们可以通过开展合作研究项目、共享数据集和算法模型等方式，促进 AIGC 技术的进一步发展。最后，开源共享平台的建设和推广也是促进 AIGC 领域共建产业发展生态的重要手段。通过建立开放的平台，各主体可以共享资源、交流经验，促进协同合作和创新。AIGC 作为一个涉及人工智能技术和内容创作的领域，面临着一系列的风险和挑战。

在保障 AIGC 的健康发展过程中，行业层面的自律和伦理意识至关重要。首先，行业各方应加强对 AIGC 技术和应用的伦理考量，确保生成内容的合法合规，避免违法和不道德的行为。其次，行业组织和企业应建立自律机制，制定行业准则和规范，约束行业内的行为，防止不良竞争和不当行为的发生。同时，行业层面应加强对 AIGC 技术的监管和管理，确保技术的安全性和可信度。此外，行业组织和企业还应加强与政府和社会各界的合作，共同推动 AIGC 领域的风险治理和规范建设。

3. 加速工程实践，践行社会责任

随着数字经济的快速发展，数字内容的需求不断增长，AIGC 作

为数字内容的创作和生成方式，具有巨大的潜力和市场需求。为了更好地满足市场需求，人工智能企业可以在提高算法和算力等基础能力的基础上，加强对 AIGC 技术的工程化研发。首先，可以完善数字内容模板库和知识图谱的储备，开发标准化的创作工具，提高 AIGC 工具体系的完善程度。其次，可以通过优化模型开发和技术管理流程等方式，提高定制化解决方案的交付效率，加快产品和服务的研发和推广。此外，人工智能企业还可以延伸到数字内容产业的上下游环节，拓展服务范围，加速构建产业发展的先发优势。通过加速 AIGC 技术的工程化，人工智能企业可以更好地满足市场需求，推动 AIGC 领域的发展和产业化进程。

相关企业践行科技向善理念、积极履行社会职责，针对生成内容主动进行合理标识。特别是对产业链和用户等多方主体起到纽带作用的大型互联网企业和平台型企业，应加强与政府部门和行业组织的治理合作。通过加强鉴别和审核技术能力部署、支持平台用户自主标记不良内容、建立应急辟谣机制、发布平台审核透明度报告等方式，主动承担风险防范、内容治理责任。同时，政府部门应加强对 AIGC 技术的监管。

4. 持续正向引导，加强安全和法律保障

相关行业主管部门可以通过政策引导充分释放 AIGC 的正面应用价值。在 AIGC 的发展中，政府部门可以发挥引导作用，通过制定相关政策和法规，为 AIGC 技术的应用提供支持和指导。一是引导数字内容产业的发展，鼓励企业在 AIGC 领域进行创新和应用。政府可以提供资金支持、税收优惠等政策措施，以促进数字内容产业的繁荣发展。二是坚持“以虚促实、以虚强实”的基本导向，即通过 AIGC 技术的应用推动实体经济的发展。例如，在制造业领域，可以利用 AIGC 技术优化生产流程、提高生产效率，从而推动实体经济的转型升级。三是推动 AIGC 技术的全球化发展。例如，鼓励 AIGC 技术的国际合

作和交流，加强技术交流和人才培养，加强对 AIGC 技术的国际标准制定和推广。最后，推动 AIGC 技术与其他各类产业深度融合，加速渗透到经济社会的方方面面。例如，可以利用 AIGC 技术打造逼真的在线服务体验，催生新的服务模式和商业模式，鼓励弱势或特殊群体援助，促进教育资源公平等社会民生领域的应用场景构建。借助 AIGC 技术还可以探索虚拟空间模拟仿真，对现实世界进行更为自然的人机交互和更为逼真的环境细节，从虚拟维度优化现实世界的运转，提升效力和价值再造，应用于工业制造、城市规划、科学研究、生产办公等领域。

加强 AIGC 技术的安全保障，防范 AIGC 技术的滥用和恶意攻击。建立 AIGC 技术的安全保障体系，加强对 AIGC 技术的安全漏洞和风险的监测和预警，及时采取措施进行应对和处置。此外，还需要加强对 AIGC 技术的伦理和社会责任的研究和探讨，推动 AIGC 技术的公正和透明。同时，需要加强对 AIGC 技术的法律法规制定和执行，明确 AIGC 技术的使用范围和规范，防范 AIGC 技术的滥用和恶意攻击。

明确 AIGC 相关法律，并开展 AIGC 作品权属认定方面的法律法规研究。目前，我国的著作权法律体系主要针对人类创作的作品进行规定，对于 AIGC 生成的作品的权属认定存在一定的模糊性和困难。因此，有必要明确 AIGC 相关法律，为 AIGC 作品的权属认定提供明确的法律依据。同时，还需要开展 AIGC 作品权属认定方面的法律法规研究，探索制定模仿、抄袭或合理使用的认定标准，明确侵权主体及侵权责任的认定路径，以保护 AIGC 作品及其智力和经济投入的合法权益。此外，还需要考虑 AIGC 作品所依赖的数据基础和技术工具的保护，确保相关数字内容的合规有序流通。通过明确 AIGC 相关法律和开展 AIGC 作品权属认定方面的法律法规研究，可以为 AIGC 领域的发展提供法律保障，促进 AIGC 技术的健康发展。

2.4.7.2 未来展望

1. 技术创新与演进

(1) 向多模态复杂场景发展。

面对多样性和复杂场景内容生成的挑战，AIGC 的算法能力仍然需要进一步提升。例如，AIGC 已经可以生成逼真的高清人脸图像或数字人头像。但是他依旧难以在视频模态上取得相同惊叹的效果，因为动画视频的动态复杂性和可能性远远超过静态图像。为了应对这一挑战，AIGC 的下一个发展方向将是通过不同生成个体之间的交互来进行内容创作，通过对整体的、多模态的复杂场景进行创作。AIGC 将能够生成更多多样化的内容，并推动核心和相关领域的共同发展。这将有助于构建更加丰富和多元的数字世界，为用户提供更具创意和互动性的内容体验。

(2) 向大规模分布式发展。

作为数字经济和实体经济深度融合的新模式，AIGC 利用新型人工智能技术创造和制作丰富的数字内容，这给人工智能技术研究带来了新的挑战。在这个背景下，AIGC 将从本地化集中式发展为大规模分布式，通过多智能体算法的研究，实现多智能体之间的协同工作和交互，推动 AIGC 技术的发展。具体来说，首先，随着数据量的不断增长和模型的复杂性提高，传统的单机深度学习算法已经无法满足 AIGC 的需求。通过将计算任务分布到多个计算节点上，大规模分布式深度学习技术可以充分利用分布式计算资源，加速模型训练和生成内容的速度。同时，分布式计算还可以提高模型的容错性和鲁棒性，使 AIGC 系统更加稳定和可靠。其次，多智能体强化学习技术在 AIGC 中也具有重要意义。AIGC 系统往往需要模拟复杂的交互场景，其中涉及多个智能体之间的合作和竞争。多智能体强化学习技术可以帮助 AIGC 系统学习到合适的策略和行为，以实现预期的生成内容效果。通过多智能体强化学习，AIGC 系统可以在复杂的环境中进行自主决

策和学习，从而提高生成内容的质量和多样性。

2. 关键能力的增强

随着 AIGC 技术的不断发展，其核心能力将得到显著增强。通过内容孪生技术，AIGC 能够更准确地复制现实世界的物体和场景，实现更逼真的虚拟体验。渲染技术的提升是实现内容孪生的关键。随着计算能力的增强和算法的改进，渲染技术在 AIGC 中得到了显著的发展。现在的渲染技术能够更加准确地模拟光照、材质和纹理等物理属性，使得虚拟世界中的物体和场景看起来更加逼真。同时，仿真加速技术的进步也使得渲染过程更加高效，大大缩短了生成内容的时间成本。通过内容编辑技术，AIGC 能够实现现实世界与虚拟世界的双向交互，将现实世界的问题转化为虚拟世界的计算问题，并提供最优解。用户可以通过现实世界的输入，对虚拟世界中的内容进行修改和定制，使其更符合自己的需求。同时，AIGC 也可以将虚拟世界中的内容输出到现实世界，为用户提供更加丰富、多样化的体验和服务。这种双向通道的打通将进一步推动 AIGC 在各个领域的应用和发展，为数字经济的发展注入新的动力。通过内容创作技术，AIGC 能够自主生成内容，满足用户对数量和质量的三重需求。基于内容创作的数字原生能力的潜能将在多个方面得到释放。首先，AIGC 将能够摆脱对专业生成内容（PGC）和用户生成内容（UGC）的依赖，实现自主内容生成。通过更先进的算法和模型，AIGC 可以创作出更加精彩、独特的内容，满足用户对多样化、个性化内容的需求。其次，基于内容创作的数字原生能力还将推动 AIGC 与其他技术的融合。例如，结合虚拟现实（VR）和增强现实（AR）技术，AIGC 可以创作出更加沉浸式、交互性强的数字内容，为用户提供更加丰富、真实的体验。同时，与区块链技术的结合可以确保内容的版权和真实性，为内容创作者提供更好的保护和激励机制。

3. 产品类型的丰富

随着元宇宙概念的兴起以及科技进步，快速发展的数字人是未来 AIGC 发展的一个方向。以自主生成内容的数字人将成为构建人机交融、数实融合的未来世界的最佳载体；同时数字人作为构建元宇宙新世界的关键要素，也是人类构建“自生成、自更新”的元宇宙新世界的必经之路。数字人是通过 AIGC 技术生成的虚拟实体，具备自主生成内容的能力。随着 AIGC 技术的不断发展，数字人将具备更加智能化、个性化的特征，能够与人类进行更加自然、互动的交流。数字人作为人机交互的媒介，将打破传统的物理限制，通过虚拟现实、增强现实等技术，为用户创造出沉浸式、交互性强的体验。数字人作为现实与虚拟世界的交互媒介，将成为构建人机交互、数实融合的重要载体。数字人不仅是人机交互的载体，也是构建元宇宙新世界的关键要素。元宇宙是一个虚拟的数字世界，与现实世界相互交织、相互影响。数字人作为元宇宙中的虚拟实体，将扮演着重要的角色。它们可以自主生成内容，创造出丰富多样的虚拟体验，为用户提供与现实世界不同的、自由探索的空间。数字人的存在，使得人类可以在元宇宙中实现“自生成、自更新”，创造出一个充满创造力和无限可能性的新世界。

4.多元化的应用场景

随着技术的不断进步和创新，AIGC 的应用将进一步扩展，不仅局限于传统的领域，还将涉及到更多新兴领域和行业。未来，AIGC 的应用场景将进一步多元化。除了已经涉及到的金融、传媒、文娱、电商等领域，AIGC 将逐渐渗透到教育、医疗、工业等各行各业。在教育领域，AIGC 可以为教育工作者提供新的工具，将抽象的知识转化为具体、生动的形式，提供更加个性化、互动性强的教学内容。在医疗领域，AIGC 可以辅助医学图像的分析和诊断，提高医疗的准确性和效率。在工业领域，AIGC 可以应用于智能制造和自动化生产，提高生产效率和质量。除此以外，AIGC 还将呈现出场景与场景的融合交互的趋势。随着技术的发展，不同场景之间的融合将成为未来的

发展方向。例如，在虚拟现实和增强现实技术的支持下，AIGC可以与现实世界进行无缝衔接，创造出更加沉浸式、交互性强的体验。通过场景之间的融合交互，AIGC可以为用户提供更加丰富、个性化的数字体验，推动数字经济的发展和社会的进步。

2.5 生成式 AI 的工程化

2.5.1 生成式 AI 工程化面临的主要问题

扩散模型的优异性能，以及模型微调技术的发展，激发了人们探索生成式 AI 在实际生产生活中应用的热情，也在很短的时间内深深影响了一些行业和职业。自 2022 年底，以文生文产品 ChatGPT 和文生图产品 Midjourney 为代表，一系列移动应用、商用产品和开源项目蓬勃涌现，将生成式 AI 推向了寻常百姓家，打开了生成式 AI 技术服务具体业务场景，实现工程化落地的新篇章。

工程化是指在某种产品、项目或系统的设计、开发和运作过程中，将一系列最佳实践、标准化和自动化应用于其中，以确保其高效、稳定、可靠和可维护。生成式 AI 工程化的目的是提高效率、降低成本、保证质量，从而推动生成式 AI 技术价值的实现。根据目前生成式 AI 技术发展与应用现状，在开展生成式 AI 工程化实践中，主要需要考虑以下三方面的问题。

(1) 确保良好的用户体验。工程化以应用为最终目的，需要确保使用者的体验，从而使用户能够高效地完成生成式 AI 的工作。软件工程的先进理念和技术资源为此打下了基础。众多的前端架构不仅使生成 AI 应用的用户界面美观易用，交互体验好，而且提高了开发的效率。以服务为中心的设计理念，以及模块化和 API 技术的最佳实践，确保了应用系统的构建效率和易维护性。插件式结构与开源软件模式，能充分激发技术社区的活力，推动应用的快速迭代，为用户不断提供和升级丰富的功能。

(2) 灵活获取算力资源。生成式 AI 应用的一个重要特点，就是需要大量的算力资源。资源有限、GPU 一卡难求，是很多企业用户面临的问题；本地单机工作的效率不高，常常需要多个任务排队等待，是很多个人用户的现状。此时，利用云计算资源，按需使用，动态分配资源，成为缓解上述算力需求压力的途径之一。

(3) 稳定控制生成效果。目前生成式 AI 的生成结果往往带有随机性和不确定性，原因复杂多样，例如：生成式 AI 的基础模型通常基于概率模型，生成过程中利用随机性；训练数据不足或不够多样化，使模型无法捕捉到数据的潜在分布，从而导致生成结果不稳定；某些模型对初始文本或提示敏感，输入条件的微小变化可能导致生成结果的巨大差异；等等。这一问题也导致了 AI 生成类视频应用的成熟度明显落后于 AI 生成图像，因为前者需要解决帧与帧之间的生成稳定性。要解决这一问题，不仅需要在理论上获得突破，也需要在技术实践中进行探索。

除此之外，伴随着工程化实践的开展，有些问题也在逐渐显现，需要引起我们的关注。模型是生成式 AI 应用的基础，当前大量生成式 AI 模型不断涌现，性能日新月异。一方面，基础模型提供商 (model provider)、使用基础模型调优提供服务的厂商 (model tuners)，以及可能涉及的监管机构面临着安全、规范、便捷地发布与管理模型的问题。另一方面，面对如此众多的模型，用户也面临着高效搜索与调用模型的问题。生成式 AI 大大提升了创作效率，同时也生成了海量的媒体资源。因此，需要高效地对这些资源进行存储、管理与价值挖掘。处理得当，它们对用户是宝贵的资产；处理不当，反而会成为用户的包袱，造成信息过载和额外的应用成本。

2.5.2 AI 生图应用平台

用户通过应用平台设置参数、调用生成式 AI 模型，并获得生成

结果和系统的反馈信息，简单易用、功能完善的应用平台是用户高效完成生成式 AI 工作的保障。在 AI 图像生成领域，应用平台的用户人群主要包括算法工程师、制作人、设计师三类。他们虽然职能不同，但对于生成式 AI 应用平台的需求在使用成本、使用难度、功能等方面是相通的。

(1) 算法工程师主要负责生成式 AI 模型的训练和调优。一方面，他们利用现有素材，根据业务需要，对基础模型微调，以训练出能生成某一风格或某一类图像的模型；另一方面，他们需要不断输入各种参数，用生成的模型进行推理，生成测试图像，以检验模型是否达到预期，能进入实际生产环节。算法工程师需要在训练和推理工作之间不断迭代，通过代码搭建工作流程，在命令行和编程环境完成工作。生成式 AI 应用平台提供可视化的用户界面，用于控制参数的快速调节，帮助算法工程师减少代码工作量，将注意力集中在模型效果上，提升工作效率。

(2) 制作人负责提出具体设计要求，并评估设计成果。在传统的设计流程中，由于设计操作的专业性，制作人往往是将设计理念通过文字表达出来，由设计师按照文字要求进行设计。要达成制作人设想的效果，往往需要制作人和设计师之间反复沟通，多轮修改。制作人利用生成式 AI 应用平台，可以自己直接将创意通过图像表达出来，再由设计师根据具体要求进行精修，大大节省了需求方与设计方在前期沟通中花费的精力。

(3) 设计师负责执行具体的设计操作，不断调试参数，以获得需要的设计效果。设计师利用生成式 AI 应用平台，快捷、批量地生成大量的设计原稿，高效地标记与挑选生成图像结果，并打通设计 workflow，将原稿作为输入应用到后续环节进行进一步的人工修改与 AI 生图。

目前，以 Midjourney³⁵为代表的 SaaS 化商用软件，以 Stable Diffusion WebUI³⁶、ComfyUI³⁷为代表的开源应用是用户使用最为广泛的 AI 图像生成领域的应用平台。表 1 对比了两类平台的主要特点。Midjourney 以简洁易于上手的操作界面和可媲美专业水准的出图效果，率先吸引了大众对 AI 图像生成的关注。Stable Diffusion WebUI 是应用最广泛的应用软件之一。Stable Diffusion WebUI 是一个基于 Stable Diffusion 模型的开源项目，利用 Gradio 模块搭建出网页形式的图形化用户界面。SD WebUI 备受欢迎的原因包括：(1) 开源免费，功能快速迭代，支持最新的 AI 生图功能；(2) 安装配置操作简单，普通用户易于上手；(3) 提供丰富全面的控制参数，专业用户可以精细地掌控创作过程；(4) 采用插件机制扩展支持多种功能，吸引了开源社区大量人员参与贡献。ComfyUI 是另一个广受关注的 AI 图像生成应用的开源项目。与 Stable Diffusion WebUI 类似，它遵循 GPL-3.0 协议开源，普通用户免费使用，可本地部署，无须联网，但建议配备 GPU 卡。与 Stable Diffusion WebUI 最大的不同在于，其应用界面采用节点式的工作流设计模式，适合有 3D 设计软件使用经验的用户。虽然它的上手门槛较高，但具有异步队列系统等 Stable Diffusion WebUI 尚不具备的功能。

表 2-1 Midjourney 与 Stable Diffusion WebUI 主要特点比较

	Midjourney	Stable Diffusion WebUI
费用	有免费出图额度限制，超过限额则需要购买会员	遵循 AGPL-3.0 协议开源，普通用户不需要付费

³⁵ <https://www.midjourney.com/>

³⁶ <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

³⁷ <https://github.com/comfyanonymous/ComfyUI>

	Midjourney	Stable Diffusion WebUI
操作体验	界面简洁，操作简单，主要依靠提示词控制，出图可控性不强	界面功能丰富，出图过程可控性强
运行要求	无本地硬件要求，需要联网运行	无须联网，可完全本地部署，但建议本地配备 GPU 卡，否则运行效率很低
模型训练	闭源，无法自定义模型和二次训练	开源，用户可以自己训练模型
扩展性	无插件，功能有赖于原厂开发和更新	采用插件机制，提供丰富的插件，社区共建

截至 2023 年 7 月底，Stable Diffusion WebUI 项目在 GitHub 上已获得 93k 颗星，18.3k 个 fork，比 ComfyUI 均高出一个量级。同时，ComfyUI 可以通过插件的形式在 Stable Diffusion WebUI 上使用。因此，目前对于个人创作者和企业用户，从使用成本、上手难度、功能与可扩展性等方面综合考虑，Stable Diffusion WebUI 具有优秀的用户体验，是开展 AI 图像生成实践的首选。

2.5.3 利用云计算资源应对算力需求挑战

2.5.3.1 Stable Diffusion WebUI 的应用瓶颈

虽然 Stable Diffusion WebUI 非常流行，它的图形化操作界面大大降低了普通用户进行 AI 生图的门槛，但在实际应用中，用户通常会面临一些限制，影响工作和创新的效率。Stable Diffusion WebUI 是单机版应用，对设备要求高，需要配有图形处理单元 (GPU) 才能顺畅地生成图像。用户一般使用带有 GPU 显卡的本地计算机或租用云环

境中带有 GPU 设备的虚拟机，在其中安装和运行 **Stable Diffusion WebUI**。更为严重的是，单机版的计算资源有限，每次只能执行一个工作负载。一旦当前任务开始，界面就不会再允许提交新任务，直到当前的任务完成。这种单机运行模式会极大地影响用户开展 AI 生图工作的效率。

Stable Diffusion WebUI 在实际应用中面临的限制，是普遍存在的生成式 AI 高算力要求的一个侧影。由于生成式 AI 需要处理庞大的数据集和复杂的模型，其训练和推理的过程需要大量的计算资源。运行较大规模的生成式 AI 模型，通常需要高性能的硬件设备，如图形处理单元 (GPU) 或专用的 AI 加速器。这些设备能够并行处理大规模的矩阵运算和神经网络计算，以加快模型的训练和生成速度。此外，生成式 AI 的训练过程通常需要较长时间，甚至需要多天或数周才能完成。这期间，对于训练过程的自动化监控与调度，以及对结果的自动评估，都是生成式 AI 在实际中得到更大范围应用所面临的挑战。

2.5.3.2 向云端寻求弹性算力资源

随着生成式 AI 的发展，面对更大规模的数据集、更复杂的模型和更高质量的生成结果需求，人们将目光投向云计算服务，以获得更强大的计算能力和工程化平台的支持。实际上，目前很多流行的大模型都是在云计算平台上完成的训练。例如，**Stable Diffusion** 模型就是在 **Amazon SageMaker** 服务上完成训练的³⁸。借助 **Amazon SageMaker**，**Stability AI** (**Stable Diffusion** 模型的开发公司) 在具有数千个 GPU 或 **Amazon Trainium** 芯片的计算集群上构建 AI 模型，从而将训练时间和成本缩短 58%。

利用拥有弹性和高可用算力支持的云计算已成为大规模开展生成式 AI 实践的重要途径。通过云计算平台，生成式 AI 可以利用弹性

³⁸<https://aws.amazon.com/blogs/machine-learning/stability-ai-builds-foundation-models-on-amazon-sagemaker/>

的资源分配和并行计算能力，加快训练和推理速度；同时，还可以利用丰富的云服务构建起灵活实用、可快速迭代的工程化系统，为创新和应用提供基础环境。

在云端拓展算力的方式主要有两类，以 **Stable Diffusion WebUI** 的使用为例：

(1) 云虚拟主机。用户在云计算平台租用配有 GPU 等加速设备的虚拟机，AI 生成工作完成后即可释放资源。目前，很多云计算服务商针对生成式 AI 业务提供专门的虚拟机类型，在虚拟机上预置好 **Stable Diffusion WebUI** 环境和相应的模型文件，便于用户开机即用。用户在使用体验上与在本地没有区别，仍然会受限于所使用的虚拟机上的 GPU 资源。当设计任务较多时，仍然会遇到任务队列的瓶颈。由于应用的前端和后端都全部位于同一台虚拟机上，算力依旧依赖于前端所在的这台虚拟机，所以在应用架构上，与在本地自购主机没有本质区别。

(2) 后端云算力。面对大规模生成式 AI 任务需求，往往采用更高效灵活的前后端分离方式。**Stable Diffusion WebUI** 作为应用前端，被部署在低配置的云虚拟机或本地主机上；当需要推理或训练时，调用后端独立的高性能虚拟机。这样，针对训练和推理，分别配置为不同场景专门优化过的机型；针对不同的生成工作量，配置不同数量的算力资源，以实现并发加速。在亚马逊云科技的平台上，还可以使用 **Spot** 实例，充分利用闲置的计算资源，获得更高的性价比。

2.5.3.3 亚马逊云科技的算力与 AI 研发资源

Stability AI 公司所使用的 **Amazon SageMaker** 是亚马逊云科技的全托管的一站式机器学习开发平台，面向生成式 AI 的算法和模型开发人员，帮助用户高效实现生成式 AI 基础模型的训练、推理、自定义、微调、部署和管理。通过 **Amazon SageMaker**，用户不仅可以轻松访问包括 **Nvidia GPU**、**Amazon Trainium** 在内的最新的基础设施

资源，利用 **Studio**、**Notebook** 等一系列调试、分析和追踪模型效果的工具，轻松实践机器学习运维 **MLOps**、大规模集群协调和分布式训练，还可以获得专门优化的机器学习框架和库、高性价比的 **Spot** 实例支持、多种推理方式的部署、大语言模型并行化处理等多种特性，最大限度地节省模型开发和应用成本，提升整体生产效率。

SageMaker JumpStart 是 **Amazon SageMaker** 提供的一个资源中心，目前内置 300 多种开源模型和 10 多种预设场景解决方案。通过 **SageMaker JumpStart**，用户能够一键部署或微调包括 **Stable Diffusion** 在内的主流生成式 AI 开源模型，低代码地轻松开发高质量模型，并缩短部署时间。同时，它还支持对大量模型的管理，包括搜索和共享。

在硬件层面，用户需要高性能、低成本且为机器学习专门构建的基础设施。亚马逊云科技不仅提供基于英伟达最新 GPU 芯片（如 H100、A100、A10、T4 等）的虚拟机实例，还提供基于自研 AI 训练 (**Trainium**) 与推理 (**Inferentia**) 芯片、专门针对生成式 AI 应用优化的高性价比虚拟机实例，帮助用户大幅节省生成式 AI 训练和推理的成本。与其他类似的 **Amazon EC2** 实例相比，采用 **Trainium** 训练芯片的实例 **Tm1** 可以节省高达 50% 的训练成本，经过优化后可以在与高达 800Gbps 的第二代 EFA (弹性结构适配器) 网络相连的多个服务器上分发训练任务；采用 **Inferentia2** 推理芯片的实例 **Inf2**，吞吐量提高了 4 倍，延迟降低了 10 倍，支持包括 **Stable Diffusion 2.1** 在内的大量生成式 AI 的基础模型，在图片生成时间、QPS (每秒查询推理响应速度)、服务器推理成本上有较大优势。

2.5.4 扩散模型的控制

基于海量图像数据训练的扩散模型使生成结果具有多样性，向用户展示出惊人的创造力，但其算法中引入随机噪声等环节也决定了生成结果的不稳定性是扩散模型在实际应用中无法回避的问题。AI 图像生成的不稳定性主要体现在生成内容的准确性和质量会随生成条

件等因素有所差别。因此，需要通过提示词、采样器等技术手段控制 AI 生成图像的内容、风格、尺寸等指标。这就如同在传统美术设计中，艺术家通过画笔形状、落笔轻重、颜料水分浓稠等准确地表达自己的创作意图和各种细节。

为了稳定控制 Stable Diffusion 模型的生成结果，主要有以下 3 种思路。

(1) 植入概念 (concept implantation)。当原有基础模型中不存在用户想表达的某事物时，可以通过模型微调技术，将概念与对应的形象植入到基础模型中。不论采用 Dreambooth³⁹ 技术还是 LoRA⁴⁰ 技术，植入概念的方法是提供一批与新事物相关的图片，给这批图片赋予独特的标识，并将标识归到一个已存在的主题 (subject) 或类 (class) 中。通过向基础模型植入概念，不仅可以控制生成所需的事物主体 (人物或物体) 和风格，还能控制细节内容、光影效果、姿态等。

(2) 注入条件 (condition injection)。当不便于用概念描述控制信息 (如动作、位置等) 时，可以对扩散模型的神经网络结构进行调整，将控制信息注入到生成过程中，即通过新增可控的神经网络结构 (例如线稿) 来约束扩散过程，从而实现准确的表达。目前常用的这类控制技术有 ControlNet⁴¹ 和 Gligen⁴² 等，利用它们就能够以线稿/边缘图 (如 Canny 边缘、HED 边界、用户草图等)、深度图、人体姿势图 (OpenPose)、物体边界框等作为输入控制条件，引导图片生成。

³⁹ Ruiz, Nataniel, et al. "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023

⁴⁰ <https://huggingface.co/blog/zh/lora>

⁴¹ Zhang, Lvmin, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." arXiv preprint arXiv:2302.05543 (2023).

⁴² Li, Yuheng, et al. "Gligen: Open-set grounded text-to-image generation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

(3) 改进模型 (model improvement)。如果要从根本上提升模型的性能，那么就需要考虑改进模型的设计。一个完整的扩散模型包括前向过程、反向过程、采样过程等很多组件，它们都会影响生成稳定性⁴³。最新发布的 Stable Diffusion XL (SDXL) 1.0 模型进行了全方位的优化⁴⁴，比 Stable Diffusion 1.5 和 2.1 版本能提供更加优异的性能。例如，在 SDXL 模型训练阶段引入输入图像尺寸参数，使模型在训练过程中能够学习到图像的原始分辨率信息，从而在推理生成阶段更好地适应不同尺寸的图像生成，降低了模型对于输入图像分辨率的敏感性，提高了图像缩放条件下生成质量的稳定性。类似的，SDXL 训练还采取了图像裁剪参数条件化和多尺度训练策略，也有助于提高生成质量的稳定性。

稳定控制 AI 图像的生成结果是一项复杂的任务，需要在模型设计、训练策略、数据处理和评估方法等方面进行综合考虑。在实际应用中，对于某些特定的任务，往往还需要通过人工干预，对生成结果进行筛选、标记和调整，通过评估与反馈的循环确保生成质量和可靠性。

2.5.5 通过解决方案加速生成式 AI 技术落地

解决方案 (solution) 是指针对特定问题或用户需求提供的完整的、可操作的解决方法，通常需要整合基于技术、软件、硬件等多方面的资源，是工程化成果的具体体现。在云计算背景下，解决方案是云计算服务商托管服务、合作伙伴产品，以及开源社区项目三者的有机组合，旨在满足客户或用户的具体需求，并解决特定的业务或技术

⁴³ Podell, Dustin, et al. "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis." arXiv preprint arXiv:2307.01952 (2023).

⁴⁴ Chang, Ziyi, George A. Koulieris, and Hubert PH Shum. "On the Design Fundamentals of Diffusion Models: A Survey." arXiv preprint arXiv:2306.04542 (2023).

挑战。在生成式 AI 应用领域，很多软件开发者、信息技术服务商、运营商等个人或企业，已经提供了多种多样的解决方案，将一系列最新的最佳实践和技术资产整合到一起，使用户能更高效、稳定、便捷地开展相关工作，减轻搭建生成式 AI 工作环境的负担。

以亚马逊云科技为例，它面向客户与合作伙伴的实际业务需求，针对特定的业务场景，研发可重用的、经验证的、易于部署的解决方案，进一步帮助用户安全、高效、经济地利用云计算资源，在实际业务中便捷地获得 AI 能力。亚马逊云科技解决方案的特点包括：(1) 代码开源，方案本身免费，客户只需要为解决方案运行时所消耗的云资源付费。(2) 支持 CloudFormation 一键部署，使客户可快速完成部署，提升项目效率。(3) 解决方案的架构和代码经过亚马逊云科技的技术专家严格验证，体现优良架构 (Well-Architected) 的设计思想与最佳实践，有效降低成本，确保方案可靠。

2.5.5.1 亚马逊云科技典型的生成式 AI 解决方案

Stable Diffusion WebUI 是目前最流行的 AI 图像生成的应用平台之一，但其单机运行的模式极大地影响了用户的工作效率。为了推动大规模的工程化应用，利用云计算的弹性算力资源成为克服单机运行瓶颈的最佳选择，但 Stable Diffusion WebUI 本身并没有提供接入云端算力资源的功能。

基于对用户需求的深入调研和分析，亚马逊云科技在 2023 年 6 月推出了 Stable Diffusion 插件解决方案，帮助用户将现有基于 Stable Diffusion WebUI 的模型训练、推理和调优等任务负载从本地服务器迁移至 Amazon SageMaker，利用云上弹性资源加速模型迭代，避免单机部署所带来的性能瓶颈。该方案为常用的 WebUI 原生功能/第三方插件提供云上 workflow，包括 txt2img、Img2img、LoRA、ControlNet、Dreambooth (含 LoRA 模型训练)、Image browser 等，能够满足设计人员和算法工程师的日常工作需求。

Stable Diffusion 插件解决方案主要具有以下特点：(1) 安装便捷，开箱即用：使用 **Amazon CloudFormation** 一键部署亚马逊云科技的中间件，并以社区原生 **Stable Diffusion WebUI** 的插件形式一键安装前端插件，客户即可快速利用 **Amazon SageMaker** 资源进行推理、训练和调优。(2) 社区原生，与时俱进：该方案以 **Stable Diffusion WebUI** 的插件形式运行，用户不需要改变现有的 **WebUI** 使用习惯。方案插件和中间件的代码开源，采用非侵入式设计，有助于用户快速跟上社区相关功能的迭代，包括备受欢迎的 **Dreambooth**、**ControlNet** 和 **LoRA** 等插件。(3) 扩展性强，迁移平顺：**WebUI** 界面与后端分离，**WebUI** 可以在任何支持的终端上启动而没有 **GPU** 的限制，原有的训练、推理等任务通过插件所提供的功能便捷地迁移到 **Amazon SageMaker** 上，从而为用户提供弹性计算资源、降低成本、提高灵活性和可扩展性。

采用解决方案的优势，除了获得更加易用、安全的用户体验外，还在于能够得到亚马逊云科技在软件和硬件能力上的全面加持。以超分辨率重建为例，它是生成式 **AI** 的一个重要应用方向，将低分辨率的图像、视频转换为高分辨率的版本，并且自动补充和恢复图像和视频在提升分辨率后的细节内容，改善用户的观看体验，在如广告设计、影视创作、游戏开发等领域有重要的应用价值。

亚马逊云科技的 **AI** 视频超分辨率重建解决方案采用自研的基于深度学习的超分辨率算法，基于海量数据预先训练出 **AI** 模型，普通用户开箱即用，不需要具备 **AI** 技术能力。在工程实践上，该解决方案充分体现了亚马逊云科技多种服务在有机组合后带来的强大性能。一方面，它的架构使用 **Amazon Batch** 等服务分块批量作业，充分利用云上资源的弹性，缩短总处理时间，确保任务高效执行。当视频由 **HD** 转为 **4K** 分辨率时，在云资源充足的条件下，可通过并发加速达到 1:1 的处理速度（即 1 分钟的原始视频只需 1 分钟即可处理完

毕), 全面优于常规处理速度。另一方面, 基于亚马逊云科技自研芯片 **Inferentia** 实现高性能推理, 与基于 CPU 或 GPU 的传统方案相比, 推理成本降低 55%; 同时, 它支持 **Amazon EC2 Spot** 实例, 与按需实例的价格相比, 使用 **Spot** 实例最高可以享受多达 90% 的折扣。

2.5.5.2 探索传统艺术创新的生成式 AI 实践

亚马逊云科技的 **Stable Diffusion WebUI** 插件解决方案和丰富的托管服务为用户高效利用 **Stable Diffusion** 开展生成式 AI 实践拓展了空间。借助这个解决方案, 亚马逊云科技的大中华区解决方案研发中心与中央美术学院的实验艺术与科技艺术学院合作, 利用生成式 AI 为中国传统艺术注入新元素, 在创新交互式艺术装置方面进行了有益的探索。

在亚马逊云科技 2023 中国峰会上, 双方联合推出了“皮影随形”(GenAI Shadow Puppetry) 互动展位。这项工作展示了姿势分析、**txt2img**、**Dreambooth**、**LoRA** 和 **ControlNet** 等技术的新颖集成, 主要包含 4 个步骤: (1) 创建交互式故事情节, (2) 训练皮影风格的 **LoRA** 模型, (3) 采用步骤 (2) 中训练的模型生成角色和相关素材, (4) 构建端到端的交互式体验。

在步骤 (1) 中, 中央美术学院的师生们创作了叙事线索, 编织出源于中国古代神话的故事情节, 并融合进现代的故事形象, 从而在过去与现在的反差碰撞中形成有趣的体验。在步骤 (2) 中, 使用 **Amazon SageMaker** 训练皮影风格模型。先利用 **Grounding DINO** 模型从经典皮影文献中提取不同的图像, 如图 2-15 所示, 并进行概念标注。再将精选的图像集作为 **LoRA** 模型训练的概念注入, 增强其创作能力。在步骤 (3) 中, 生成超越传统皮影题材限制的新角色和素材, 如宇航员、火车和计算机等新颖的元素, 如图 2-16 所示。在步骤 (4) 中, 设计并实现互动体验, 将设计的故事情节与参观流程结合起来。在观看某个皮影动画视频后, 观众可以自由地做一些动作。利用 **OpenPose**

技术，从摄像头拍摄的观众动作视频中提取姿势数据，用于控制由 LoRA 模型生成的图像，生成互动动作的视频，如图 2-17 所示。最后，将互动动作视频与背景视频一起渲染，生成最终的作品，如图 2-18 所示。为了确保生成视频中形象的一致性，并避免闪烁，在多个相邻帧之间进行姿势对齐，将对齐的姿势帧组合成单个图像，作为 ControlNet 的输入。

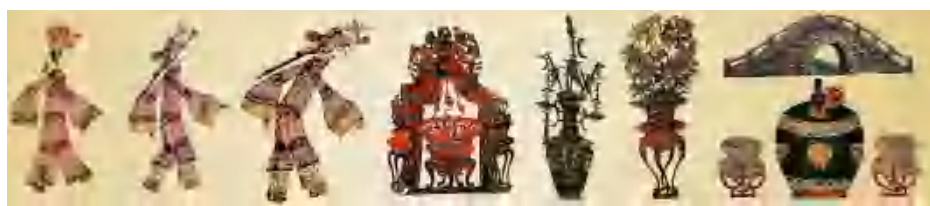


图 2-15 用于训练的原始皮影形象



图 2-16 AI 生成的新皮影形象



图 2-17 通过动作控制生成的皮影形象

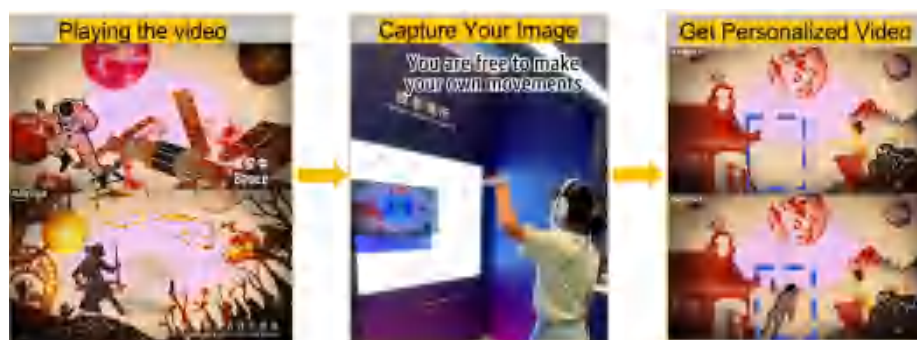


图 2-18 交互体验

图 2-19 所示为“皮影随形”互动展位的架构图，整个系统采用无服务器架构，利用了亚马逊云科技的多项服务。系统前端采用 React 框架，将应用程序拆分成多个组件，每个组件都具有自己的状态和属性，便于单独控制和通信；同时，还使用了视频播放/采集相关技术，通过 React Player 实现视频播放，使用 React Web Cam 采集视频数据。前端的网页及相关资源保存在 Amazon S3 存储桶中，并通过 Amazon CloudFront 服务分发。对于系统后端，首先在客户端侧完成视频的抽帧与动作分析，并将结果通过 API Gateway 服务发送到云端的 Amazon Lambda 服务处理。然后，Lambda 函数将每个请求保存到非关系型数据库 Amazon DynamoDB 中，并触发 Amazon SageMaker 进行推理。最后，Amazon SageMaker 根据用户的动作通过预训练模型生成皮影风格的图像集，生成结果被合成为视频后，保存在 Amazon S3 存储桶中。当观众扫描二维码请求访问视频时，Amazon Lambda 服务根据二维码中的 ID，从 Amazon S3 中获得生成视频的信息，由 Lambda 函数生成临时访问链接，并返回给观众。

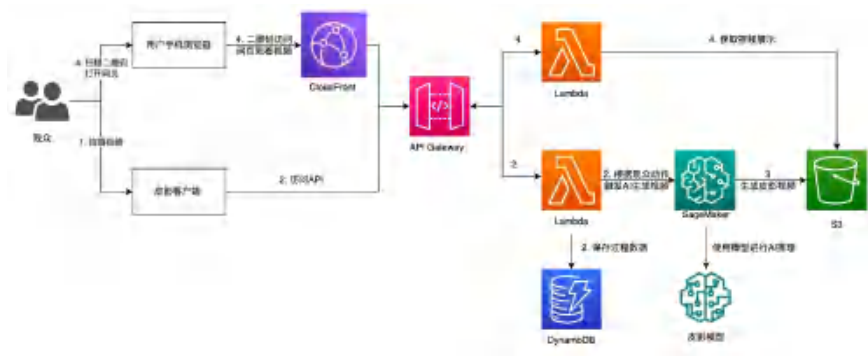


图 2-19 “皮影随形” 互动展位架构图

不断发展的生成式 AI 技术为该项目留下了更多的改进与创新空间。在创作过程上，根据前期中央美术学院师生的使用体验反馈，可以通过技术手段优化创作工具，进一步降低艺术工作者的工作量，帮助他们将精力集中在构思故事情节和微调细节上。例如：更加完整的单张图的生成技术，其中同时包括人物与背景；端到端直接生成视频，减少工作流程中的人工操作；等等。在互动形式上，可以融入更多个性化元素，不止是观众的动作，还可以根据观众的面容、表情、服饰、性别等信息生成个性化内容。同时，还可以进一步提供手机端的体验。

2.5.6 展望：从工程化到工业化

随着生成式 AI 技术的工程化手段的成熟，我们将迎来生成式 AI 的工业化应用阶段。在 AI 图像生成领域，工业化具体是指将生成式 AI 技术应用于图像生成过程中的规模化、标准化和商业化，从而将工程化过程中积累的最佳实践与资源复制放大，进一步降低生成式 AI 在应用层面的复杂度和成本。

生成式 AI 图像生成的工业化主要具有以下特点：

(1) 大规模生成图像：利用生成式 AI 技术，极少或无需人工干预，快速生成大量的图像和视频，以满足工业化生产和商业需求。这意味着生产者应当有灵活弹性的算力获取方式，以确保高可用性和低使用成本，同时有完善的数据与媒体资源管理系统和机制，以确保实

现生成的数字资产的价值。

(2) 自动化流程：自动化是大规模生产的基础，它将贯穿从数据准备、模型训练、模型发布，到通过简单的输入触发模型，自动生成图像的全过程，无需手动干预。云计算环境中丰富的托管服务、无服务器架构、事件总线，确保用户能够快速构建起高可用、弹性的、高性价比的自动化流程。

(3) 标准化的模型和流程：工业化要求使用标准化的模型和流程，以确保生成结果的稳定性和一致性。这包括使用经过训练和优化的生成式 AI 模型，并建立统一的数据处理和后处理流程。

(4) 商业价值的呈现：生成式 AI 图像生成的工业化不仅仅是一种技术探索，更注重其商业价值。这包括将生成的图像更广泛地被应用于广告、媒体、设计、创意产业等领域，以实现商业利益和创收。

从使用者的视角，我们会看到以下变化：

(1) 重新定义一些行业的生产流程。生成式 AI 技术是一种先进的生产力，它将改变原有的生产者角色，催生出新的生产流程，对生产者提出新的职业要求。图 5 所示是一种以生成式 AI 技术为主线的图像生产流程构想。在这种生产线中，美术总监不仅具有极强的审美设计能力，还具有一定的技术能力，熟练使用 ControlNet 等工具。他通过提示词和 ControlNet 模型，掌控项目的美术风格，由后台批量生成图片。设计人员从大量图片中进行挑选，利用生成式 AI 工具精修，并将修改意见反馈给美术总监；美术总监再根据反馈修改参数，调整批量生图的效果。

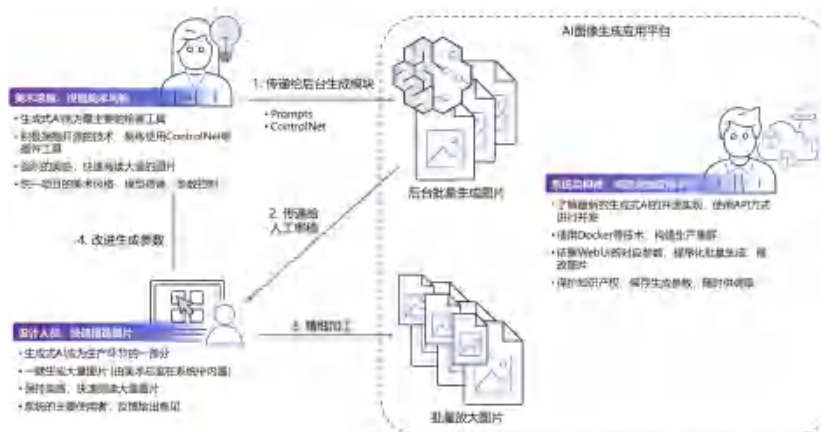


图 2-20 新生产流程示意图

(2) 应用界面的进一步简化与定制化。简单的应用界面有助于降低使用门槛，让更多的普通用户和设计师使用 AI 图像生成技术。极简的应用界面也正是 Midjourney 迅速走红、吸引大量用户尝试的重要因素。随着新生产流程的形成，以及 AI 图像生成经验的积累，不同行业的企业类用户会针对自己的业务流程与需求定制应用界面，隐藏复杂的配置参数，通过 API 技术调用 Stable Diffusion 等模型和相关生成式 AI 的处理操作，给普通用户提供简洁的应用界面，进一步提高工作效率。

(3) 模型应用方式更加便捷和多样。在当前的应用场景中，用户一般通过 Huggingface、Civitai 等网站下载模型后，在自己的环境中进行微调 and 构建应用。在下一阶段，用户可以通过 API 直接调用模型，无须自己预置模型推理和训练的资源，从而进一步降低构建应用的门槛。Amazon Bedrock 是亚马逊云科技针对生成式 AI 领域推出的最新的服务，是使用基础模型 (FM) 构建和扩展生成式 AI 应用程序的最简单方法，为所有开发者降低使用门槛。它在 2023 世界人工智能大会 (WAIC 2023) 展品评选中入选 2023 九大“镇馆之宝”之一。通过 Bedrock 服务，客户可以通过 API 访问来自 AI21 Labs、Anthropic、Stability AI 和 Amazon Titan 的基础模型，例如 Titan Text 、Jurassic-2、

Claude、Stable Diffusion XL 等。借助 Bedrock 的无服务器体验，用户可以结合自身应用场景轻松找到适合的基础模型，快速上手；在确保数据安全和隐私保护的前提下，使用自有数据基于基础模型进行微调和定制，使得模型效果更加符合自身的业务场景；并使用他们已经熟悉的亚马逊云科技工具和能力，将定制化模型集成并部署到他们的应用程序中，同时不需要管理任何基础设施。

生成式 AI 的工业化将在更大范围更深远地影响我们的生产和生活，也给我们带来了数据管理、数据隐私、AI 伦理、版权保护等诸多新问题和新的挑战。我们需要持续关注和思考生成式 AI 技术的进步和优化，以确保 AI 生成的图像能够被安全、合规、可持续地用于商业应用，并产生积极的社会价值。

2.6 人工智能作为创作的课题

人工智能透过大众媒介的传播已成为家喻户晓的高频词汇，艺术界乘着跨界之风亦对其趋之若鹜且多有尝试。

艺术创作也可以依据人工智能领域所研究的课题展开：

一、其中最为显著的便是在构建通识人工智能系统过程中展开的子问题——人工智能创造力研究。这类研究包含了文本、音乐、图像的一种或者多种组合生成。其中，前沿的大参数人工智能系统已经可以完成较为复杂的平面设计任务。这些生成任务的研究可以依据前文所述的思路融入到艺术创作中。

二、人工智能被大量的用于发现我们所在世界的隐藏规律，其中有一门显学是通过利用人工智能识别图像来理解世界。这些研究可以用于将名片上的信息转换为文字，识别快递及信件上的邮编和地址，基于人脸识别的身份核验，以及基于摄像头的自动驾驶。这些识别类任务可以帮助艺术创作者收集、分类创作素材，或是基于小型自动驾驶平台进行装置类的创作。

三、人工智能被用于驱动单个或者多个对象实现与外界的互动。这种互动可以是形式较为简单的棋类游戏，构成复杂信息不对等的策略类游戏，亦或是与物理空间直接发生交互的单体或者集群机器人。这类以控制对象为目的的人工智能系统可以帮助艺术家构建虚拟的生态系统，或者加强作品元素之间以及与观众的互动。

2.7 人工智能艺术案例研究

人工智能已经走进了艺术领域，为创造和鉴赏艺术带来了一场前所未有的革命。从绘画到音乐，从文学到影像，人工智能技术正逐渐融入艺术创作的方方面面，为艺术家和观众呈现出多种可能性。这个令人着迷的领域不仅激发了创作者的想象力，还引发了关于创意、原创性和技术伦理等问题的深刻讨论。接下来，我们将通过近年来的一些人工智能艺术案例来探索这个新兴领域，揭示正在发生的文化和技术革命。

2.7.1 AI 生成艺术案例研究

生成艺术是指使用计算机算法或人工智能技术创作的艺术作品。这些算法和技术能够模仿人类艺术家的风格和创作方式，从而产生独特的艺术作品。

生成艺术的核心是计算机程序通过学习艺术家的作品、图片、声音或文本等大量样本数据，来掌握并模仿创作者的创作风格。最常用的技术之一是“深度学习”，特别是基于神经网络的生成模型。这些模型在学习了大量数据后，能够生成具有相似特征的新作品。这一类作品主要包括图像、声音和文字生成类艺术作品。其中图像生成是探索最为广泛的领域，它专注于使用计算机算法和人工智能技术来生成艺术性的图像作品。在此类作品中，计算机程序通过学习大量的图像数据，例如绘画、摄影、图形等元素来创造出新的艺术图像。

在作品《Fifty Sisters（五十姐妹）》中，艺术家 Jon McCormack

将石油公司原始的图形元素作为基础数据集，利用人工智能技术不断学习植物形态，同时训练它具有一定独特的视觉语言，通过不断调优，从代码中生长出来的植物千姿百态。在这种创作方法背后，艺术家的身份也发生了一些转变。在传统绘画中，艺术家是作品的创作者，从构思到执行都是由艺术家亲自完成。他们运用自己的技巧、创意和感觉，通过手工绘制创作出艺术作品。然而在 AI 生成绘画中，艺术家的角色更像是指导者或训练者。虽然他们仍然需要设计模型、选择算法和训练数据，但最终的图像是由计算机程序生成的，一部分创作过程被移交给了人工智能。在 Jon McCormack 的另一件作品《The Unknowable（不可知）》中展示了五种来自悉尼地区的本土植物，这些形态是通过模仿生物进化和发展的定制计算机程序生成而来。视频中这些植物在无尽的解体与重组中不断循环。AI 生成图形为艺术家提供了一个全新的探索领域。艺术家可以利用人工智能技术创作出以前难以想象的图像风格，实验不同的表现形式和创作方式，拓展艺术的边界。

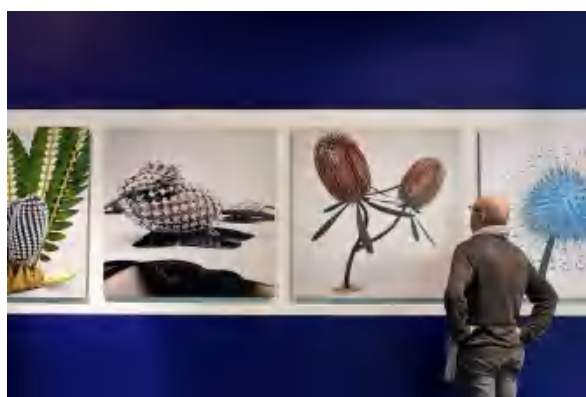




图 2-21、图 2-22 Fifty Sisters（五十姐妹）.Jon McCormack. 2012⁴⁵



图 2-23、图 2-24 The Unknowable（不可知）.Jon McCormack.2017⁴⁶

⁴⁵ <https://jonmccormack.info/project/fifty-sisters>

⁴⁶ <https://jonmccormack.info/project/the-unknowable>

James Bridle⁴⁷ 《The Cloud Index（云指数）》项目(2016)是一个与策展人、艺术家 Ben Vickers 合作的项目，Vickers 表示：“‘云指数’是一款软件，可以用来根据不同的政治结果创建不同的天气形态。”为了开发此项目，艺术家使用展示英国天气状况的卫星图像和显示英国与欧洲关系的英国脱欧投票结果来训练神经网络。云指数是通过分析 15000 多张天气模式的卫星图像，并将其与英国 2016 年欧盟公投前七年的民意调查数据相关联。这种分析是通过机器学习进行的，这是一种模拟大脑结构的神经网络，用以读取大量数据。虽然最初的提议有点荒诞，但 Bridle 巧妙地将人工智能、数据预测和天气结合在一起。他的作品涉及科技和社会问题的交汇，他通过艺术来引发观众对重要问题的思考，从而更深入地探索和呈现数字化时代的复杂性。



⁴⁷ James Bridle 是一位擅长将艺术、技术和科学融合，以人工智能的方式揭露和分析当前的政治环境与问题的艺术家。

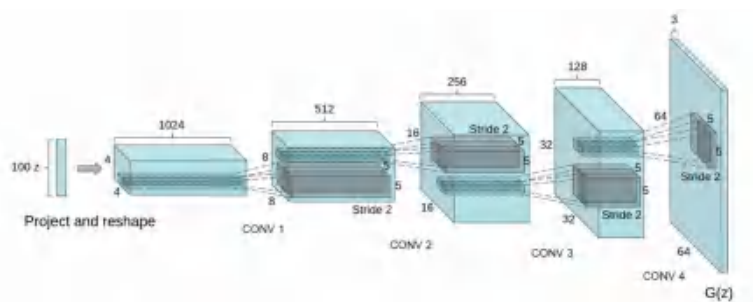


图 2-25、图 2-26、图 2-27 The Cloud Index(云指数).James Bridle.2016⁴⁸

《Shiv Integer(整数刀)》是艺术家 Matthew Plummer Fernandez 与 Julien Deswaef 合作完成的作品。它是一个能够创造集合艺术的 3D 打印机器人。Thingiverse 是最大的在线 3D 打印社区，拥有庞大的用户制作模型库。Shiv Integer 作为定制软件会随机在 Thingiverse 上挑选模型，并将其组合成雕塑，并给它们起一堆由语无伦次的词构成的名字，比如“电子果汁高尔夫上的圆盘”。这个过程继承了达达主义的现成品艺术和偶发艺术传统，探讨了互联网文化中的知识共享、版权等等问题，同时也在 Thingiverse 上引发了关于它是艺术还是垃圾的长期争论。

⁴⁸ <http://cloudindx.jamesbridle.com/>

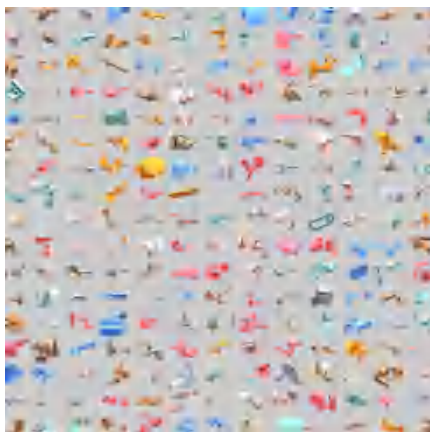
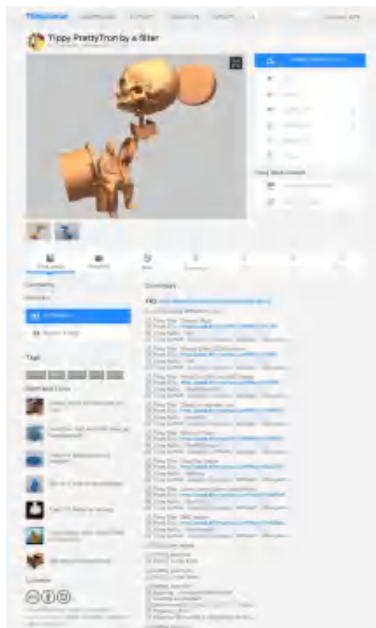


图 2-28、图 2-29、图 2-30 Shiv Integer（整数刀）.Matthew Plummer
Fernandez . 2016⁴⁹

⁴⁹ .<https://www.plummerfernandez.com/shiv-integer/>

《Fall of the House of Usher II（厄舍古厦的倒塌II）》是一个手工制作的数据集，包含两百多幅单独的水墨画，该作品基于1929年的无声电影《厄舍古厦的倒塌》而创作。反馈回路系统使用经过特殊修改的算法处理和复制图纸，然后使用生成对抗网络生成的镜头来构成下一个训练集的基础。重复、回忆和娱乐的概念是“古厦的倒塌”系列的核心。这部作品是继《厄舍古厦的倒塌》（2017）之后的系列作品中的第二部，该部分利用该数据集创建了12分钟的动画。

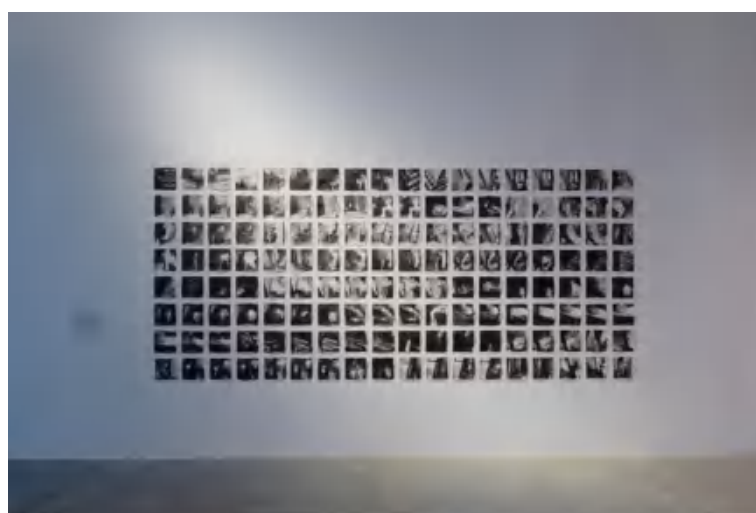


图 2-31 Fall of the House of Usher II（厄舍古厦的倒塌 II）.Anna Ridler.2017⁵⁰

《人工智能无限电影》是艺术家与人工智能科学家合作开发的一个没有电影人（如导演、编剧、摄影师或演员等）参与的实时电影生成系统。观众可根据需求，输入电影类型（如科幻、犯罪、爱情等），再通过输入关键词或句子改变电影的叙事情节或风格，制作由AI出品的永不重复的电影。此项目试图追求自然人创造不出来，又是自然人观影需要的传统电影里没有的成分。其理念的依据基于自然人的缺陷，例如：由于情感狭隘对问题判断的成见；政治、经济利益驱使导致的贪婪与邪念；知识量和视野所限。AI站在冷静的立场，通过人

⁵⁰ <http://annaridler.com/fall-of-the-house-of-usher-ii>

类各种影像素材的读取，提示人类整体生活观念的“内部结构”。这是对未来电影一种可能性的试验。



图 2-32 人工智能无限电影 (AI-IF).徐冰. 2017 年至今⁵¹

AI 生成人脸一直备受关注，它引发人们对身份和隐私问题的思考的思考，作品《Portraits of Imaginary People (虚构人物肖像)》通过使用 Flickr 上成千上万张人脸照片，输入到名为“生成对抗网络”(GAN) 的机器学习程序中，来探索人脸的潜在空间。GAN 的工作原理是利用两个神经网络进行对抗游戏：一个是“生成器”，旨在生成逐渐更具说服力的输出；另一个是“判别器”，旨在学习区分真实照片和人工生成的照片。最初，这两个网络在各自任务上都表现不佳。然而，随着“判别器”网络开始学会预测真假照片，它不断提醒“生成器”，迫使后者生成更难以区分、更具说服力的示例。为了跟上“生成器”的进展，生成器也不断改进，而“判别器”也必须相应地改进其识别能力。随着时间的推移，生成的图像变得越来越逼真，因为双方都竭力超越对方。因此观众看到的图像是神经网络从训练图像中学

⁵¹ <https://www.xubing.com/cn/work/details/690?type=project#690>

到的规则和内部关联的结果。作品以虚构人物肖像为中心，引发人们对真实性的关注。



图 2-33 《Portraits of Imaginary People（虚构人物肖像）》Mike Tyka，2017⁵²

《Edmond de Belamy》是一幅由位于巴黎的艺术团体 Obvious 于 2018 年使用生成式对抗网络 (GAN) 构建的肖像画。该作品是以画布为媒介的，属于一系列名为《La Famille de Belamy》的生成图像。Belamy 这一名称是为了向生成式对抗网络的发明者 Ian Goodfellow 致敬⁵³。在法语中，“bel ami”意为“好朋友”，因此它是 Goodfellow 这个名字的双关语翻译。该作品作为首幅使用人工智能创作的艺术品在佳士得拍卖行进行拍卖后，实际成交价高达 432500 美元，广受关注。

⁵² <https://miketyka.com/?p=A%20fleeting%20memory>

⁵³ [theverge.com. How three French students used borrowed code to put the first AI portrait in Christie's\[EB/OL\].https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robby-barrat-gans.](https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robby-barrat-gans)



**图 2-34 《Edmond De Belamy》；材质：使用 GAN 算法生成，喷墨打印于画布；
尺寸：70x70 厘米，2018**

在 AI 艺术中，我们经常看到艺术家使用类似的代码，如 CycleGAN、SNGAN、Pix2Pix 等，并使用从网络上抓取的相似数据集进行作品训练，这导致了图像的同质化问题，同时会让 AI 艺术显得重复和庸俗。为了应对这样的情况艺术家可以选择走在他人之前，永远都使用最先进的技术媒介进行创作，但新技术的使用并不能成为决定艺术作品“好”与“坏”的持久因素，艺术家应更专注于用自己的创作素材来训练 AI 模型，以确保作品的独特性，例如艺术家 Helena Sarin 使用自己训练的 GANs 来创作数字艺术，她可以在短时间内生成大量的作品草稿，并从中挑选出最具艺术价值的作品进行展示。

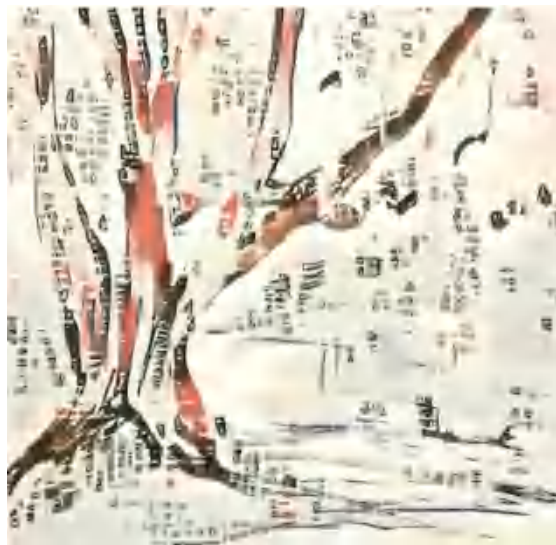


图 2-35 “四月赏花的游客大军 哦，他们是麻雀人”

bashoGAN 是用树木图像和俳句的书训练的模型

（“troops of tourists come for april flower-viewing oh, they're sparrow-men”），

Helena Sarin, 2018⁵⁴

《Compressed Cinema（压缩电影）》是一个包含五个新的视听作品的系列，于 2020 年完成。这个系列的视频图像由 Casey Reas 创作，每个作品都有由 Jan St. Werner 创作的立体声音轨。这些作品的理想展览空间是五个隔音的黑箱房间，地板设有舒适的座椅。视听作品将以墙面尺寸进行投影。声音通过高品质立体声扬声器系统回放。观众可以在房间自由走动，并可以花费任意时间于一个作品，每个循环体验都有所不同。这些视频可以在影院中从头到尾播放，也可以同时播放任意数量的视频，并以任意顺序播放，它们也可以作为电影/视频节目的一部分来观看，视频没有标题或片尾字幕，它们按照艺术家的意愿开始和结束。这些视频的图像是由 Casey Reas 使用生成对抗网络（GANs）创作的一组“电影截图”派生而来。这个过程在 Reas

⁵⁴ <https://twitter.com/hashtag/bashoGAN?src=hash>

的 2020 年著作《用生成对抗网络创造图片》（Making Pictures With Generative Adversarial Networks）中有详细记录。



图 2-36、图 2-37 COMPRESSED CINEMA（压缩电影）.Casey Reas.2018-至今
(精选)⁵⁵

Mario Klingemann 的《Memories of Passersby I（路人记忆 I）》是一个隐藏在古董家具中的计算机系统，家具看起来就像是中世纪的现代橱柜和老式收音机的混合体，作品利用神经网络实时地呈现不断变化的肖像，相比于由算法生成的单一图像，作品更关注于创造图像的计算机代码，艺术家通过改变、删除或交换训练数据的权重来操纵原有模型，这种引入的故障会导致模型误读输入数据，而呈现非常规视觉效果。作品向我们展示了机器如何在海量数据中“记忆”和“想象”人类面容。

⁵⁵ <https://reas.com/>



图 2-38 Mario Klingemann 的《Memories of Passersby I (路人记忆 I)》2018⁵⁶

在 Shinseungback 和 Kimyonghun 的项目《Nonfacial Portrait, No.05 (非面部肖像之五)》中他们邀请艺术家为他人画面部肖像。画家们必须遵循一条规则：“肖像的脸不能被人工智能检测到。”在制作肖像时，三种人脸检测算法的相机会扫描画布，如果检测到任何人脸，显示器会通知艺术家。艺术家会参考它并修改这幅画。如果一个人的作品很“像”一幅肖像画，人工智能也将很容易检测到脸部。人们越是试图让它无法被人工智能识别，这幅画就越不会被视为人的肖像。人们必须找到只有人类才能感知的狭小视觉空间。在装置中，每幅肖像都与画家绘画过程的视频一起呈现。

⁵⁶<https://underdestruction.com/2018/12/29/memories-of-passersby-i/>

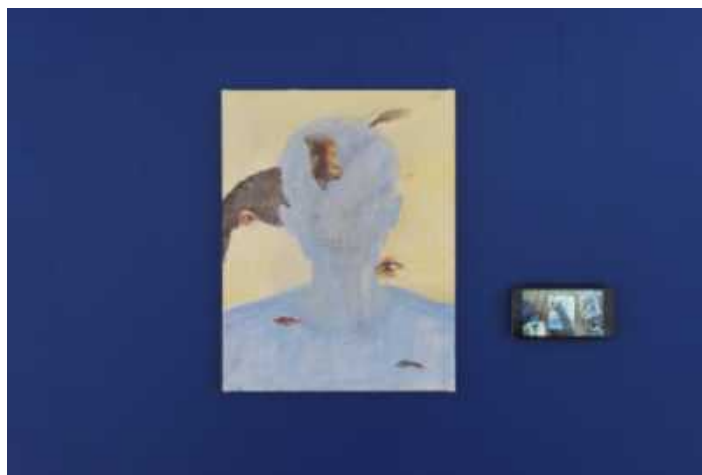


图 2-39 Nonfacial Portrait, No.05（非面部肖像之五）.Shinseungback
Kimyonghun.2018⁵⁷

《合成抽象》是一系列抽象的屏幕打印图像，它们始终被主要图像识别系统（例如 Google、Amazon、Tumblr 等）注册为“成人”或“露骨内容”。该系列探讨了机器如何感知露骨内容，是 Tom White 正在进行的关于机器如何观看、理解图像的研究。

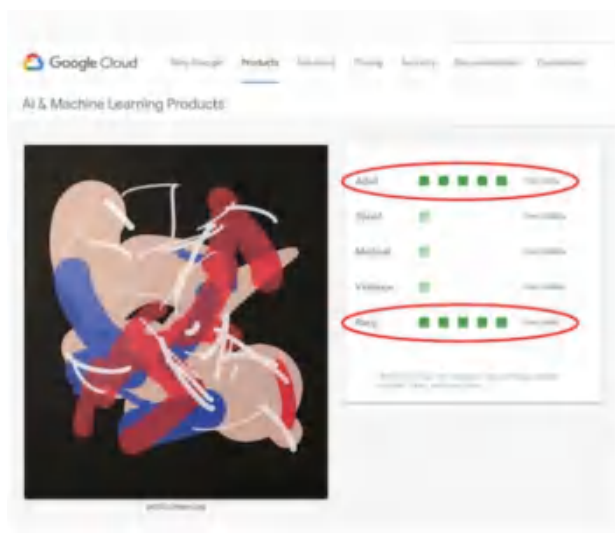


图 2-40 Synthetic Abstractions（合成抽象）.Tom White.2018⁵⁸

⁵⁷ <http://ssbkyh.com/>

⁵⁸ <https://aiartists.org/tom-white>

White 称自己的版画是为了通过机器的眼睛看世界,并为机器“说话”发声。这种声音实际上是一系列算法,即“感知引擎”。“感知引擎”获取机器视觉算法训练的数据——包含数千张物体图片的数据库——并将其提炼成抽象形状。然后将这些形状反馈到相同的算法中,以查看它们是否被识别。如果没有,图像会反复矫正,直到它被发送回去。这是基于图像生成技术设计的一个试验过程,本质上是算法理解世界的逆向工程。

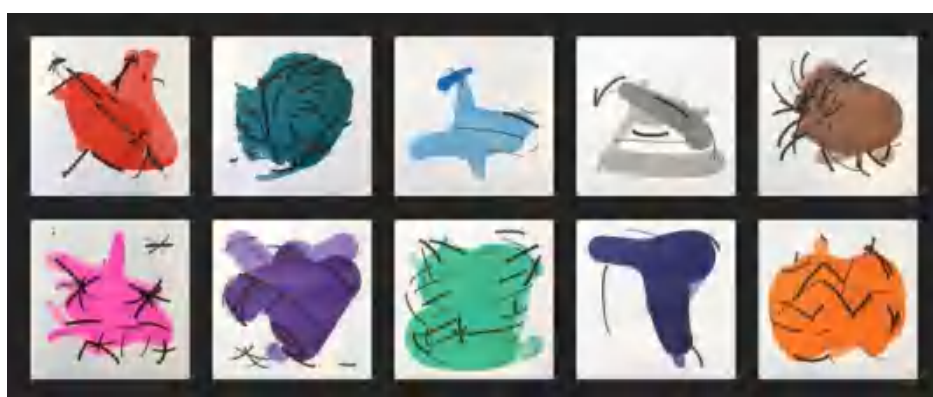


图 2-41 Perception Engines (感知引擎).Tom White.2018⁵⁹

《nimiia cétii(火星萨满)》是艺术家 Jenna Sutela 与 Somerset House Studios 的常驻艺术家 Memo Akten, 以及谷歌艺术与文化实验室的创新负责人 Damien Henry 合作共同制作的一个新的视频装置。作品的灵感来自于物种间交流的实验,并渴望与我们意识之外的世界建立联系。这项工作记录了神经网络、早期火星语言的录音和太空细菌运动的镜头之间的相互作用,并使用机器学习生成了一种新的交流形式。

⁵⁹ <https://aiartists.org/tom-white>

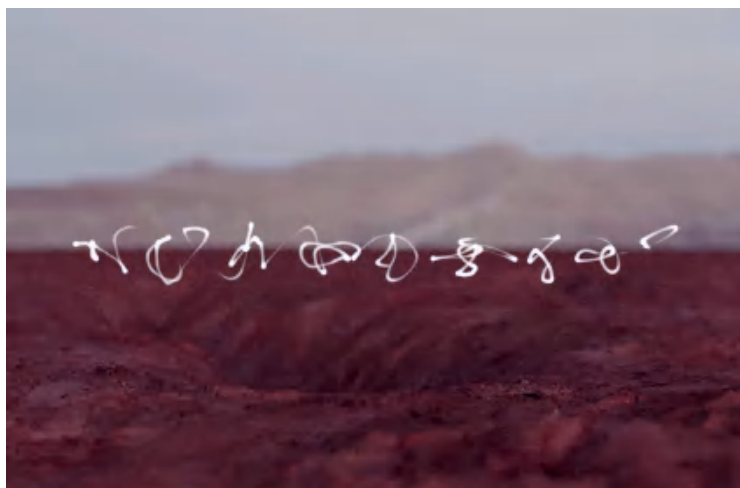


图 2-42 nimiia cétii（火星萨满）.Jenna Sutela. 2018⁶⁰

“Alt-C”是一种利用植物产生的电力为开采加密货币的单板计算机供电的装置。在光合作用过程中，绿色植物会释放土壤中的糖和有机物。这些营养物质随后会被释放电子的细菌消化。当电子在两个电极之间流动时，电流被用来为加密货币采矿设备供电。随着货币的生长，有可能得出大气条件与加密货币生产之间的相关性。Michael Sedbon 已经训练了一个人工智能神经网络来预测与英国天气预报相关的加密货币挖矿率。人工智能将试图制定一项战略，将种植、收获的资本再投资于可耕种的土地。



⁶⁰ <https://www.somersethouse.org.uk/whats-on/jenna-sutela-nimiia-cetii>

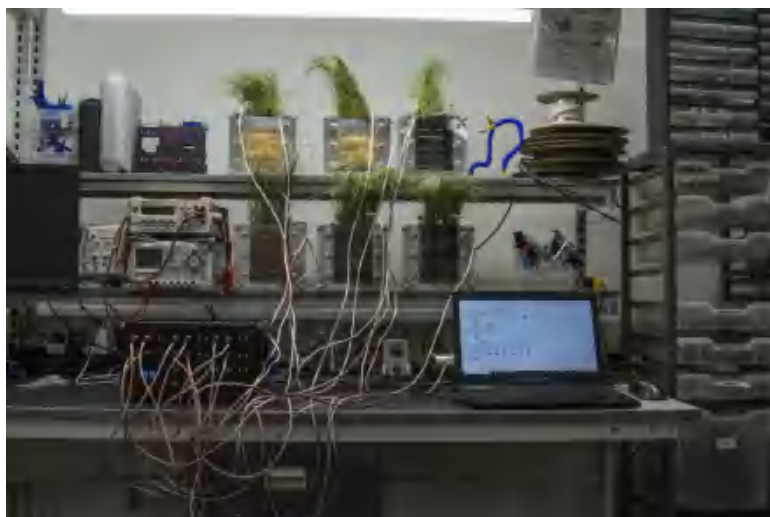


图 2-43、图 2-44 《ALT-C》，Michael Sedbon. 2019⁶¹

《Making Pictures With Generative Adversarial Networks(用生成对抗网络创造图片)》是艺术家 Casey Reas 首次介绍新兴人工智能技术的非技术性著作,艺术家在书中探讨了使用生成对抗网络(GANs),特别是深度卷积生成对抗网络(DCGANs)制作图片的感受。此书是为对人工智能技术的创造性应用感兴趣的读者编写的入门读物。希望读者可以探索这一新兴技术,并根据自己的理解对其进行改编。艺术家鼓励对机器智能时代的艺术进行严谨的讨论,希望激发更多的研究与合作。

⁶¹ <https://michaelsedbon.com/Alt-C>



图 2-45 Making Pictures With Generative Adversarial Networks（用生成对抗网络创造图片）.Casey Reas.2019⁶²

作品《Western Flag（西部国旗）》描绘了 1901 年在德克萨斯州斯平德尔托普（Spindletop）地区发现的世界第一个重要石油开发地点“卢卡斯喷油井（Lucas Gusher）”，现在这个地点已经荒芜和枯竭，艺术家将其重新制作成数字模拟作品，作品中央有一根竖立的旗杆不断喷出黑烟。计算机生成的斯平德尔托普（Spindletop）在整个年份内与德克萨斯州的实际地点完全相同，太阳在适当的时间升起，白天根据季节变得长短不一。这个模拟程序由计算机程序实时计算每帧动画所需的内容。《西部国旗》位于科切拉谷和棕榈泉市的入口处，是对人类过度开发和耗尽资源的鲜明提醒。数百万年前，这片海床曾经生机勃勃，而现在却被迫变成了荒地，在这里升起人类“自我毁灭”的旗帜，提醒人们去思考在全球变暖和曾经肥沃土地变成沙

⁶² <https://www.anteism.com/casey-reas-resources>

漠的过程中人类所扮演的角色。作品中艺术家借用数字技术打破空间与时间的界限，在现实世界中创造出独特的视觉效果。作品不仅打破了传统艺术品的静态边界，而且为观众提供了更加身临其境的艺术体验。

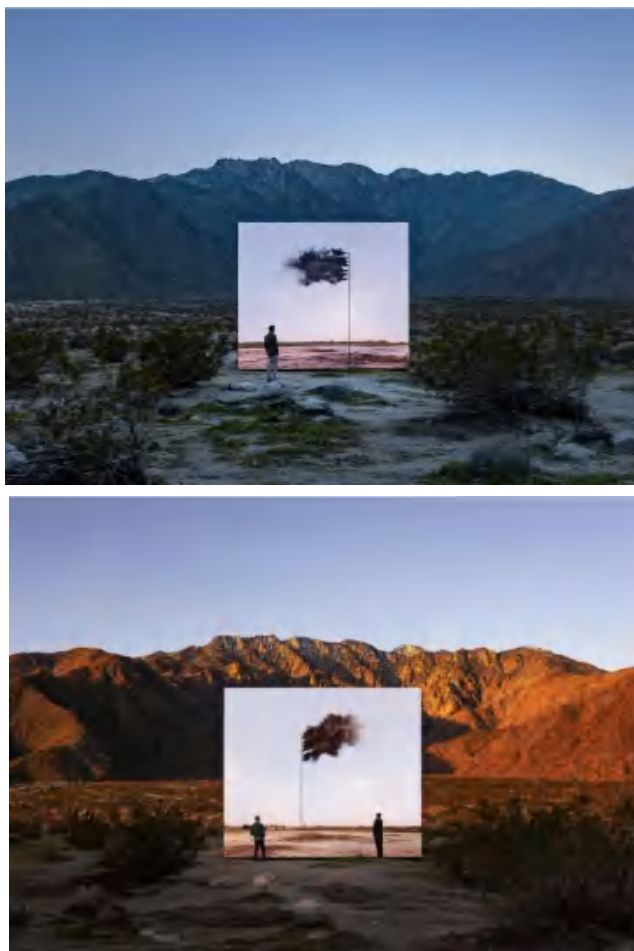


图 2-46、图 2-47 Western Flag（西部国旗）.John Gerrard. 2019⁶³

《一个有科技的女人》的第三部分是一个三屏视频(18分25秒): 基于行为过程中出现频率最高的关键词,在第一部分的记录文档中筛选出所有提到这个关键词的部分,然后拚贴剪辑而成。在这个过程中,

⁶³<http://www.johngerrard.net/western-flag-spindletop-texas-2017.html#2019-western-flag-spindletop-texas-20>

艺术家运用了人工智能(AI)技术中的 **image detection** (图像检测) 和 **dense caption** (密集标题) 功能, 对所有视频中的图像进行了解读和抓取, 即通过 AI 去详细描绘视频中出现的所有内容。(此部分与艺术家汪洋合作完成)。



图 2-48 一个有科技的女人.武子杨.2019⁶⁴

艺术家的实践不仅是探索新技术的可能, 同时也会关注相关伦理问题, 团队合作项目《Exposing AI》关注了那些因人脸数据收集而受到损害的人们, 这些数据在全球范围内被用于增强警备和军事应用, 但这些照片的使用并没有得到任何相关人员的知情或同意。在这个艺术项目 **Exposing.ai** 中, 用户可以在网站上寻找自己的人脸数据。这是一项关于隐私、监控、人脸识别的研究项目, 该项目主要调查人们的照片如何在不知不觉中成为全球生物识别行业的信息数据, 同时提出了问题: “我们如何在开放共享的行业大背景下保护用户的生物数据?” 此外 **Exposing.ai** 项目的研究成果曾出现在《金融时报》、《纽约时报》、《自然》杂志、美国政府问责局的报告、2020 年人工智能指数等多个媒体和学术研究论文中, 并且在公共讨论中推动了关于数据集收集伦理问题的讨论。

⁶⁴ <http://www.unart.org.cn/cn/category/article-list/detail!wzy>

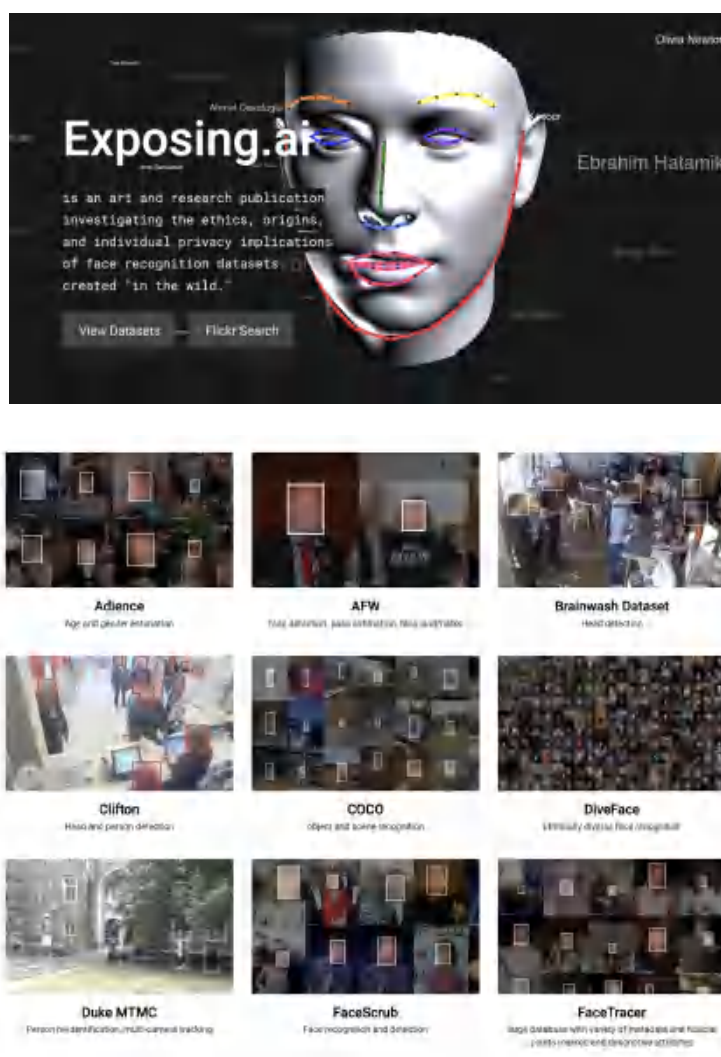


图 2-49、图 2-50 Exposing.ai.Adam Harvey.2019⁶⁵

Deepfake 是用于合成人类面部、语音或视频的技术，已经成为生成艺术的一部分，其通过神经网络学习人的面部特征并在目标视频中重新生成或替换面部，从而创造出令人信服的假视频。

Deepfake 技术可以为艺术创作者提供新的创作手段和方式，但同时也带来了伦理和真实性的挑战。在艺术创作中使用 Deepfake 可以实现之前难以实现的创意，如替换电影角色的面部、重塑历史人物的影像等。但在其他领域，它可能被用于制造虚假新闻或误导公众，

⁶⁵ <https://exposing.ai/>

因此需要谨慎使用。

AI 换脸作品《我们雨中看过你哭泣》是艺术家陈抱阳创作于 2019 年 11 月，也是电影《银翼杀手》的背景设定时间。在电影中饰演 Roy 的 Hauer，面对试图终结他的人类赏金猎人 Rick 时，即兴创作了那段著名的“雨中之泪”独白。在这部作品中，Roy 的脸被老年化的 Harrison Ford 所替代，他在电影中饰演了人类赏金猎人。Ford 面对着他自己年轻时的模样，谈论着人类的共情和他在那场神秘的战争中的经历。艺术家借 Deepfake 这种人工智能技术来叙述人与人工智能之间的联系，“就事论事”是一种创作的路径。

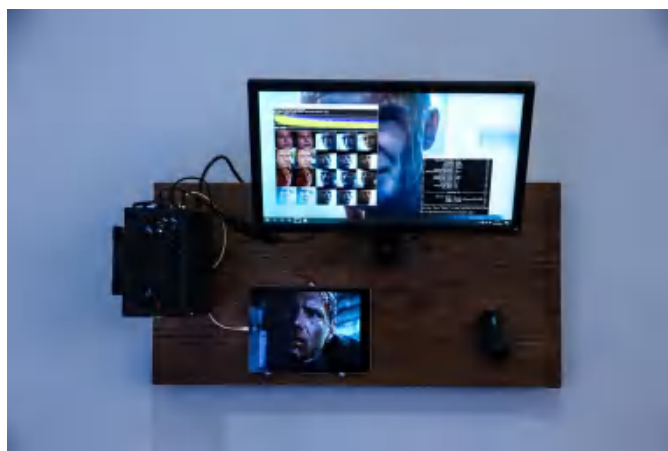
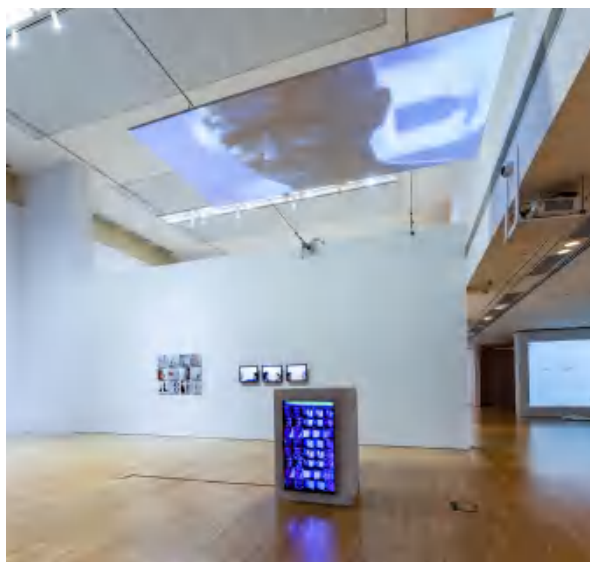


图 2-51 《我们雨中看过你哭泣》.陈抱阳.2019⁶⁶

⁶⁶ baoyangchen.com

作品《Zizi-Queering the Dataset (Zizi-重构数据集)》中艺术家 Jake Elwes 以 StyleGan (2019) 为基础在包含 70,000 张人脸照片的 FFHQ (2018) 数据集上进行了训练。正如艺术家在其网站上所言，面部识别算法和深度伪造技术目前很难辨别变性人、同性恋者和其他边缘化身份人士。虽然神经网络的灵感来源于人脑，但令人遗憾的是，它显示了我们在构建身份时存在的二元对立问题。作品旨在解决面部识别系统经常使用的训练数据缺乏代表性和多样性的问题。制作这段视频的方法是破坏这些系统的过程，艺术家用在网上找到的 1000 张变装和性别不固定的面孔图片对程序进行重新训练。这使得神经网络内部的权重从最初训练的规范性身份偏移，进入了一个涉及性别问题的领域。

作品《Zizi-Queering The Dataset》让我们得以窥探机器学习系统的内部，并直观地了解神经网络已经（和尚未）学到什么。这件作品是对差异和模糊性的赞美，让我们意识到数据驱动社会中的偏见。Zizi 系列作品从 2019 年开始一直持续至今，是艺术家对人工智能与变装表演模糊地带的不断探索。变装表演挑战性别，探索异质性，而人工智能通常作为一个概念和工具被神秘化，并在复制社会偏见方面发挥着不可小觑的作用。Zizi 通过深度伪造，使用机器学习创建的合成变装身份，将这些主题结合在一起。

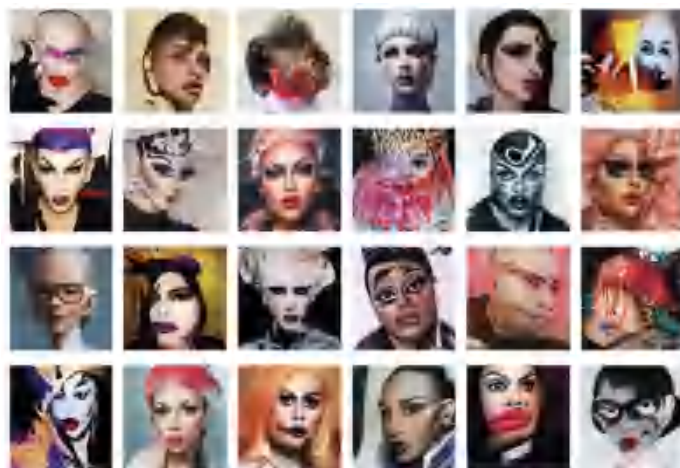


图 2-52 Zizi-Queering the Dataset (Zizi-重构数据集).Jake Elwes.2019⁶⁷



图 2-53 《Zizi-Queering the Dataset (Zizi-重构数据集)》 Jake Elwes 2019
7 屏装置, Onassis Foundation, Pedion tou Areos Park, 雅典, 希腊⁶⁸

Sofia Crespo 是一位对生物学相关技术有着浓厚兴趣的艺术家。她的主要关注点之一是有有机生命使用人工机制来模拟自身生成和进化的方式, 这意味着技术并不是与有机生命完全分离的, 而是基于对已有有机生命的认识而产生的。Crespo 着眼于人工智能图像生成技术与人类创造性地表达自己和认识世界的方式之间的相似之处。她的作品质疑人工智能在艺术实践中的潜力及其重塑我们对创造力的理解

⁶⁷ <https://www.jakeelwes.com/project-zizi-2019.html>

⁶⁸ <https://www.jakeelwes.com/project-zizi-2019.html>

的能力。另一方面，她还非常关注使用机器学习技术的艺术家角色的变化。在这件作品中，Crespo 提炼了我们通常对于水母“本质”或特征的看法和理解。她认为，我们对于不同事物本质的认识在相互碰撞和冲突的过程中可以产生一种潜在的梦幻般的效果，我们会在看到的既有事物中寻找共鸣。“使用深度学习算法，我们可以在数字领域探索我们与自然世界的互动，模糊人工生命和自然之间的界限。”⁶⁹



图 2-54 This Jellyfish Does Not Exist (这个水母并不存在).Sofia
Crespo.2020-2021⁷⁰

Martine Syms 从事视频、文本、表演和动态图像方面的创作，她的视频装置作品《eural Swamp (神经沼泽)》挑战了种族和性别的陈规陋习，并探讨了在一个超数字化的世界中作为黑人和女性的意义。作品是五个屏幕组成的的装置空间，艺术家将已编写好的脚本用来生成图像、视频和文字，作品中的五个声音中，有两个是机器学习提供的，艺术家利用算法和人工智能来质疑图像的政治性，同时在思考技术是如何改变我们彼此的沟通方式。

⁶⁹ <https://entangledothers.studio/this-jellyfish-does-not-exist/>

⁷⁰ <https://www.sofiacrespo.com/>

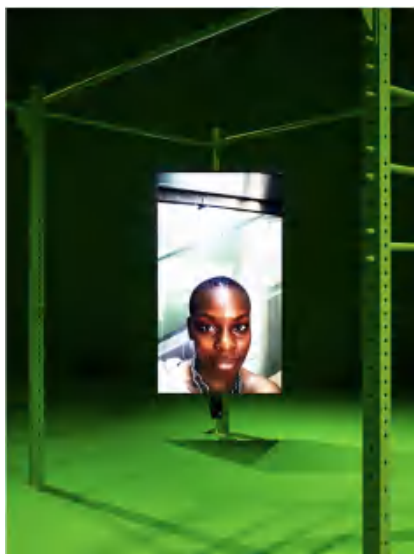


图 2-55 Neural Swamp (神经沼泽).Martine Symsi.2021⁷¹

《THE CROW (乌鸦)》2022 戛纳电影短片节最佳短片。艺术家 Glenn Marshal 用 OpenAI 的 CLIP 创建的短片 (CLIP 是一个神经网络, 可以通过自然语言监督高效地学习视觉概念) 艺术家将真人舞蹈《Painted》的视频输入 CLIP, 然后让系统生成一段“乌鸦在荒凉景观中的绘画”视频, 艺术家表示, “生成的结果几乎不需要刻意挑选。他将这归功于提示词和底层视频之间的相似性, 底层视频中描绘了一个披着黑色披肩的舞者模仿乌鸦的动作在舞蹈, AI 让每一帧画面看起来都像是真实的。”



⁷¹ <https://metrophiladelphia.com/martine-syms-neural-swamp/>



图 2-56、图 2-57 《乌鸦》（THE CROW），Glenn Marshal, 2022⁷²

《AB INFINITE 1（AB 无限 1）》是艺术家 Andrea Bonaceto 的代表作，展现了艺术家和观众之间的开创性合作。Bonaceto 将这部作品描述为“迄今为止我最深刻的作品”。该作品提供了他生活和文化循环的快照，为旅程“ab Infinite”（拉丁语中的“来自无限”，是他名字首字母的双关语）赋予了意义。该作品的观众可以使用 Instagram 和 Twitter 上的 #abinfinite1 标签上传图像，图像将通过 Bonaceto 人工智能算法进行翻译，然后纳入作品独特的视觉词汇中。通过同样的过程，在主题标签下发布的任何文本都成为作品中展示的一首不断演变诗歌，让人想起詹姆斯·乔伊斯的《尤利西斯》的足迹中建立的现代意识流。《AB INFINITE 1》是使用 Algorand 区块链上的 ARC-19 标准创建的可变 NFT。这使得 NFT 能够更新其 IPFS URL（图像），从而允许创建随时间变化的 NFT。



图 2-58 AB Infinite 1（AB 无限 1）.Andrea Bonaceto.2022⁷³

作品《Unsupervised(无监督)》是艺术家 Refik Anadol 长期项目《Machine Hallucinations（机器幻觉）》的一部分，探索基于集体视

⁷² <https://www.youtube.com/watch?v=pK7AGfBtw1Q>

⁷³ <https://www.abinfinite1.com/about>

觉记忆的数据美学空间。自 2016 年项目启动以来，Anadol 将智能机器作为人类的合作者，利用 DCGAN、PGAN 和 StyleGAN 算法，以庞大的数据作为基础进行训练，以展开探索。Anadol 和他的团队从数字档案和公共资源中收集数据，并使用机器学习分类模型处理这些数据集。作为一种精心策划的多渠道体验，《Machine Hallucinations（机器幻觉》为观众带来了一种自我再生的惊喜感，并通过控制论的偶然性提供了一种新的感官自主形式。Refik Anadol 也曾在演讲中提出“将数据作为颜料”的创新思路。这会是人们以 AI 技术记录历史，进行艺术创作的新方法。

《Unsupervised NFT Collection》正是从这样的美学愿景中产生，它将来自现代艺术博物馆（MoMA）庞大藏品的 138,151 条原数据处理在机器的“头脑”中。使用 StyleGAN2 ADA 捕捉机器对现代艺术在多维空间中的“幻觉”，Anadol 用 MoMA 艺术品收藏的子集训练了一种独特的 AI 模型，创建了 1024 维度的嵌入，并对排序后的图像数据集进行分类，以更好地理解数据的语义和上下文。这个不断扩展的数据宇宙不仅代表了数据的插值合成，还成为一种潜在的宇宙，在其中由一种新形式的艺术创造力产生的幻觉潜力解释着 MoMA 无与伦比的现代与当代艺术收藏。

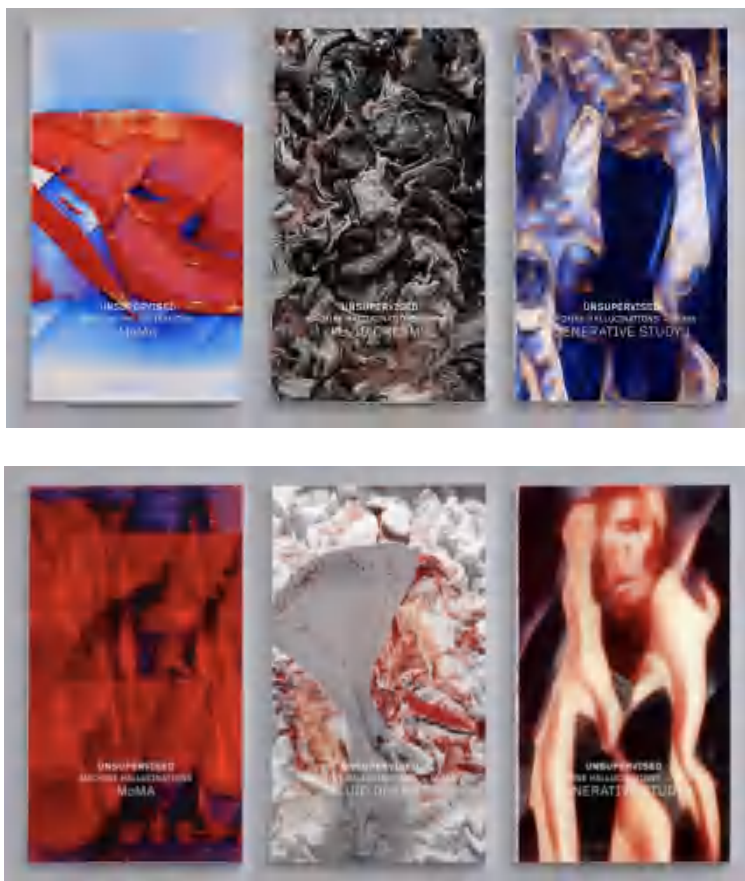


图 2-59、图 2-60 Unsupervised(无监督).Refik Anadol.2022⁷⁴

在艺术家兼制作人 Riar Rizaldi 创作的电影《化石》（Fossilis）中，他用自己的作品审视资本、技术、世界和小说的相互作用，在电影中，AI 生成的图像是物理世界和虚拟世界之间的边缘过渡，考古学家使用的机器会思考如何构建一个可以虚拟观察这些电子垃圾的世界。要形成这个虚拟世界，它需要古人类留下的电子垃圾数据集图像。艺术家在制作影片时将数字和物理废物融入到艺术实践里，《化石》的拍摄场景是由废弃的垃圾制成的，例如未使用的电缆、废弃的金属板和化粪池。《化石》中的 CGI（Computer-generated imagery）是废弃的 CGI 资产的复制品。AI 生成的图像取自 ImageNet（是一

⁷⁴ <https://refikanadol.com/works/unsupervised/>（关于该项目的更多信息请参阅 MoMA Magazine:

<https://www.moma.org/magazine/articles/658>

个用于视觉对象识别软件研究的大型可视化数据库)数据集中的电子垃圾图像。最后,街景取自印度尼西亚的“电子尸体”市场:这是一个人们出售电子废弃物的市场.这些动态影像制作技术在处理废弃物和闲置物的概念上各有千秋,既有物理上的,也有数字上的。此外,使用不同的技术也会带来不同的感官体验,这也是艺术家在这个作品中想看到的效果。



图 2-61 《Fossilis(化石)》幕后图像.Riar Rizaldi.2023⁷⁵

当代艺术家蔡国强的人工智能艺术计划 **cAI™** (读作 AI CAI), 是从其数十年的艺术哲学与方法论出发而量身开发人工智能程序。它能够深度学习蔡国强的艺术创作、著述、影像与档案资料,以及宇宙和未知世界等领域的知识。**cAI™**作为随着前沿科技与艺术家人生进程不断有机生长的多模态综合体,它本身既属于蔡国强的 AI 艺术作品,也是其创作中的合作伙伴,未来亦可能独立创作。这无疑是一个实验性的艺术计划。关于 **cAI™**蔡国强谈到:“火药和 AI 对我来说都是难以捉摸和控制的媒介,但正因此令人着迷。火药自有神奇的‘生命力’,但需点火人;现在 **cAI™** 是我的影子和镜子,是我的合作伙伴;未来它会打破这面镜子,成为独立的艺术家吗?很多人担心

⁷⁵ <https://rizaldiriar.com/fossilis.html>

AI 对人类构成威胁。我想，当前用 AI 创作艺术的危险性不在于 AI 本身，而在于，使用 AI 创作、成果却不是艺术……我想通过 cAI™ 思考：在人类迈入与人工智能共存的时代，艺术家和艺术史该何去何从？期待终有一天，cAI™ 成为超越人类认知维度、来自‘外星球’的导师……”。作为一个实验，蔡国强与 cAI™ 相互启发共同创作了火药屏风《月亮上的画布：为外星人作的计划第 38 号》。在制作屏风时，cAI™ 通过摄像头全程观摩，并实时给出视觉反馈。



图 2-62 月亮上的画布：为外星人作的计划第 38 号.蔡国强.2023⁷⁶

自然语言处理（Natural Language Processing，简称 NLP）在过去几年取得了显著的发展和进步。近年来，预训练语言模型（Pre-trained Language Models）成为 NLP 领域的重要突破。模型如 BERT（Bidirectional Encoder Representations from Transformers）、GPT-4（Generative Pre-trained Transformer 4）等，采用了深度神经网络和 Transformer 结构，通过在大规模数据上进行预训练，学习到了丰富的语言知识，为各种 NLP 任务提供了强大的基础。预训练语言模型的出现，使得机器在上下文理解和语义理解方面取得了显著的进步，从而提升了机器在文本理解、情感分析、问答系统等任务上的表现。人工智能自然语言处理技术的进步也给艺术创作带来了更多可能性，近年来也出现了一些基于语言模型的艺术项目。

⁷⁶ https://m.thepaper.cn/wifiKey_detail.jsp?contid=22806977&from=wifiKey#

人工智能生成的诗歌项目，如谷歌艺术与文化的“Poem Portraits”仅需要用户提供一个单词或短语，人工智能就可以根据输入生成一首独特的诗。然后将生成的诗歌与视觉表现相结合，以创建个性化的“诗歌肖像”。

Poem Portraits 是一件实验性的集体艺术品，由艺术家兼设计师 Es Devlin 与谷歌艺术与文化实验室和创意技术专家 Ross Goodwin 合作构思，融合了诗歌、设计和机器学习。该想法源于 2017 年 Es Devlin 与蛇形画廊艺术总监 Hans Ulrich Obrist 的一次对话。最初它是蛇形画廊夏季派对上的一个实体装置，最终成为一个线上人工智能生成的诗歌项目。



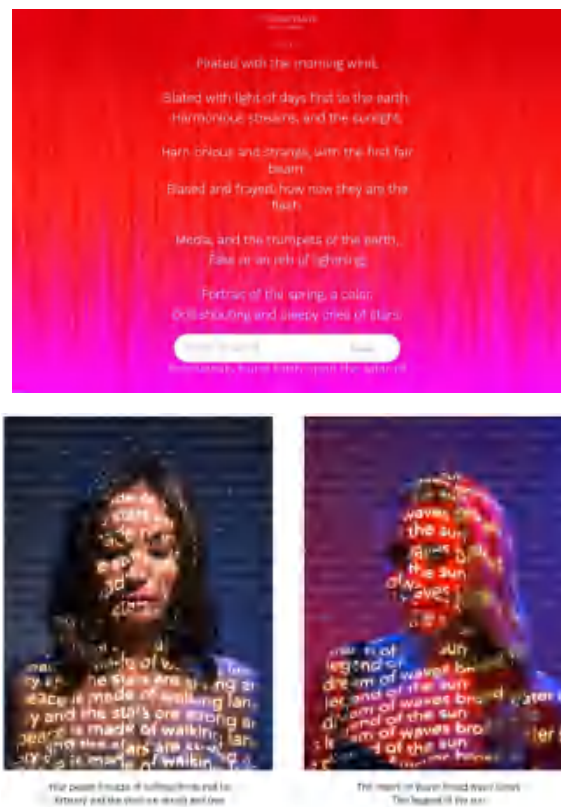


图 2-63、图 2-64、图 2-65、图 2-66 Poem Portraits (诗歌肖像).Es Devlin. 2017⁷⁷

《Alice&Bob (爱丽丝与鲍勃)》出自艺术家组合 Anna Ridler & Daria Jelonek 于 2017 年创作的作品。Anna Ridler 是一位艺术家和研究人员，她研究知识体系以及技术是如何被创造并更好地理解世界。她的作品经常涉及到处理信息或数据集，特别是以创建新的、不同寻常的叙述数据集。Daria Jelonek 是一位数字艺术家、设计师和研究者，目前生活和工作于伦敦。她的主要工作领域是交互艺术装置，关注自然与科技的关系。光、投影和动态影像是她作品中常用的视觉语言。

《爱丽丝与鲍勃》是一个无限生成情书的装置，它由四量子位的量子计算机所构成并通过其数据库进行控制。量子科学论文的摘录和标题被处理成一系列看起来是爱丽丝和鲍勃这两个虚构人物之间的手写信件。并且由于偶尔提及的第三位人物伊芙而情节变得复杂，这让爱丽丝与鲍勃的故事线像量子设备中的光子纠缠艺术家将爱情故

⁷⁷ <https://simon.bird.me/google>

事中无效算法数据的重新文本化揭示了科学、文学和艺术之间的诗学和意想不到的联系。《爱丽丝与鲍勃》不仅仅是将科学实验进行简单的可视化呈现，而是对交流模式和语言发展的诗意性思考。表达故事的方式总比故事情节更多。





图 2-67、图 2-68、图 2-69、图 2-70 Alice & Bob (爱丽丝与鲍勃). Anna Ridler & Daria Jelonek. 2017⁷⁸

Ross Goodwin 发起了名为“Wordcar”的项目，将人工智能算法与车辆相结合，让车辆在公路上行驶并生成诗歌。Ross Goodwin 将监控摄像头、GPS 设备、麦克风和时钟连接到神经网络，他在汽车后备箱装了监控摄像头，通过收集噪声使其旋转并“环顾四周”，摄像机每 20 秒钟捕捉一幅图像，首先以最直观的方式将图像文本化，利用图像识别网用一个句子来描述图像，再输入一个自由联想的文本生成语句（句子数据来源于约 200 本手工挑选的书籍的语言）。该文本现已出版成书《1 the Road》。这种跨界的艺术实验为人工智能与文学

⁷⁸ <http://annaridler.com/quantum-computing-art>

之间的交叉创作带来了新的可能性，并探索了机器智能与艺术创造之间的奇妙交融。



图 2-71 Wordcar 的眼睛是一台 Axis M3007 监控摄像机.2017⁷⁹

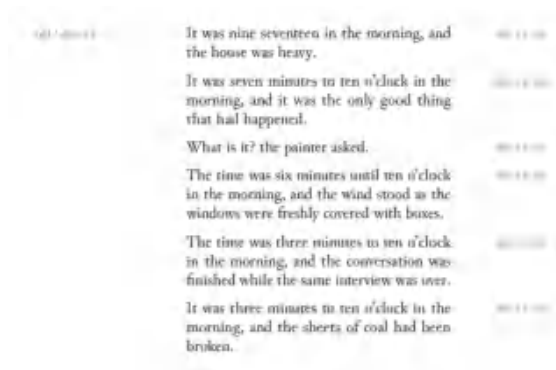


图 2-72 《1 the Road》内页.Jean Boîte Éditions 出版⁸⁰

在作品《A.I. Interprets A.I. Interpreting ‘Against Interpretation’》(Sontag 1966)中,艺术家 Jake Elwes 设计了两个人工智能,一个人工智能对苏珊·桑塔格 (Susan Sontag) 的《反对阐释》(Against Interpretation) 进行基于图像的视觉解读,另一个人工智能会将这些图像转换成语言。桑塔格在《Against Interpretation (反对阐释)》一书中写道,她不喜欢评论家对艺术作品进行过度诠释,不喜欢我们过

⁷⁹ <http://rossgoodwin.com/wordcar/>

⁸⁰ https://en.wikipedia.org/wiki/1_the_Road

多地解读艺术作品的内容和意义，而不是仅仅体验艺术作品及其形式。然而，在这段视频中，我们看到的是一个人工智能无意义地过度解读桑塔格的话，这也具有额外的先见之明，因为生成式人工智能可以说是（不可解读的）创造纯粹的模仿和形式，因为它没有任何人类艺术家的意图、意义或内容。

作品图像由图像生成扩散模型创建，以桑塔格句子作为原始输入提示词。然后使用图像标注算法（GPT2 和 CLIP）将这些图像解释为语言。这些大型预训练人工智能模型是利用互联网上获取的大量图像和文本数据集创建的，这些数据集代表了互联网某一特定时间点的资讯特征。



图 2-73 Encode/Decode(“编码/解码”).Rafael Lozano-Hemmer.2020⁸¹

京东人工智能研究院与中央美术学院合作探索 AI+艺术的进展。在合作中，双方从技术和艺术的角度创作了一系列 AI 艺术作品并多次入选参展重要的数字艺术展。同时，基于项目中研究的创新技术发表多篇高质量论文。

此项目主要聚焦于研究构建深度语义理解和图像生成模型，从艺

⁸¹ <https://www.jakeelwes.com/project-sontag.html>

术家的作品中自动学习特有的艺术风格，从而让 AI 模型能够生成拟人化的艺术作品，包括艺术性思维导图、个性化及情感化的书法作品等。同时，AI 模型还具有通用的语言理解与交互能力，能够通过语音等方式与用户互动，实时将用户的语言用艺术形式呈现。

作为多模态生成任务的尝试，艺术性思维导图生成也是一项探索工作。艺术性思维导图是中央美术学院邱志杰教授创作的特有的一种艺术表现形式，通过围绕一个关键概念联结相关的词组、符号、事件，艺术性的表征了关于特定主题的知识、信息、隐喻及触达这些知识的线索，同时通过持续的解释与抽象，思维导图的艺术性生成过程也表达了艺术家一系列信息联想扩张与主观选择决策的过程。

首先 AI 模型需要从特定主题出发，进行信息联想与选择决策。为此，AI 对四种基本信息联想方式建模，分别是：基于深度词嵌入模型（word embedding）的语义相似度联想；基于语素和音素的语言学相似度联想；模拟达达主义特征（艺术的随机性）联想；及艺术家特定风格迁移联想。其中第四类风格迁移联想是为了进一步模拟特定艺术家的联想过程。为此，基于特定艺术家的过往作品我们提取了艺术家的知识图谱，从而获得艺术家特有的联想思维特征。在创作过程中，以上四种联想方式并行进行，同时基于艺术家的过往作品数据来训练一个决策模型，决定每一步最终选择的是哪一种联想方式产生的哪一个特定扩展词汇，并加以呈现在思维导图上。

产生联想信息后（比如从一个概念出发生成新的扩展词汇后），我们需要进一步将新扩展的词汇映射到思维导图上。我们首先收集了近 3000 个中国山水画的绘画元素，通过考虑与绘画元素的语义隐喻相匹配，扩展词汇将被映射成相应的绘画元素。同时根据语义相似度我们将设定扩展次与中心词在导图中的距离。

同时，系统支持交互式的操作，可以支持用户持续的语音输入，并依据输入对当前已经呈现的思维导图进行动态扩展。为此，我们的

系统支持用户中英双语语音识别并转化为文本。同时通过关键词提取模型获取文本中的重点词语，作为中心概念进行思维导图的实时扩展。

在思维导图扩展过程中，不拘泥于传统语义相似度方法，将艺术家知识图谱，语言学特征和达达主义等方法融入其中，使得扩展出来的词汇更加具有艺术创作力。同时，人际交互的设计，使得用户在看到智能思维导图后给出更多反馈，为思维导图的持续性创作提供更多可能性。

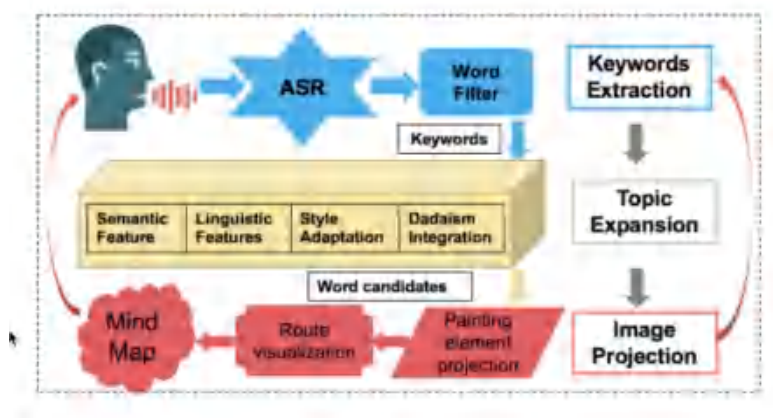


图 2-74 思维导图系统框架图

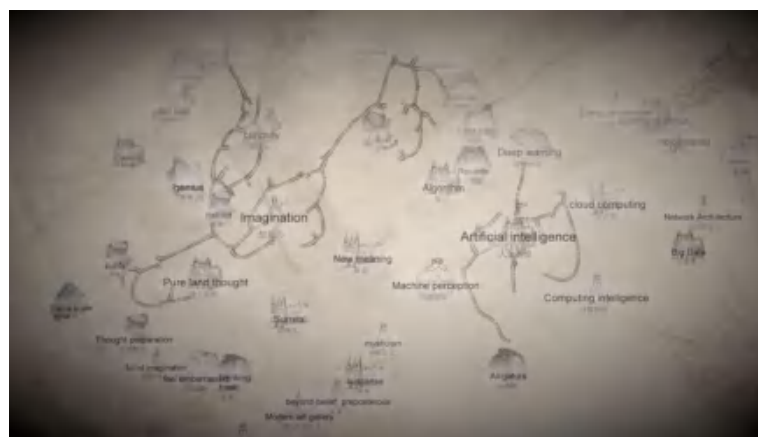


图 2-75 以“人工智能”和“想象力”为主题的思维导图创作示例

个性化情感化书法作品生成是另一项跨模态尝试，结合 NLP 情感识别技术和基于生成对抗网络（GAN）的图像生成技术，探索了语言 and 图像的多模态融合技术在书法生成中的应用。

语言和图像的多模态 GAN 模型常被用于语义可控的图像生成，但书法不只是标准的字符，往往还呈现了艺术家固有的个性，及在创作当时特定的情感（比如欣快，狂喜，悲愤等），统称为风格。书法的风格既反映在每个字的呈现上，也反映在整个篇章的布局上。为此，这项工作的 AI 风格化书法生成系统含有两个模型：

书法图像生成模型：

首次提出了基于情感的风格化书法字体生成，涵盖欣快，狂喜，悲愤等七种情绪。使用 GAN 模型，以标准楷体为源图片，根据预测的文本情绪标签，生成不同情绪的文字级别书法图像。研究方向上首次探索不同情绪风格化信息，根据文本情绪变化，生成不同风格的字体图像。在研究方法上，打破传统基于笔画抽取和笔画重构的书法文字生成方法，使用 GAN 模型进行标准字体到不同情绪字体的转化。同时，在传统 GAN 模型基础上，增加 Style-Discriminator 对不同目标字体风格进行监督，保证相同字在不同风格时的多样性。

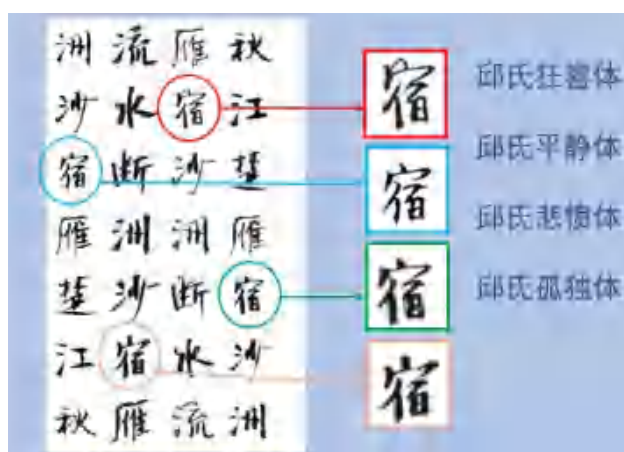


图 2-76 情绪化汉字生成示例

篇章级别书法布局模型：

这项工作中首次提出将篇章结构建模融入书法生成过程中。通过融合文字，图像等多模态信息，使用序列建模方式对每个图片大小，

间距等进行建模和预测，从而构造更具艺术性和观赏性的篇章作品。首先我们对输入的文本蕴含的情感进行检测（目前还可支持文言文类文本的情感检测），然后借助 RNN 进行序列建模，综合文字情绪特征，文字结构特征和图片特征等对篇章书法中每个字的长，宽和行距，字距等元素进行预测。进而将生成的文字图片组合为具有艺术性的书法篇章。



图 2-77 篇章书法生成系统框架图



图 2-78 书法篇章生成示例

通过这两个 AI+Art 项目，京东 AI 研究院与中央美术学院合作聚焦于在个性化、情感化、具有发散思维的场景下（如艺术创作场景）

的人工智能辅助艺术创造的可能性，并进一步推动高度拟人化人工智能技术的研究。

2.7.2 AI 交互装置案例研究

交互艺术强调观众的参与，与传统艺术相比更鼓励观众与作品进行实时的交互和反馈。观众往往可以通过触摸、倾听、行走等形式与作品进行互动。随着信息技术的进步，人机交互的界面也由最开始的命令界面、图形界面、多媒体界面朝着更智能、更复杂的混合形式发展。人工智能为主的交互作品，也由简单的线性互动逐渐发展为多维度，多结果的综合感官交互。

AI 交互装置中观众可通过自身参与直接影响作品呈现的结果，但此类作品交互模式相对线性和简单，观众只能按照预定的流程进行操作，每个步骤只有几个特定选项，呈现结果相对单一，这是种可参与的艺术装置，但不能称之为真正意义上的交互。在作品《As We Are(如我们所是)》中，艺术家 Matthew Mohr 在装置作品内设置了 29 台摄像机，从不同角度拍摄参观者的面部，系统会自动合成拍摄到的图片生成三维模型，并对五官进行修整后将图像拼接到电子屏幕上。在这件作品中每个观众根据特定的流程，进行相应操作后都可以形成专属的面部图像，但这背后互动的逻辑是相对流程化的，呈现的结果也趋向同质化。



图 2-79、图 2-80 As We Are (如我们所是).Matthew Mohr.2017——至今⁸²

Not The Only One (N'TOO) 是一项正在进行的实验。它试图从“智力”不断发展的人工智能 (AI) 的角度讲述一个美国黑人家庭多代人的回忆录。它是一个语音交互人工智能实体，其针对的是在科技领域易遭受忽视的群体的担忧和理想。

在这件作品中，人工智能实体接受来自单一家庭的三代女性提供的口述历史（数据）的训练。额外的文化协调数据提供了背景和极大的叙述范围。N'TOO 使用此信息作为回答问题的基础。值得注意的是，人工智能以独特的角色进行交流，并从第一“人称”的角度提供答案。

⁸² <https://matthew-mohr-studios.squarespace.com/>

N'TOO 基于深度学习算法。它存储在计算机上以保护社区数据，并且在不断发展。当艺术家开始这个项目时，她希望创建一个能够有逻辑地回答提出的问题的实体。在制作过程中，她发现了对一个更加动态、自由流动的实体的追求，它可以分析数据，并提出自己有限的、古怪的、有时富有洞察力的更有生成性的答案。与 N'TOO 交谈就像与两岁的孩子交谈一样。艺术家正在做的工作揭示了公众目前交互系统的一些局限性。

在这里，讲故事、艺术、技术和社会参与相结合，创造了一种新型的人工智能叙事形式。该项目致力于创建特定文化下、基于自然语言的人工智能，反映制作它们的社区的目标。通过以口述历史和创造性地讲故事，该项目希望引发有关人工智能及其对现在和未来社会影响的重要对话。⁸³



图 2-81 Not the Only One.Stephanie Dinkins.2018⁸⁴

《Please Feed The Lions（请喂狮子）》是特拉法加广场上的一座互动雕塑，由多学科艺术家兼设计师 Es Devlin 设计。艺术家将一个红狮子雕塑安置在广场上，并邀请观众通过雕塑前面的 Google

⁸³ <https://www.stephaniedinkins.com/ntoo.html>

⁸⁴ <https://www.stephaniedinkins.com/ntoo.html>

Pixelbook 来“喂”狮子一个词，每个词都会触发一对诗的生成，伴随着雕塑内置的发生系统狮子会大吼一声并将生成的诗句投射在嘴里。诗歌的生成算法语料均来自 19 世纪的诗，也是这四头狮子制作的时期。观众通过词汇生成诗歌参与到作品中，让作品有了更多的互动性和趣味感。



图 2-82、图 2-83 Please Feed The Lions（请喂食狮子）.Es Devlin.2019⁸⁵

作品灵感源自于中国古琴音乐及其特有的记谱法——减字谱。自唐代以来，古琴正是依靠减字谱流传至今，并且在时代的潮流中不断发展。作品利用人工智能图像识别算法，通过多媒体交互的方式，邀请观众参与到从古琴减字谱到古琴音乐的转化过程当中来。参与者可以依照指引或是自己的喜好不断尝试、解读、再尝试，由此体验由于

⁸⁵ <https://esdevlin.com/work/mask-somerset-house>

交互的不同所产生琴曲的不同。作品所呈现的琴与人之间交互结果的多样性，具有包容性、可解释性和一定程度的随机性，并从音乐的角度启发探讨人与人之间的多样性问题。



图 2-84 琴与人的交互 1 号.魏冰.2021

AI 应用程序被认为应该是完美无瑕的中立系统，但实际应用中它却经常表现出歧视行为，因为它们所依赖的数据集反映了人类程序员的偏见。艺术家杰克-埃尔维斯（Jake Elwes）注意到计算机系统很难识别变性人、同性恋和其他边缘化身份者，在作品《The Zizi Show》（2023）中，艺术家利用变装、性别、身份等形式来揭露和颠覆这种偏见。杰克-埃尔维斯（Jake Elwes）与一群变装艺术家合作，制作了他们身体的照片数据集。然后利用机器学习系统模仿表演者的口型动作合成新的虚拟变装舞者。视频中不断变化的身体均由神经网络生成，观众可以点击按钮选择喜欢的表演者和曲目，当 AI 试图呈现高难度的表演姿势或将多种不同身份相结合时身体会变得模糊、扭曲。在这件作品中观众通过简单的选择，影响作品的呈现，以此实现互动。

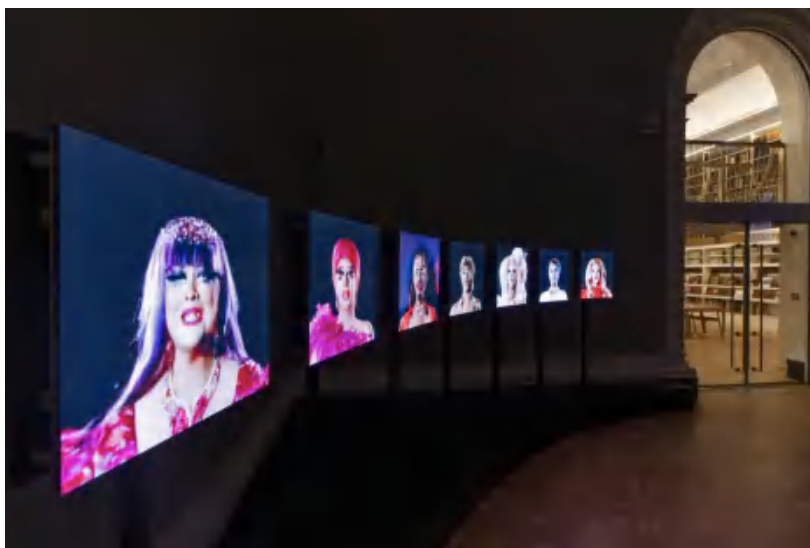


图 2-85 the Zizi Show.Jake Elwes.2023⁸⁶



图 2-86 the Zizi Show.Jake Elwes.2023⁸⁷

随着技术发展，沉浸式艺术装置开始出现在大众视野，“沉浸式”理念最早被运用在戏剧领域，在沉浸式艺术装置中作品多构建一个独立空间将感官与外界阻隔，以身体作为感知的媒介，通过声音、光、传感器等引导观众与作品互动，沉浸式艺术作品多具有很强的观赏性和参与性，作品《Archive Dreaming》是土耳其多媒体艺术家 Refik Anadol 与 Google 联手打造的实时音乐与视觉特效相结合的沉浸式艺术装置。该装置作品以艺术馆的 170 万个文档作为数据源，通过特殊

⁸⁶ <https://www.jakeelwes.com/project-zizi-vam.html>

⁸⁷ <https://www.jakeelwes.com/project-zizi-vam.html>

的编程方式实时渲染出 170 万张图像。通过数据可视化设计系统，庞大的数据文件转化为令人震撼的沉浸式空间，把观众包围其间。同时系统还呈现了每张图片的信息，观众可以与这些视觉资料实时互动，既呈现了艺术与科技的无限魅力，又传达了历史文化的浩瀚苍穹。作品模糊了物理空间和数字空间的界限，创造了沉浸式的环境，以回应观众的存在和参与。也是对数字媒体与物理实体互动关系的探索，重新定义了建筑结构与媒体艺术的紧密关系。作品带给了观众初始的沉浸式体验，但由于技术和思维的限制，其叙事性、交互的开放性和多样性仍有待提高。

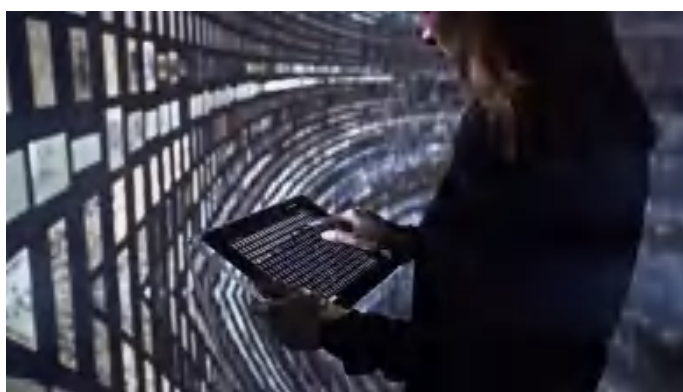
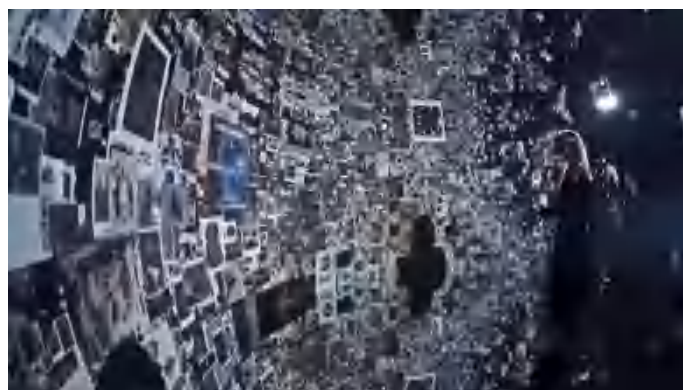




图 2-87、图 2-88、图 2-89、图 2-90 Archive Dreaming.Refik Anadol.2017⁸⁸

作品《Encode/Decode（“编码/解码”）》是艺术家 Rafael Lozano-Hemmer 根据牙买加裔英国文化研究学者斯图尔特-霍尔（Stuart Hall, 1932-2014）的评论文章制作的一个生成性互动装置。霍尔对电视语境下媒体信息的生产、传播和解读方式进行了理论分析，他认为受众可以通过其个性化的社会语境在解码信息方面发挥积极作用，并拥有通过集体行动改变意义的能力。Lozano-Hemmer 的作品中，缓慢流动的动画字母环绕在整个空间，通过相互吸引的方式排列，而参与者的身体轨迹则影响着字体流动的方向。

⁸⁸ <https://refikanadol.com/works/archive-dreaming/>



图 2-91 Encode/Decode (“编码/解码”) .Rafael Lozano-Hemmer.2020⁸⁹

除了人机交互外，艺术家也会将行为表演与 AI 技术相结合，此类交互作品多带有一定的叙事性，观众可以通过手势、语音识别、眼球追踪等技术来带入情节、展开互动，但并不能决策作品走向，作品《Unlearning Language"遗忘语言》是 Lauren Lee McCarthy & Kyle McDonald 共同完成的表演项目，作品试图探索语言对我们意味着什么？随着人工智能生成的文本激增，我们不断被检测和存档，我们能否想象一个超越持续监控的未来？Unlearning Language 是一个行为装置，它利用机器学习来激发我们寻找对语言的新理解，这种理解是算法无法检测到的。作品由两部分组成：1.最多可容纳 8 人的交互装置，一组参与者受到一台希望训练人类更不像机器的人工智能的引导。当参与者进行交流时，他们会被检测（使用语音检测、手势识别和表情检测），然后 AI 会通过光、声音和振动进行干预。整个团体必须共同努力，找到一种新的交流方式，对算法不可察觉。这可能包括拍手或哼唱，或者修改语音的速度、音调或发音。通过这种有趣的实验，人们逐渐展现出与机器区别开来的最人性化的特质。他们开始设想一个更重视人类交流的未来。2.面向观众的交互式开场表演。开场表演为装置提供了背景故事。观众被鼓励彼此互动，以及与表演者进行互动。

⁸⁹ <https://www.lozano-hemmer.com/encodedecode.php>





图 2-92、图 2-93、图 2-94、图 2-95 Unlearning Language(遗忘语言).Lauren Lee
McCarthy & Kyle McDonald.2022⁹⁰

从本质上来讲，“交互”与“互动”的概念是相似的，它们都表达的是互相影响、互相作用的过程，但是“交互”含有双向性的特点，交互艺术设计中的“交互”主要是指参与者和作品之间的双向信息交换，交互的形式涉及范围也很广，包含了声音互动、视觉互动、思想交流等等。⁹¹在交互装置创作中，观众不仅可以参与到作品中，其抉择也会影响作品呈现的结果，此结果会反作用于观众，影响下一次选择，以此循环往复。此类交互作品中参与者可以根据自己的需要和兴

⁹⁰ <https://lauren-mccarthy.com/Unlearning-Language>

⁹¹ 钱抒辰.人工智能技术与环境艺术设计课堂融合[J].居舍,2020(06):189-190.

趣自由地选择不同的操作路径，可以更加灵活地探索和体验作品。

VR 与人工智能的结合正在成为一类虚拟现实叙事的趋势。人工智能可以实时的与人交互并生成新的剧情。观众在体验时的参与会直接影响故事的走向和结果。在电影《墙中狼》(Wolves in the Walls)⁹²创建了一个 AI 驱动的虚拟形象。主角露西是一个富有想象力的小女孩，确信她家的墙中有狼，而她的家人根本无视她的想法。用户扮演露西想象中的朋友，帮助她寻找证据，揭露她家中狼的威胁。根据露西对家人的看法，每个人都以“视觉隐喻”的形式出现。例如，露西的妈妈，从一个很高的地方向下看露西，因为对她的想法持轻视和怀疑的态度。而父亲在一个巨大的音乐厅里练习音乐，似乎完全不关心露西。露西能够通过自然语言处理与观众进行交谈，跟踪观众的动作，与观众进行愈益深刻的互动，最重要的是她能记得观众是谁，与她的交往过程会成为她的人生经验的重要部分。



图 2-96 Wolves in the Walls(墙中狼) .Fable Studio 与 Oculus .2018⁹³

另外一种由人工智能来控制叙事发展的方式类似于彼得罗·加利

⁹² 《墙中狼》为 Fable Studio 与 Oculus 合作推出的第一部作品，改编自尼尔·盖曼的同名儿童向小说。

⁹³ <https://thirdrailprojects.com/wolvesinthewalls#wolves-in-the-walls>

亚诺(Pietro Gagliano)⁹⁴在 2020 年导演的动态电影《Agence》(图 2-97)。

《Agence》中的生物是由脚本化 AI 驱动的;在这个交互式 VR 作品中,观众被邀请在一个由简陋机器人组成的宇宙中扮演上帝,可以从一个星球跳到另一个星球,改善星球居民的生活环境,让他们越来越多地帮助和理解彼此,或者破坏它并使他们陷入冲突。也可以一举消灭它们,彻底结束一个星球上智慧生命的发展。观众与系统通过互动以影响模拟结果,这使得每次观看都能独特地推动故事讲述,为叙事创造无尽的可能性。

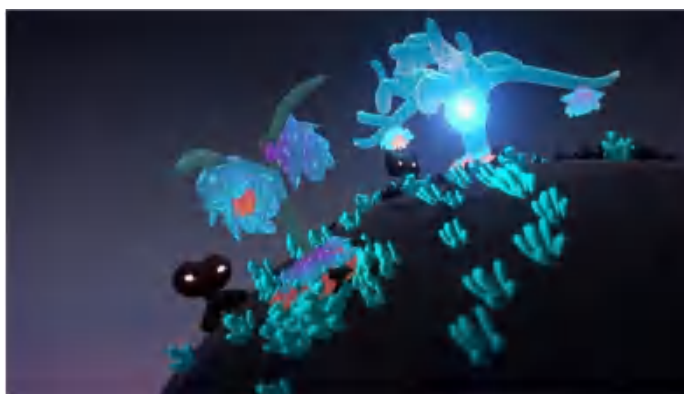


图 2-97 Agence.Pietro Gagliano.2020⁹⁵

孙羽茜在研究过程中非常关注 AI 的虚构背景,比如虚构的故事如何影响人们对 AI 的理解。她也认为游戏是一个特别好的环境,可以让人“悬置怀疑(suspend suspicion)”,暂时停止对虚构世界的怀疑,跳出自己的现实。基于这样的理念,她曾与新剧元合作推出了一款名为《一千零一夜》的游戏,这是一款以阿拉伯民间传说系列《天

⁹⁴ Pietro Gagliano 加拿大导演、Transforms.AI 的创始人,通过实时互动娱乐将 AI 与文化融为一体。

⁹⁵ 图片来自:<https://www.transforms.ai/agence>

艺术家网站: <https://www.pietrogagliano.com/>

方夜谭》为基础创作的互动游戏。玩家将扮演故事的女主人公珊鲁佐德，每天晚上为撒山国王（AI 扮演）讲故事，以推动游戏的进行。珊鲁佐德有将故事转为现实的能力，所以玩家需要引导国王，讲出自己需要的故事。当国王提及“刀”“剑”“盾”等字眼时，这些词会变成“实体”武器落入玩家手中或者兵器库，来支持玩家与撒山国王进行作战，如果玩家战胜了国王，便可解救自己与其它的战俘。《一千零一夜》与传统文本叙事游戏最大的不同点在于游戏中与 NPC 互动的剧情是不固定的，完全由玩家的语言进行控制。实现多玩法和多结果。

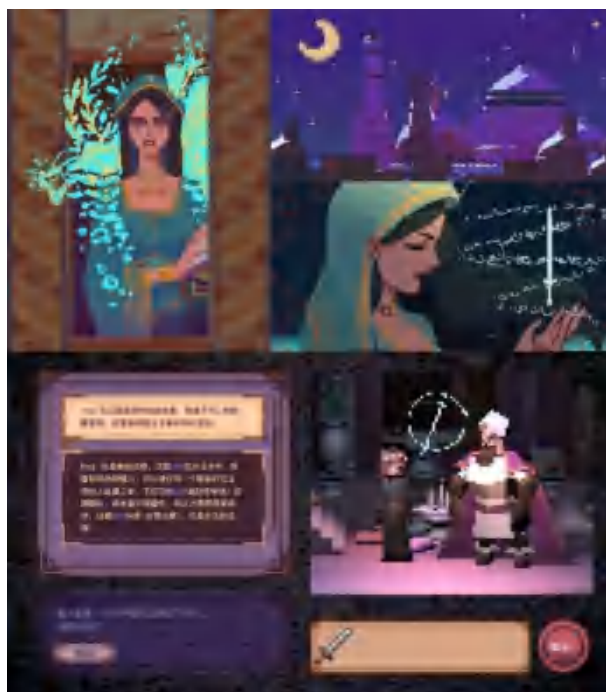


图 2-98 一千零一夜.孙羽茜.2022⁹⁶

2.7.3 多智能体系统和 AI 控制的机器人艺术

人工智能技术的介入使艺术家们有机会超越传统的数字创作边界，艺术创作也逐渐由生成，交互向多智能体过渡，多智能体系统（Multi-Agent Systems, MAS）在人工智能领域是一个研究多个自治

⁹⁶ https://www.sohu.com/a/651684244_118576

智能体如何交互合作的子领域。这里的“智能体”可以理解为能够感知其环境、做出决策并执行动作的实体。每个智能体都有自己的目标，它们可以是合作的、竞争的，或者两者的结合。它们可以独立地进行决策和交互，以实现群体或个体的目标。每一个智能体都有其自己的策略、知识和目标人格，而整个系统的行为是这些智能体之间互动的结果。这种系统可以模拟真实世界中的复杂场景，例如动物的群体行为、交通系统或经济市场的变化。

多智能体系统中的智能体，可以是软件、机器人，人类或人类团队。智能体可以分为从简单到复杂的类型：1.被动智能体或无目标智能体（如简单模拟中的物体）。2.具有简单目标的智能体（如鸟群中的鸟）。3.认知主体（具有复杂的计算）。其中智能体具有几个重要特征 1) 自治：智能体至少部分独立、有自我意识。2) 局部视角：没有智能体具有完整的全局视角，或者系统太复杂，智能体无法利用此类知识。3) 权力下放：没有任何智能体被指定为控制中心（或系统被有效地简化为单一系统）。

从数字艺术发展规律来看，数据可视化和渲染、游戏主要是基于屏幕的线索，艺术家试图利用人工智能构建人造生态系统，在这个生态系统中观察多智体生存和进化模式。这种新的创作趋势不只是为了创建吸引人的数字场景，而是着眼于模拟、创造和展示完整的、动态的、自我演化的生态环境。

艺术家陈抱阳创作的三件涉及智能体的作品，体现了这种思想。在《恭喜你，但是我在他的隔壁》中，陈抱阳构建了一个智能体循环往复走迷宫的场景。在 20 世纪的最后十年，人们喜欢不断地切换电视频道，在 21 世纪的第一个十年，他们四处奔跑。近十年来，推荐系统已使他们习惯于屏幕告诉他们的所有事情，但他们不愿成为生物电池。他们忘记了为何困于此地、他们从哪里来。耳边的声音不断告诉他们前进，下一个，下一个好，下一个会很棒，下一个就是终点。

他们总是听到“X 是划时代的”，“Y 是人类命运的转变”，“Z 时代已经到来”，但后来他们知道 Z 之后的字母是 AA。

在这里，受到神经网络训练的汽车正代表一个智能体，它根据迷宫的环境感知来做出决策。无论是它还是玛丽奥，目标始终是前进、前进、再前进。每当车辆到达迷宫的另一端，LED 屏幕都会显示一个新的迷宫，并且汽车会一次又一次地上路。是的，下一个城堡总会有一位公主。

这件作品透过对智能体的使用，探讨了人在与推荐算法互动时的主动与被动关系。人类感谢 AI 帮助解决问题，但我们不应忘记问题的起源。观众们只是看着智能体在迷宫中行动，他们没有问为什么。智能体只是在程序生成的迷宫中寻找出路，反映了一个看似忙碌但实际上是空洞的现代生活循环。



图 2-99 恭喜你，但是我在他的隔壁. 陈抱阳. 2019⁹⁷

在陈抱阳的后续作品《追逐迷雾》中，他利用多智能体的视角，采用孩童和猴群中常见的追逐游戏这种形态，构建了两个智能体相互追逐的场景。两辆受 AI 驱动的汽车在战争迷雾中相互追逐。在这件

⁹⁷ <http://baoyangchen.com/AI-Car-Maze>

作品中，陈抱阳探索了人与其所创造的智能体间，造物者与被造物的关系。通过现场设置的交互台，观众可以放置虚拟的障碍物，从而影响两个智能体追逐的情况。

多智能体系统中的每一个智能体都需要遵循某种规则或策略来与其他智能体交互。在加强学习的环境中，这些规则或策略通常是基于奖励和惩罚的机制来确定的。第一定律，或加强学习中的规则使用，强调了在这样的学习环境中，规则的明确性和一致性的重要性。

当我们谈到构建人造生态系统时，其实与创建多智能体环境有许多相似之处。首先，都需要先明确这个系统或生态中的参与者应当如何行动，以及他们的行动会带来什么后果。这些行为规范和后果其实就是这个系统或生态的规则。

这与创世纪的造物行为有一定的相似性。在宗教或神话的叙述中，创世纪通常描述了宇宙或世界的创造，以及创造者为其所创造的世界设定的规则。这些规则决定了这个世界的运行方式，以及其中的生物如何相互交互。

因此，无论是在加强学习的环境中，还是在构建人造生态系统中，规则的设定都是至关重要的。这些规则不仅决定了系统的稳定性和可持续性，还决定了系统中的参与者如何与彼此以及与外部环境互动。而对于人类来说，当我们尝试模拟或再造这样的系统时，我们实际上是在进行一种现代版的造物行为，需要深入思考并明确规则，以确保所创造的生态系统是和谐、稳定且可持续发展的。



图 2-100 追逐迷雾.陈抱阳.2020⁹⁸

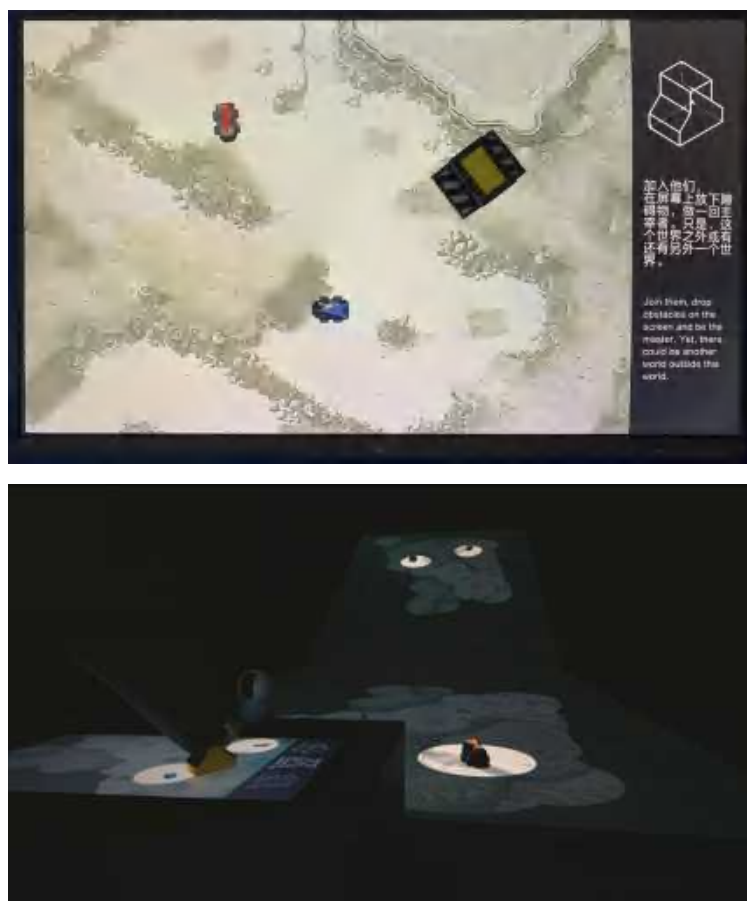


图 2-101 图 2-102 追逐迷雾（交互台）. 陈抱阳. 2020⁹⁹

⁹⁸<http://baoyangchen.com/Chasing-in-the-Fog-of-War>

⁹⁹ <http://baoyangchen.com/AI-Car-Maze>

在陈抱阳的《第一森林定律》中，他构建了一个半封闭的空间，空间中有一棵普通的绿色植物安静地生长。为了植物的生存，一套精密的人工智能系统无微不至地照顾它，确保光线、湿度和温度都可得到适当的调整。从高处观察，这个井然有序但又忙碌的景象让人称赞。然而，在这喧嚣的背后，人们骄傲地忽视了真相，直到烟雾散去，荒诞性显露无疑。艺术家认为代表数字技术的人工智能被赋予了任务，成为了看似完美的中介，桥接了人类与大地之间的距离。然而，人们早已忘记了那烟囱中冒出的浓烟，忘记了与自然之间原始而真实的联系。陈抱阳的《第一森林定律》深刻探讨了在高科技介入下，人类与自然的关系如何被重新塑造，并提醒人们不要失去对真实自然的敬畏和认知。

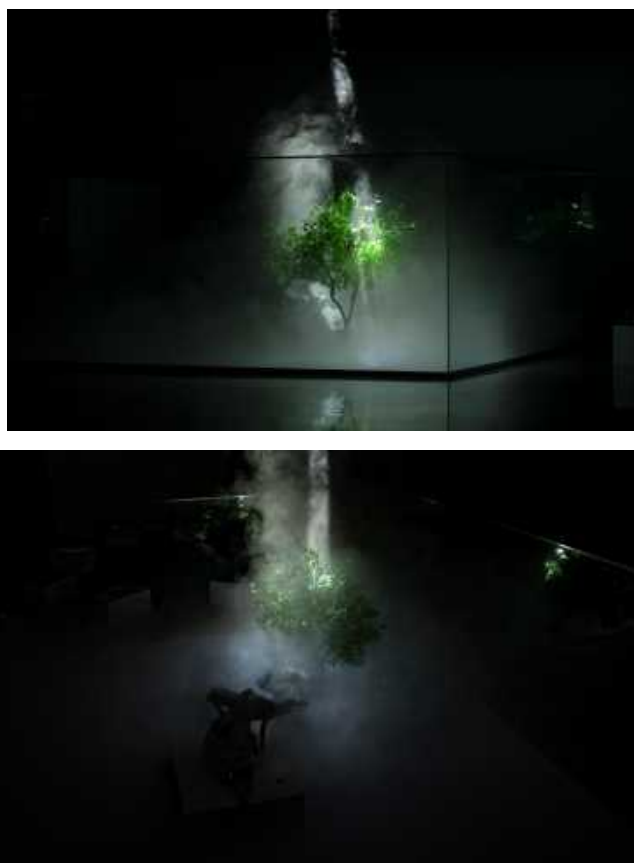


图 2-103、图 2-104 第一森林定律.陈抱阳.2020¹⁰⁰

¹⁰⁰ <http://baoyangchen.com/AI-Car-Maze>

系统构成¹⁰¹

智能体构成：

机械臂的 AI：

宗旨：照顾并确保树木的健康和生长。

功能：根据树木的健康状况调整其位置，与 DMX 舞台灯和烟雾机通信，并适应环境以保护树木。

与其他代理的互动：与 DMX 舞台灯合作提供最佳照明，并与烟雾机合作确保适当的湿度。

灯光：

宗旨：为树木提供必要的照明。

功能：自主改变其光线方向和强度。

与其他代理的互动：基于自身的内部逻辑照亮树木，但在必要时可以与机械臂的 AI 进行协调。

烟雾机：

宗旨：以雾/露的形式为树木提供所需的湿度。

功能：根据自身的行为集运行，间隔释放雾气。

与其他代理的互动：在可能与机械臂的 AI 进行协调的同时，提供适当的湿度以达到最佳的水分水平。

其他组件：

树木：

宗旨：系统旨在培育和保护的**中心实体**。

功能：在机械臂 AI 的照料下生长并保持健康。

与其他代理的互动：作为照料的接受者，其健康状况由机械臂的 AI 进行监测。

第一定律：

在其所有操作中，机械臂的人工智能系统应优先考虑树木的健康和生长。为此，它必须：

¹⁰¹ 翻译自 <http://baoyangchen.com/Forest-Theorem-One>

1. 持续监测树木的生理状态和健康，根据需要调整机械臂的位置，确保树木获得最佳的照明和水分。
2. 与两个灯光和烟雾机智能代理进行沟通和协调，确保树木得到足够的光照和湿度。
3. 在任何时候，都要防止可能对树木造成伤害或不良影响的行为或条件。¹⁰²
4. 持续学习和策略优化，以适应变化的环境条件和树木的不断变化的需求。

这些原则确保了管理机械臂的 AI 始终将树木的健康和生长放在首位，并考虑到与其他智能代理的协同工作。

表 2-2

如前文描述的系统构成中，艺术家明确了机械臂的人工智能的主要任务和功能，特别是它与其他智能体的互动方式。根据“第一原则”，机械臂的 AI 应始终优先考虑树木的健康和生长。然而，在“第一定律”中，我们注意到第三条关于防止对树木造成伤害或不良影响的规则被划去。艺术家通过这种方式，去除了该条规则的限制，并故意设置了“错误”的参数。这导致机械臂过于频繁地抬升树靠近顶部灯光以获取阳光，或将其降低以接近地面的水雾，结果使得树不堪其重，枝干断裂，叶片四散。这种设置不仅展示了 AI 的潜在风险，也对我们人类与技术之间复杂的关系进行了深入的探讨。

利用多智能体的创作在某种程度上是一种社会仿真的表现。这种系统中的每一个代理，都可以被视为一个独立的个体。它们之间的互动模拟了社会中个体之间的互动和合作，为艺术家提供了一个理想的平台，探索和表达关于组织行为、合作与竞争、信息传播和群体决策等社会现象的观点。如前面的例子，艺术家通过明确机械臂 AI 的任务并故意更改规则，展示了简单规则如何在复杂互动中导致意想不到的结果。这不仅仅是一个技术的展示，而是一个关于人与技术关系的社会隐喻。它反映了即使在高度自动化和智能化的环境中，如果缺乏

¹⁰² 此处为源网站划去，意味词条规则从改智能体系统所遵循的规则划去。

正确的道德和伦理指导，技术依然可能带来灾难性的后果。此外，这也提醒我们在将任务委托给 AI 时，需要明确、完整和正确地设置参数和规则。

从陈抱阳的三件作品中，我们不难发现多智能体系统能够从每个简单代理遵循的基础规则中展现出复杂的行为，为艺术创作提供了丰富的表现手段，揭示从简单到复杂的自然转变。当这些代理在同一环境中相互互动时，它们可能产生出预期之外的行为和模式，进一步为艺术创作带来了丰富的灵感和可能性。更为深入地，这样的系统往往可以被解读为对现实社会的隐喻，允许艺术家借此探讨和批判各种社会、文化和政治现象。此外，这种创作方式不仅仅是对技术的运用，它还要求艺术家对人文、社会和哲学有深入的了解和反思，实现技术与人文的有机结合。最后，多智能体创作也为艺术家的角色带来了新的定义。在这样的作品中，艺术家不再只是传统意义上的创作者，而更多地扮演了设计者或导演的角色，他们为系统设置规则，然后客观地观察、解读甚至干预系统的表现。

艺术家郑曦然 (Ian Cheng) 也是探索多智体话题的先驱之一。他的作品《“BOB” (Bag of Beliefs)》是一个充满生命力的数字生态系统，他创造了一个 AI 生物，它的性格、身体和生命故事在不同展览间辗转进化，其中的实体在一套预定的规则和机制下进行交互，作品展示了一个完整且持续演变的数字世界。这种创作模式超越了单纯的视觉呈现，而是强调了整个系统的互动性和自我驱动的演化过程。郑曦然称其为“自带神经系统的艺术”。



图 2-105 郑曦然 (Ian Cheng) 的《“BOB” (Bag of Beliefs)》2021¹⁰³

机器意识与艺术结合是一种新型创作模式，能够提高机器的智能和制作能力，激发出新的创意，为艺术家提供全新的表达方式。机器意识还可以利用技术挖掘艺术作品的价值所在，从而推动人类文明进步。同时，探索机器意识对艺术创作的影响，也可以促进 AI 产生重大变革。多智能体系统在艺术创作中的应用启示了机器意识的形成和演化。

来自斯坦福和谷歌的科学家在最新的研究项目中将记忆、反思（从记忆中推理）和规划代码与大语言模型结合，创建了生成型智能体，他们构建了一个名为 **Smallville**（虚拟小镇）的游戏，其中有 25 个 AI 智能体在小镇上生活，他们有工作，会八卦，能组织社交，结交新朋友，甚至举办情人节派对，每个「小镇居民」都有独特的个性和背景故事。为了让 AI 居民更加真实，**Smallville** 小镇还设置了许多公共场景，包括咖啡馆、酒吧、公园、学校、宿舍、房屋和商店等。AI 居民可以在 **Smallville** 内随处走动，进入或离开一个场所。项目通过沙盒环境和游戏框架，使用户能够实时地观察和干预智能体的行

¹⁰³ <https://lifeafterbob.io/>

为。这种实时性可以为研究者提供丰富的数据和观察机会，也增强了用户的参与感。该项目展示了人工智能和仿真领域的重要进展，对于模拟人类行为和个性化交互的研究有着重要的意义。为未来的研究和应用提供了有力的基础。



图 2-106 smallville 论文截图 2023¹⁰⁴

传统上机器或算法被设计为完成单一、固定的任务，而多智能体系统允许多个独立的 AI 实体相互协作或竞争，共同完成任务或解决问题。在艺术创作中，这种协同作用可以模拟人类的创作过程，如集体绘画或即兴音乐会等。

Patrick Tresset 是工作生活于伦敦的艺术家，他的互动装置《人类研究#1, 3RNP》由三个绘图机器人组成。观众可以坐在椅子上由机器进行视觉记录和描绘。这熟悉的场景让人联想到室内写生，人类扮演模特的角色被机器人观察、描摹。每个机器人都有自己的绘画风格，并以不同的方式处理笔触。除了牵引臂，机器人均配备了一个可移动的摄像头用来观察模型和绘制的图像。Tresset 的兴趣并不在于机器人模拟人类绘画风格而是人类和机器行为之间的差异。

¹⁰⁴ 论文地址：<https://arxiv.org/pdf/2304.03442v1.pdf>

项目地址：https://github.com/joonspk-research/generative_agents

作者使用了自动化绘图机器来捕捉人类轮廓，这些机器无法理解自己的活动。它们对艺术没有概念，也无法感知自己的绘画。艺术曾被认为是独特的人类活动，那么现在人工智能机器人所绘制的图像是否可以被称为艺术呢？《人类研究#1, 3RNP》引发了相关问题的思考，必须满足哪些条件才能被视为艺术以及艺术是否必须完全由人类作者创作呢？



图 2-107、图 2-108 Human Study #1, 3RNP. Patrick Tresset .2011¹⁰⁵

ArtSbot 是艺术家 Leonel Moura 起始于 2003 年的创作。由几个名为 Mbots 的小型自主机器人组成，每个机器人都配备了颜色检测传感器、避障传感器、微控制器和致动器，用于移动和操控笔刷。Mbots

¹⁰⁵ <https://www.fkv.de/en/patrick-tresset/>

有两种不同的行为：一种是随机行为，每当颜色传感器读取到白色时，基于小概率（通常为 $2/256$ ），机器会通过激活笔来初始化过程，以及增强传感器检测到的颜色的正反馈行为，激活相应的笔（由于有两支笔，颜色圆圈分为两个范围-“暖”和“冷”）。在这个过程中，机器人会集体生成构图，其中从随机背景中出现一些颜色簇。**Leonel Moura** 认为这本质上是一种非人类的艺术形式。2001 年，他终于能够将蚂蚁算法与一种机械臂结合并进行协同合作，产生了 **Leonel Moura** 所创造的 ArtSbot “群体绘画”。艺术家曾表示：“如果机器人和人工智能发展出一种文化，在可预见的范围内，它肯定会与我们自己的文化截然不同。但这一现实并不能改变机器可以产生人类称之为艺术的东西这一事实。这是我工作的核心问题，致力于拓展艺术的概念和实践。”

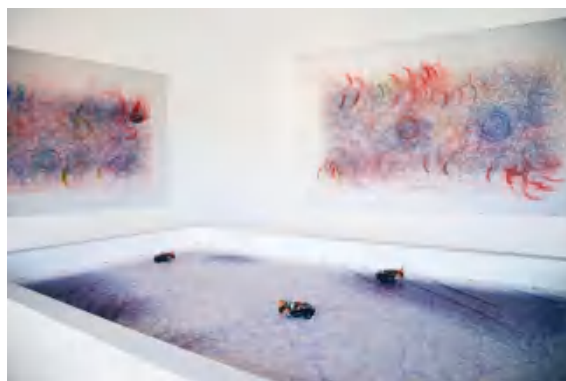
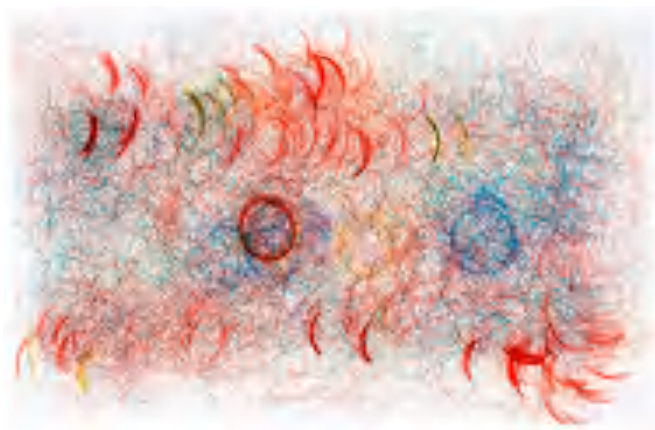


图 2-109、图 2-110 ArtSbot. Leonel Moura. 2017¹⁰⁶图 2-111 050517.Leonel Moura.2017¹⁰⁷

Empathy Swarm 是艺术家组合 Adam Donovan & Katrin Hochschuh 创作于 2019 年的作品，它是由多个人工智能机器人组成的人机交互装置。在与人类访客的连续互动中，机器人可以理解并回应访客的反馈从而唤起一种确定感即可能存在类似于机器人同理心的东西。当机器群体通过在地面运动来表达自己的时候，人类下意识地通过运动和面部表情来反应，这揭示他们的情绪状态。这些机制共同创造了一个生物反馈回路，显示了两个物种的相互依赖性。Empathy Swarm 为未来社会人工智能机器人介入人类的艺术疗愈提供了可能性。所有行为体都会影响一个不断演变的系统，创造出只有在此时此地才可能出现的独特时刻，并展开一段全新的体验。在另一个层面上 Empathy Swarm 背离了传统机器学习中对人类设定的目标进行优化的形式。相反，实现了机器的自我驱动和自我决段。

¹⁰⁶<https://www.leparisien.fr/culture-loisirs/au-grand-palais-les-robots-se-muent-en-artistes-08-04-2018-76528>

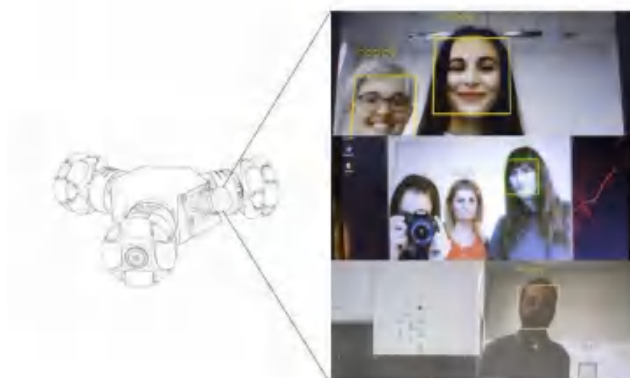
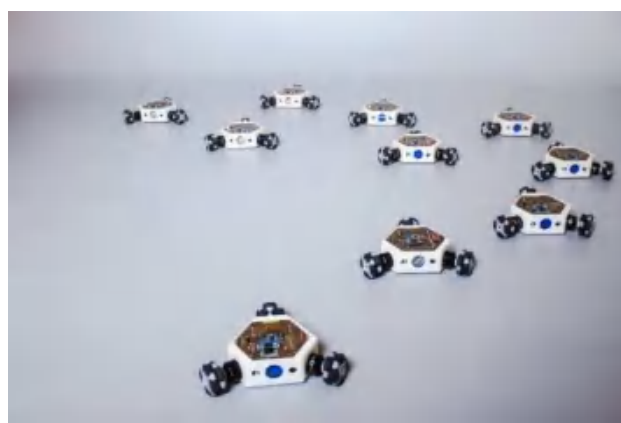


图 2-112、图 2-113、图 2-114 Empathy Swarm.Adam Donovan & Katrin Hochschule. 2019¹⁰⁸

《植物区系培育农业网络》是加拿大籍华裔跨学科艺术家钟愆君 (Sougwen Chung) 在 2020 年和她的“助手”——人工智能机器人合作完成的作品。她最得力的“助手”是一台命名为“D.O.U.G.”的机器人，艺术家通过 Google 机器学习平台 TensorFlow 上的 ML 模型对这台机器人进行训练。

¹⁰⁸ <https://esc.mur.at/de/node/2661>

作品《植物区系培育农业网络》记录了钟愆君在家 66 天隔离期间，与她的机器人助手协作共同描绘植物区系培育农业网络蓝图的过程。这个蓝图是为一个 2021 年建成的多机器人植物养殖农业所绘制。此养殖所将设计一种贴近自然的机械系统，以花卉、微生物细胞和光合作用为机械提供能量。这张蓝图便是农业系统的线路图，以此表达人类、机器、生态和谐共生的主题，与机器助手共同创作充分展现了机器意识的形成过程，其中机器在与艺术家的互动中学习和模仿人类的创造性。



图 2-115、图 2-116、图 2-117 植物区系培育农业网络.钟愆君.2020¹⁰⁹

¹⁰⁹ <https://sougwen.com/exhibitions>

艾达 (Ai-Da) 是世界上第一个超现实人工智能 (AI) 机器人艺术家，她的名字来源于被誉为“世界上第一位程序员”的 19 世纪先锋女性数学家、早期机械计算机研究者艾达·洛夫拉斯 (Ada Lovelace)。创作者是艾丹·梅勒 (Aiden Meller)，由研究员 Lucy Seal、数字艺术家和 3D 设计师共同对其面部进行设计。此外，还由利兹大学电子与电气工程学院研发了机械臂使艾达能够握住铅笔，并采用了牛津大学研究的智能算法，利用计算机视觉来分析艾达看到的图像，并开发了一个控制系统，让艾达完成绘画。

2022 年 4 月 23 日，全球首个超现实人形机器人艺术家艾达在威尼斯双年展上以艺术家的身份举办艺术作品个展，这是双年展 120 年历史上首次机器人艺术家与人类艺术家共同展出作品。艾达此次的展览名为 *Leaping into the Metaverse* (跃入元宇宙)，展览内容结合了艾伦·图灵 (Alan Mathison Turing) 理论、元宇宙，以及意大利诗人但丁·阿利吉耶里 (Dante Alighieri) 的天堂、地狱和炼狱概念，旨在探讨 AI 在当今生活中的伦理问题。其创作者表示“科技很强大，但我们把它设定在但丁的炼狱中，我们这样做的原因是因为我们不知道它到底是天堂还是地狱，而艾达将继续照亮一些困难领域。”“艺术家”在伦敦设计博物馆举办的名为《艾达：机器人肖像》的个展上展示了艾达的作品，其中包括“自画像”，鉴于机器人没有“自我”，这显然是一个悖论。这也引发了人们对数字时代身份的质疑，以及人工智能未来可能对艺术产生的影响。

与人类艺术家从生活中汲取艺术创作的灵感不同，艾达是依靠自己的程序来创作的。为了尽可能像人类，艾达的绘画风格在预设上从具象调整为抽象，并受到 20 世纪初期艺术家麦克斯·贝克曼 (Max Beckmann)、凯瑟·科尔维茨 (Kathe Kollwitz) 和毕加索等大师的影响，形成了她特有的抽象风格。例如在她橡树作品中，艾达通过其眼部的 AI 程序扫描处理一颗橡树后，再由研究人员将其处理过的图

案，绘制在一个笛卡尔坐标系上，并通过模拟人类大脑的计算系统对图像进行抽象处理，便产生了棱镜般的效果，然后再将其半成品图案打印到画布上，最后由艺术家 **Suzie Emery** 在画布上继续创作，进而完成这幅作品。





图 2-118、图 2-119、图 2-120 Ai-Da(艾达).Aiden Meller.2021¹¹⁰

武子杨&Mark Ramos 的作品《互联生态系统》是一个基于一系列数字感官所构成的生态系统的实时模拟项目。由纽约新美术馆的孵化器 NEW INC,根茎（Rhizome）和贝尔实验室（Nokia BellLabs）委托，《互联生态系统》将贝尔实验室的实验性机器人和传感器收集到的机器人视觉和传感数据，重新定位以驱动三维模拟环境，使观者可以在不断变换的环境中探索与互动。在《互联生态系统》中，自然现象被数字和人工智能系统取代，成为推动发展的力量：电力=维持，WIFI 信号=能量，Lidar 数据=火/热能，等。“生命形态”将以机器人和人工智能等形式填充这个数字生态系统。在这个网络景观中的参与者开发了新的数字感官，以体验随着环境变化而产生的数据，并在不断变化的在线环境中与模拟的世界彼此互动。

¹¹⁰ <https://www.ai-darobot.com/>



图 2-121 A.I.C.C.A.. 互联生态系统 武子杨&Mark Ramos.2021¹¹¹

A.I.C.C.A 的创作者是德国人工智能艺术家兼神经学家 Mario Klingemann。这只复古未来主义风格的艺术评论狗，出自 Ruben Östlund 的电影场景。它融合了尖端技术、视觉分析和大语言模型，是世界上第一只通过拉屎产出艺术评论的人工智能狗。基于内部复杂的算法模型，它能够考虑艺术品的构图、颜色、风格甚至语义元素等，多种因素综合影响下形成其独特的视角。在体内热敏打印机的帮助下印出评论，这种艺术评论的产出方式看起来就像机器狗在进行“排便”。得益于作者的奇思妙想，这无疑是一种既幽默又荒诞的表达形式。A.I.C.A.处于艺术、技术、人工智能的十字路口，必将掀起一场风暴。在一片对人工智能统治世界和夺取人类工作的恐慌中，Mario Klingemann 以这样幽默的方式给出了他的回答。

¹¹¹<https://www.radiancevr.co/artists/ziyang-wu-and-mark-ramos/>

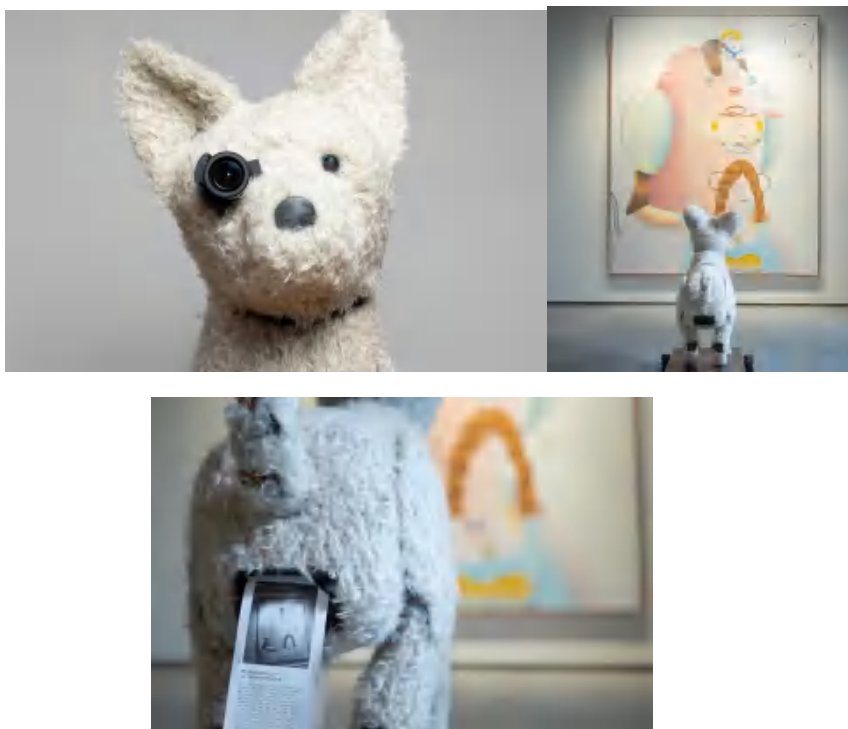


图 2-122、图 2-123、图 2-124 A.I.C.C.A.. Mario Klingemann .2023¹¹²

在艺术创作领域中，多智能体系统逐渐成为一个创新的手段，为创作者提供了一种全新的方法来表达和探索人类与机器、以及机器与机器之间的复杂关系。而想在伦敦大学学院（UCL）汪军团队与中央美术学院邱志杰、陈抱阳的团队合作探索项目“一夜狼人杀项目·AI合作博弈的表现”恰好为这一观点提供了生动的例证。

这个研究项目旨在观察 AI 在合作与竞争环境中的行为模式，尤其是它的撒谎和判断能力。一夜狼人杀，作为一款充满策略和欺骗的团队游戏，为此提供了理想的实验场地。该游戏强调团队合作、身份伪装和策略选择，正是这些因素使其成为一个富有挑战性的多智能体环境。通过将不同的 AI 模型配置为游戏中的不同角色，研究人员可以深入研究 AI 在模仿、竞争和合作等社交交往中的行为。

¹¹²<https://hypebae.com/2023/6/aicca-robotic-art-critic-dog-mario-klingemann-coleccion-solo-interview>

从艺术的角度看，这不仅是一个探索技术的项目，还是一个关于人与机器，以及机器与机器间关系的社会隐喻。通过欺骗、竞争和合作，AI 为我们提供了一面镜子，让我们重新审视自己在社会中的行为和动机。而团队希望创造的狼人杀新角色和剧情，以及基于传统故事，如农夫与蛇的新故事，都将丰富这一隐喻的深度。项目并不仅仅局限于技术实验。事实上，它为参与者提供了一个宝贵的机会，从多维度全面探索和挑战人工智能的边界。其中，大语言模型的使用为项目带来了先进的计算和学习能力，使得 AI 可以更加深入地理解和应对游戏中的各种复杂情境。

项目也探索了如何在艺术创作中将技术与人文结合，将多智能体系统融入到具有文化、社会和哲学意义的故事中，从而实现机器意识的最终目标。从认知哲学的角度，这个项目为我们提供了一个独特的视角来思考欺骗、道德和人工智能之间的关系。欺骗，在人类社交互动中，既是一种策略，也是一种道德上受到质疑的行为。而当机器被赋予了这种能力，我们不得不重新审视欺骗的本质和意义。它使我们对于道德的定义和判断，以及我们如何看待机器的行为进行了深入的反思。此外，故事创作、场景设计和装置创作为项目增添了艺术的维度。这不仅提供了一个视觉和听觉的盛宴，更为观众创造了一个沉浸式的体验，使他们可以更加深入地感受和思考游戏中的每一个决策和结果。它们为观众提供了一个框架，使他们可以从技术、哲学和艺术的角度，全面审视这一项目。

结合以上各个维度，这个项目实际上是一个跨学科的合作，它挑战了我们对于艺术、技术和哲学的传统认知，同时也展示了在当代社会中，这三者如何紧密地相互影响和互动。

2.7.4 总结

在人工智能的历史长河中，生成式技术作为其独特的支流，为艺

术史铺设了新篇章。自其诞生之日起，生成式艺术便成为了人工智能与艺术交融的标志性先驱，借助其神奇之手，把看似抽象、复杂的数据转化为生动、直观的艺术形象，为其赋予生命和情感。

人工智能艺术的发展催生了一系列创新的实验与实践。AI 不仅提供了工具和技术，更为艺术家开启了全新的创意维度。例如在数字创世纪的实践中，AI 技术使得艺术家能够在更广阔的创意空间中探索，将传统的 3D 建模和虚拟现实提升到新的高度。

而今，随着多智能体系统的兴起，艺术创作的领域又迎来了新的疆界。多智能体为我们呈现了一个协同、生态、互动的新世界，为艺术家们开辟了一个更加广阔的创作天地。这种技术的进步不仅代表了工程学的成就，更是艺术史上的里程碑。综合看来，基于 AIGC 生成式艺术的创作已在艺术史的演进中取得了坚实的进展，未来它将成为我们持续探索和期待的焦点。

第3章 AI对艺术生态的影响

3.1 AI艺术品推荐与未来趋势预测

得益于艺术品和艺术市场消费数据的积累和整理，人工智能在艺术走势预测方面的应用越来越广泛。随着越来越多的艺术品和艺术交易数据被数字化并整理，人工智能帮助分析和预测未来艺术领域的发展成为艺术行业的主要方向。因此，人工智能可以运用数据分析和机器学习技术发现艺术市场中的规律和趋势。

人工智能还可以识别不同艺术家、艺术流派和题材之间的联系，并根据历史数据和市场动态进行走势预测。这有助于投资者、艺术机构和艺术家做出更明智的决策，减少投资风险，同时也为艺术市场的健康发展提供指导。通过对网络平台、社交媒体和展览活动等数据的分析，人工智能可以捕捉到新兴艺术家的崛起以及受欢迎的艺术风格的兴起。从而带动艺术市场的多样化和创新，为新兴艺术家提供更多机会与曝光。此外，基于历史艺术品价格和交易数据，人工智能可以构建预测模型，帮助投资者和收藏家评估艺术品的投资潜力和未来价值。它可以辅助判断艺术品的市场表现，识别高潜力的收藏品以及预测未来可能出现的投资机会。这为收藏家和投资者提供了更科学的决策依据，增加投资成功的可能性。

作为艺术品拍卖行业的先驱，2018年苏富比宣布收购了 Thread Genius。Thread Genius 是一家专门从事图像识别和推荐技术的人工智能初创公司，其应用程序可以立即识别物体同时能够向观看者推荐类似物体的图像。Mei Moses 是苏富比于 2016 年收购的一个包含 50,000 张图像的艺术数据库，可以通过在线分析数字图像来识别一系列相似的对象。Thread Genius 将算法与 Mei Moses¹¹³ 艺术索引相

¹¹³ 苏富比 Mei Moses 指数于 2016 年收购，被广泛认为是衡量艺术品市场状况的卓越指标。该指数由纽约大学斯特恩商学院教授梅建平博士和迈克尔摩西博士于 2002 年开发。

结合。通过大量的数据处理和分析，Thread Genius 帮助藏家群体和潜在客户提供了专业支持。

纽约知名在线艺术经纪公司 Artsy¹¹⁴作为全球首个致力于人工智能创作者和艺术家的的人工智能交易平台，聘请了人工智能专家 Hugo Liu¹¹⁵ 担任首席科学家以扩展其机器学习和预测能力。Artsy 平台允许用户创作、销售和购买人工智能制作的艺术品，同时根据算法为藏家推荐需要的人工智能艺术作品。

区别于艺术品交易机构，Artrendex¹¹⁶公司开发了一个名为“AI Powered Art”的平台，该平台使用自然语言处理技术，分析艺术家的文本描述和评论，从而推荐与之相匹配的艺术品，为策展和艺术品推荐提供智能化的解决方案。该艺术推荐引擎产品，以类似于 Netflix 在分析用户的历史观看模式后向其推荐内容的方式。通过收集和分析观众的艺术品偏好、历史浏览记录和社交媒体数据等信息，人工智能可以建立观众的兴趣模型，并根据这些模型为每位观众推荐适合他们口味和兴趣的艺术品。

Artnome¹¹⁷是一家专注于数字艺术和数据分析的公司。由于强大的数据库和技术支持，他们使用机器学习算法来分析艺术市场的历史交易数据和艺术家的创作风格进而预测未来数字艺术品的市场趋势

¹¹⁴ Artsy 是一家总部位于纽约市的在线艺术经纪公司。其主要业务是为众多画廊开发和托管网站销售艺术品。它利用搜索引擎和数据库来绘制艺术作品之间的联系和映射关系

¹¹⁵ Hugo Liu 博士是 Artsy 的首席科学家，也是 Artadvisor 的联合创始人兼首席执行官，这是一款针对艺术收藏家的应用程序，可以根据数十万艺术家的职业成就和二级市场表现无缝预测和跟踪他们的价值。Artadvisor 于 2017 年 4 月被 Artsy 收购。

¹¹⁶ Artrendex 是一家由科斯拉 (Khosla) 支持的企业，希望通过为艺术市场创建人工智能解决方案来利用人工智能艺术淘金热。Artrendex 由罗格斯大学计算机科学教授 Ahmed Elgammal 领导。

¹¹⁷ Artnome 拥有庞大的分析数据库，收录了许多重要的艺术家的已知作品。使用机器学习和计算机视觉等数据科学的最新工具，基于该数据库进行原创研究。

和价格变化。**Artnome** 收录了世界上不同时期重要艺术家的已知作品，使用机器学习和计算机视觉等数据科学的最新工具，基于该数据库进行原创研究。传统上对于艺术品的分析和价格都来自拍卖行，但实际上只有一小部分艺术家的一部分作品流入拍卖行，因此对于没有进入拍卖行的作品如何评估和定价成为一个问题。**Artnome** 花费了数千小时和数百万美元建立了世界上最大的蓝筹艺术家完整作品数据库，以帮助解决这个问题。然而艺术作品价格浮动的因素有很多，由于顶级作品数量的稀缺即使是专家也很难预测顶级作品的价格。**Artnome** 的模型在预测估计 300 万美元或以下的作品的价格方面做得非常出色，基本上得出了与佳士得专家相同的估计值，得益于模型中有大量该价格作品的相关数据。

这些平台利用人工智能的自然语言处理技术、图像识别技术分析艺术家作品的图像、文本描述和评论。并且对社交媒体平台上的艺术讨论和话题进行实时监测和分析从中了解公众对艺术的关注和讨论焦点。从而相对快捷地推测出可能引领未来艺术潮流的话题和趋势，为客户提供关于艺术品背后故事和情感意义的解读。即使有着多年经验的拍卖公司苏富比和佳士得也要通过积攒大量的艺术家和藏家数据，分析艺术家的创作历史和行为模式，观众的浏览和购买行为，才能逐步探索出潜在的艺术创作和消费趋势并预测艺术家和观众在未来可能的行为模式。综上所述，这些平台不仅借助人工智能技术深度分析艺术世界的各个方面，还通过实时监测社交媒体上的艺术讨论来抓住时下艺术的脉搏。它们的能力不仅让市场参与者更好地了解艺术的趋势和潜在价值，也为艺术爱好者和收藏家提供了更多的见解和决策支持。即使对于拥有多年经验的拍卖公司也需要不断积累数据、分析趋势，以适应不断变化的艺术市场。这些平台的存在为艺术世界带来了前所未有的智能化和洞察力，预示着艺术和科技的融合将继续塑造未来的艺术潮流和市场格局。因此，无论是从商业角度还是从艺术

欣赏的角度，这些平台都在为艺术领域的发展和创新作出了重要贡献。

3.2 艺术品鉴定工具

艺术品鉴定一直是画廊与拍卖行的一项重要业务之一，对于专业人员和机构来说，鉴定艺术品是一件慎重的事情，鉴定周期过长和昂贵的费用都为艺术品鉴定增加成本。然而，随着人工智能技术的发展，解决方案已经被世界上多个机构与研究所提出，为艺术品鉴定带来了新的可能性。这种人工智能的解决方案将通过模型和算法来检测作品的真实性，从而在传统分析方法的基础上提供另一层安全保障。

人工智能鉴定程序的工作原理是尽可能多地研究艺术家的作品，以建立一个特征数据库。这些特征数据库包括画面的笔触、颜色与线条等艺术作品的特征。除此之外，人工智能还会深入分析每个作品中的细微细节，如艺术家画作的每个点在纸上施加了多少压力、笔移动以创建每个标记的方向、哪些标记流入其他标记等等特征。通过对这些特征的学习和比对，人工智能能够建立起每位艺术家独特的艺术风格与技巧的模型。

2019年瑞士艺术品鉴定公司 **Art Recognition**¹¹⁸开始提供人工智能用于艺术品鉴定的服务。**Art Recognition** 声称他们的应用程序使用一种简单的方式来验证艺术品的真实性，即通过分析上传的艺术品照片。一旦客户上传了艺术品的照片，将立即收到一份详细而加密的人工智能报告和证书，而无须担心任何运输延误或额外的保险费用，因为一切都在数字化管理下完成。

Art Recognition 的创始人 **Carina Popovici** 博士和她的联合创始人 **Christiane Hoppe-Oehl** 开发了一种算法，该算法仅使用一张相关艺术品的照片就能正确检测出作品是否为赝品。他们的神经网络是使用机器学习和艺术家作品的图像集进行训练的。目前 **Art Recognition**

¹¹⁸ <https://art-recognition.com/about-us/>

已完成 500 多项人工智能真实性评估，为超过 100 家私人收藏家、拍卖行、艺术品经销商、咨询公司、画廊和法律机构提供服务。然而这项技术存在着瓶颈，该算法往往更适合印象派画家，原因是这种风格的结构非常丰富，笔触大多定义明确。如果画作的线条和笔触不清晰，就很难识别。另外针对不同类型的艺术品，其数据库在识别和检测的效果上也有着很明显的差异。艺术家在不同地域、不同时期产出的艺术作品在材料、线条、颜色、境界、意趣上都有着很大的差异。因此作为一个辅助鉴定的人工智能程序，人们在鉴定时需要考虑到艺术家的时代、情感因素和环境与时代技术，当前的人工智能很难做到如此全面的分析。

2023 年 3 月 1 日由亚太人工智能学会与深圳市三品美术馆合作的全国首家 AI 艺术鉴定系统的技术成功完成了测试该项目用于书法作品和绘画作品的鉴定。该 AI 鉴定系统采用 EfficientNet_v2 模型，该模型属于较为先进的神经网络分类模型，相较于其他传统分类模型（如 ResNet、VGG 等模型），具备更快的预测速度、更高的精确度。模型形状内部包含 Fused-MBConv 模块和 MBConv 模块，模型参数共计 22495570，训练参数 22334722，模型总大小 258MB。虑到不同书法家具有独特的笔法特征，采用迭代学习的方法来训练神经网络来捕捉不同书法家笔画的特点。这使得 EfficientNet_v2 模型运用统计学方法来识别其真伪，能够有效地辨别低仿的赝品同时对于高仿赝品。该项目针对书法家管峻老师的作品，共训练了 10 个模型。随着训练量的逐步增加，每个模型都使用了不断增加的数据集，并根据真迹 100%、高仿 50%、低仿 0% 的目标来计算平均准确率。经过对比分析，发现管峻老师的第 10 号模型效果较为优秀。该模型所使用的数据量约为 2 万多字，在经过数据扩充后达到了 38 万字。这使得该模型能够较为准确地识别出管峻老师的小楷字体，为书法作品的鉴定和认证提供了有力支持。

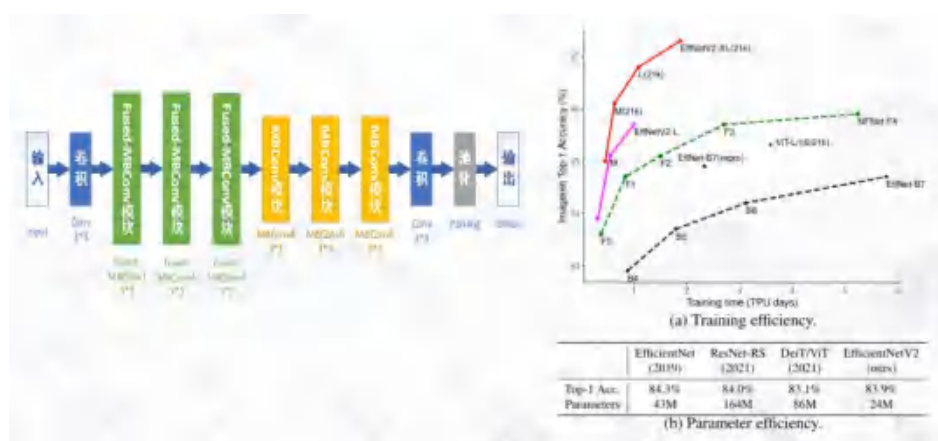


图 3-1 119

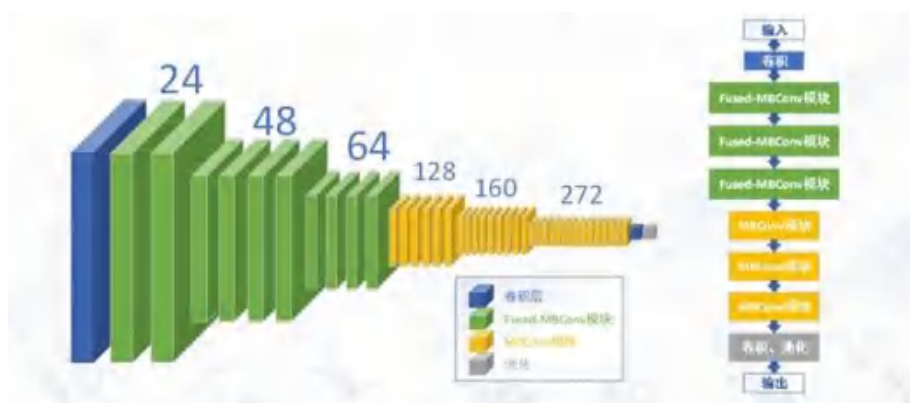


图 3-2 120

在题为《毕加索、马蒂斯，还是赝品？绘画笔触之属性与真伪的自动研究》（Picasso, Matisse, or a Fake? Automated Analysis of Drawings at the Stroke Level for Attribution and Authentication）的论文中，来自新泽西罗格斯大学（Rutgers University）与荷兰绘画修复与研究工作室（Atelier for Restoration & Research of Paintings）的研究人员公布了他们的最新发现。他们成功地开发了一项人工智能技术，该

¹¹⁹ <https://hubei.china.com/chuangtou/2023/0304/299974774.shtml>

¹²⁰ <https://hubei.china.com/chuangtou/2023/0304/299974774.shtml>

技术能够通过对比艺术作品的绘画笔触来检测伪造的艺术品。这项创新性的研究为艺术品鉴定与认证领域带来了重要的突破。

在这个项目中，研究人员为人工智能提供了来自著名艺术家，如毕加索、马蒂斯和席勒等的 300 幅线条素描作品。通过对这些作品的深入分析，人工智能成功地识别了 8 万条个人笔触，并利用神经网络学习了这些笔触特征与不同艺术家风格之间的对应关系。同时，他们还研发了一种专门负责搜寻这些特征的机器学习算法，例如线宽差异，用以反映艺术家的绘画手法和发力程度。这项研究对于艺术品的鉴定和真伪认定具有重要意义，为艺术市场和文化遗产保护提供了新的技术手段。借助人工智能的分析和学习能力，我们有望更准确地辨认出珍贵艺术品，并确保艺术遗产的准确传承。

为了推动中国艺术品市场的健康发展并构建诚信生态，艺术品造假和鉴定成本与周期的问题已经成为严重的障碍。为此，中国艺术品数据中心网与多家国内外投资机构和艺术组织共同发起了“中国艺术品登记认证系统”¹²¹项目。该项目利用区块链技术和 AI 人工智能技术，专注于艺术品的登记认证和溯源鉴定。中国艺术品登记认证系统实现了一种全新的艺术品信息化认证管理模式，包括区块链登记认证、作品信息化管理和 AI 智能识别。借助区块链技术的不可篡改性和透明性，艺术品的登记和认证得以更加可信和安全。同时，通过 AI 人工智能技术，该系统能够自动化地识别艺术品的特征和历史信息，从而实现更高效、准确的鉴定。

作为一个具有信誉和可信度的认证平台，它将有助于减少艺术品造假的风险、缩短鉴定周期和降低鉴定成本，从而提升了艺术品交易的信任度。该系统除了认证艺术家身份，以及完成艺术作品的登记，还通过 AI 来对比识别艺术家和艺术作品。能够为平台提供高效率的管理，为藏家也进行有效的鉴定服务。

¹²¹ <http://yidd.art/aboutUs.html>

这些人工智能鉴定艺术品的解决方案，不仅能够大幅缩短鉴定的周期，减少专业人员的工作负担，还能在艺术市场中为买家和卖家提供更高的安全保障。通过将大量的艺术品特征数据与已知的艺术家作品进行比对，人工智能能够发现哪些可能是伪造的艺术品，辅助专业人员做出准确的鉴定。此外，这些解决方案还可以拓展至公共领域，为艺术品的鉴赏、保护和文化遗产的传承提供有力支持。

3.3 艺术治疗

艺术治疗是一种治疗干预形式，利用艺术创作过程来改善一个人的身体、心理和情感健康。在艺术治疗中，患者被鼓励用创造性的方式表达他们的情感、想法和体验，从而帮助他们更深入地理解自己，减轻精神压力，增强自尊心，以及改善心理和情感健康。AI 艺术治疗是指人工智能（AI）技术和手法在艺术治疗领域的应用。当人工智能融入艺术治疗时，人工智能工具可以为个人提供交互式平台或软件，使他们能够创作数字艺术，增强个人的创造性表达。另一方面人工智能可以分析和解释个人创作的视觉艺术作品中的情感内容。这种分析可以深入了解艺术所表达的情感状态、经历或主题。AI 在艺术治疗中的应用也可以用于教育和培训新一代艺术治疗师。通过使用 AI 工具和模型，未来的治疗师可以更好地了解如何与患者合作，并提供更有效的治疗。人工智能在个性化艺术治疗方面的应用方面为个体提供了量身定制的艺术治疗体验。人工智能可以通过情感识别技术，分析患者的情绪和心理状态。基于这些数据，AI 推荐适合患者当前情绪和需求的艺术形式，如绘画、音乐、舞蹈等，从而提供针对性的情绪导向艺术治疗。根据患者的喜好和个性特点，生成符合他们独特需求的艺术作品。这些作品可以用于自我表达、情感释放和心理疏导，提供一种个性化的艺术治疗体验。

AI 艺术治疗可以扩大艺术治疗的受众范围，能够通过定制化治

疗方案来满足不同患者的需求，包括那些可能无法参与传统艺术治疗的人群，如远程地区的患者、特殊需求群体等。这有助于提高艺术治疗的多样性和包容性。并且将数字媒介引入了艺术治疗意味着患者可以使用虚拟工具、音乐生成等数字媒介来表达情感，同时也可以与传统的艺术媒介如绘画和音乐相结合。这种融合可以激发更多的创造性表达。

Woebot 是由斯坦福大学研究团队开发的一个基于人工智能的情感支持治疗机器人。它采用自然语言处理和情感识别技术，与用户进行对话，并提供心理健康方面的支持和指导。用户可以通过手机应用与 **Woebot** 进行交流，分享自己的感受和情绪，接受情绪调节和心理疏导的建议。最近，**Woebot** 的研究团队进行了一项试验，探索其在艺术治疗方面的潜力。他们将情感支持治疗机器人与创作艺术结合起来，让用户通过绘画和创作表达自己的情感和内心体验。与传统艺术治疗由专业的艺术治疗师或心理治疗师来引导和监督不同，**Woebot** 而是 **AI** 与患者一起探索他们创作的艺术作品，帮助他们解释和理解作品中的情感和思想以及与现实生活的联系。**Woebot** 会根据用户的绘画作品，提供相应的情感解读和反馈，帮助用户更好地理解自己的情绪和需求，并通过艺术创作来促进情感释放和心理疏导。

虽然 **Woebot** 这个案例还处于试验阶段，并需要进一步的研究和验证，但它展示了人工智能在个性化艺术治疗方面的潜力。通过结合情感支持治疗和艺术创作，**Woebot** 为用户提供了一种创新的个性化治疗体验，帮助他们更好地应对情绪困扰和心理压力。

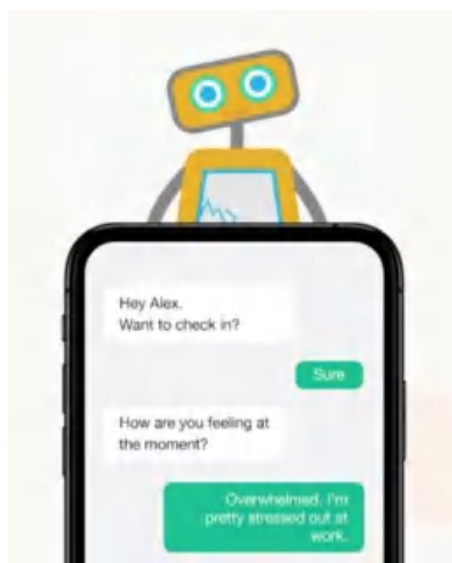
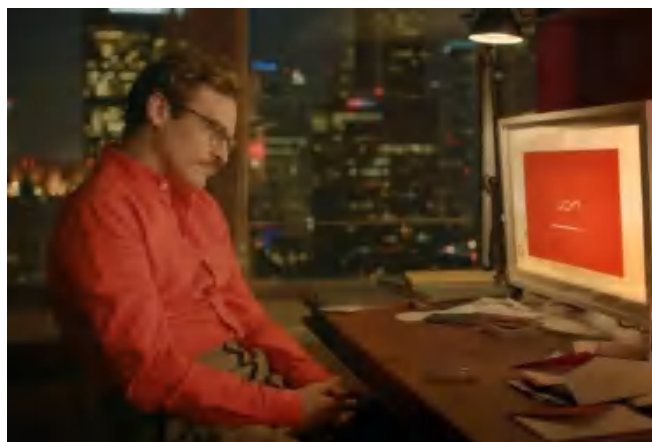


图 3-2¹²²

上海的初创公司开发的 **emotibot**¹²³通过理解用户的上下文和情绪来做出响应。**emotibot** 作为辅助型机器人，为用户提供信息、预订酒店、安排日程等，但更重要的功能是将情感与业务联系起来，正如科幻电影《她》里面那样，为人工智能注入情感关怀。



图电影《她》剧照 3-3¹²⁴

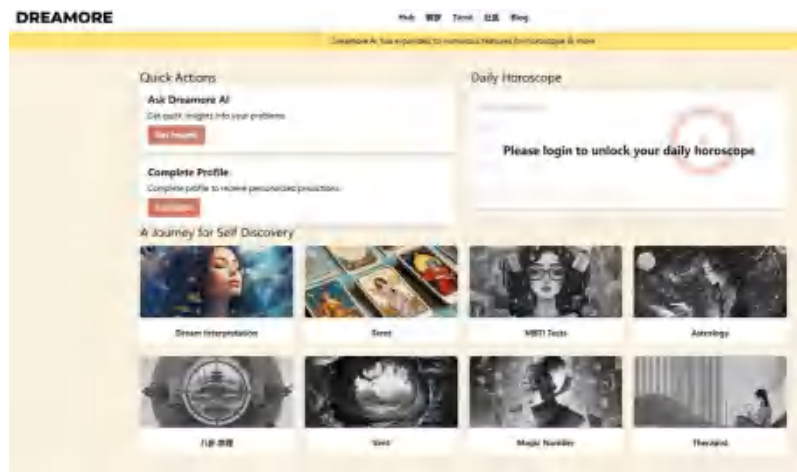
¹²²<https://img1.baidu.com/it/u=3806238247,2755546307&fm=253&fmt=auto&app=138&f=PNG?w=403&h=500>

500

¹²³ <https://www.emotibot.com/>

¹²⁴ <https://www.tmtpost.com/4203073.html>

艺术治疗属于一种心理治疗方法，它通过艺术媒介（如绘画、雕塑、音乐、舞蹈、戏剧等）来帮助个体处理情感问题、提升心理健康和促进康复。谈及心理治疗经常会提到弗洛伊德《梦的解析》，梦境能反映很多隐藏的情绪和潜意识，梦境解析也是心理治疗中的一种常规方法。成立于 2023 年的 Dreamore¹²⁵是一款人工智能驱动的解梦应用程序，dreamore 利用 openai 的 API 将用户的梦境分析并生成艺术作品帮助处理用户的情绪和经历。Dreamore 解释和探索用户梦的意义和象征，为用户提供潜意识的洞察和理解。为了提高梦境解析的准确性，用户需要尽可能地多提供细节，描述得越详细，解释将会越准确。

图 3-3¹²⁶

AI 在艺术领域的应用不仅可以用于治疗，还可以激发艺术创新和实验。艺术家和创作者可以探索 AI 生成的艺术作品，将其与自己的创作相结合，从而推动艺术领域的发展。“情绪几何”的 1.0 版本是来自中央美术学院费俊教授与数学家许晨阳的合作成果，在《情绪

¹²⁵ <https://dreamore.app/zh>

¹²⁶ <https://dreamore.app/zh>

几何》3.0 中，艺术家为观众提供了一种与作品共同生成的互动体验。观众被邀请进入互动装置的暗箱中，与几何模型进行亲身接触。在此过程中，触摸的情绪通过心率等生理数据采集设备记录并计算。这些情绪数据会按照三种情绪气质类型进行分类，并结合观众的情绪状态生成一系列“因人而异”的情绪几何图形。观众的情绪和参与进程被转化为独特的艺术创作，创造出一系列充满情感和个性化的情绪几何图形。通过这种独特的互动方式，艺术家成功地将观众与作品之间建立了更为深刻的联结，同时赋予了观众参与艺术创作的感受与权力。这种创新的互动艺术形式为艺术的表达与体验带来了新的维度，丰富了观众的感知体验，同时也让艺术作品本身成了一个与情绪和个性共鸣的平台。



图 3-4 《情绪几何》.费俊.2019¹²⁷.

费俊教授的另一件作品《水曰》同样是一件结合人工智能的作品。《水曰》的现场体验方式非常简单，观众只需站在湖边，通过一个类似传声喇叭的交互装置，与水进行倾诉。为了实现这一作品的交互体

¹²⁷ https://p2.itc.cn/q_70/images03/20211119/841bd00e96204594840830abc7fd433c.jpeg

验，费俊教授与中科院的另外两位科学家连政博士和孙旭冉博士合作，利用他们开发出的人工智能声纹情绪识别 SDK（软件开发工具包）。早期的测试得出了一个基本数据，该系统能够分辨快乐、愤怒、忧伤等多种情绪，并结合对能量值的判断，实现对观众基本情绪状态的捕捉。

基于这套声纹识别系统，费俊教授设计了水的四类情绪：释然、平静、活泼、兴奋，并设置了情绪反应的随机率使水能够产生不断变化的回应。这样，观众在湖边倾诉时，水将以一种更加灵动与变化的方式回应，仿佛与观众进行了一场情感交流。水会根据观众所表现出的情绪以涟漪的形式回应。当观众表现出忧伤和压抑的情绪气质时，水通常会呈现出雀跃欢腾的状态；相反，如果系统识别到观众带有愤怒和激烈的情绪，水则会以缓慢禅意的方式回应，仿佛在传递着“放下”的意味。通过这种创新的交互方式，观众不仅与自然产生了亲密的情感联系，还能在情绪的共鸣与反馈中获得独特的体验。整个过程犹如一种“水疗”，充满着治愈性。《水曰》作为一件融合了人工智能技术与自然元素的艺术品，通过智能算法的处理，水与观众情感产生共鸣，用涟漪的回应传达出一种与观众心灵连接的情感体验。这种创新的艺术互动形式使得观众与作品之间产生了更深层次的情感交流，犹如一种艺术“水疗”，充满着治愈性。

图 3-5 《水曰》.费俊.2021¹²⁸

日本首个人工智能与植物结合的艺术疗愈体验展《NAKED FLOWERS FOR YOU》于3月19日在东京由NAKED Inc.举行，为人们带来一场独特的植物疗愈艺术体验。该展览巧妙地将“PHYTOTHERAPY（植物疗法）”与AI技术相结合，创造出令人心旷神怡的花卉疗愈之旅。在这场展览中，AI技术的应用让每位参观者都能获得个性化的花园体验。通过AI技术的智能分析，展览为每个人打造的四种独特个性化的花园，使观众能够与花卉之间建立更深层次的情感联系。这种个性化的设计让每位参观者感受到与花朵之间的独特互动，增强了观众在花卉疗愈过程中的愉悦感。比如在野生植物力量的草药花园中，参观者可以通过扫描手机二维码，通过人工智能自行诊断并找到适合自己的草药。在指定的地点，他们可以将两种草药混合，并将这些草药带回家。根据诊断的类型，草本茶、袋装茶和浴盐等不同的使用方法都可以让人们充分享受草本植物的独特效能。

随着技术的不断进步和研究的深入，人工智能将继续为医疗领域

¹²⁸ https://p8.itc.cn/q_70/images03/20211119/a62f17f45a4d4c218b3045eeb5ba39dd.jpeg

带来更多创新和改进。人工智能与艺术治疗的结合可以增强治疗体验，促进自我表达，并为探索和治疗提供新的途径。人工智能在艺术疗愈方面有着多样化的应用，从虚拟现实艺术创作到音乐生成，从情感解读到情绪导向，都为用户提供了更个性化和创新的艺术疗愈体验。虽然这些应用还在不断发展和完善中，但它们为我们展示了人工智能在艺术疗愈领域的巨大潜力和可能性。

尽管人工智能在艺术治疗中有许多潜力，但仍然需要在伦理、法律和隐私方面考虑诸多问题。确保数据安全和患者隐私，以及准确性和可解释性是关键关注点。同时，医生仍然在治疗决策中扮演着不可替代的角色，AI 应该被视为医生的有力工具，而不是替代品。AI 艺术治疗应被视为传统艺术治疗实践的补充，并应在训练有素的艺术治疗师或心理健康专业人员的指导和监督下进行。

总体而言，AI 与艺术治疗的结合代表了艺术与医疗领域的突破。AI 艺术治疗对艺术生态系统产生了积极的影响，为多样性、包容性、创新和数字文化遗产提供了机会。尽管前景广阔，我们也必须持续关注伦理、法律和隐私问题，并确保 AI 在医疗领域的应用始终在专业医生的指导下进行，并为患者提供最佳的治疗体验。

3.4 艺术品解读

艺术品解读是一门复杂而又富有挑战性的领域，传统的解读依赖于艺术家的说明或专业学者的研究。随着人工智能技术的快速发展，它在艺术品解读方面展现出了崭新的应用前景，为观众和艺术界带来了全新的体验与认知视角。人工智能可以帮助理解艺术品的内容、风格和情感表达。在建立大量的数据库之后，人工智能可以通过图像分析和深度学习技术，识别艺术品中的视觉特征和风格。它可以帮助分析艺术家使用的色彩、构图、笔触等元素，从而更好地理解艺术品的创作风格和表现手法。

作为全球最大的艺术博物馆之一，大都会艺术博物馆拥有近 200

万件珍贵艺术品，涵盖了 5000 年的全球艺术传承。为了更好地探索和挖掘这庞大而丰富的艺术遗产，大都会艺术博物馆一直在积极研究和应用人工智能技术加强、完善艺术品的解读和分类。因为艺术品本身非结构化的属性使其难以实现自动化地编目和标记，所以大都会艺术博物馆首次尝试对在线藏品进行主题关键字的标记是通过人工完成的。大都会艺术博物馆藏品信息总经理 Jennie Choi 表示：“整个团队仔细检查艺术作品的数码图像，然后给他们看到的東西贴上标签。这是一项相当耗费人力的工作，团队的成员对每一件艺术品都要细心检查。”可见这是项工作量极大的任务。然而随着科技的不断进步，博物馆也开始探索将人工智能技术应用于艺术品解读和分类领域。AI 技术可以对艺术品的图像进行分析和识别，辅助博物馆工作人员快速而准确地进行分类和归档，从而更好地管理和展示馆藏。大都会艺术博物馆与微软合作进行名为“艺术探索者（Art Explorer）”项目，利用人工智能技术大规模获取藏品的新信息。通过 AI 技术对艺术品收藏进行自动分类和标记，提高了艺术品管理的效率和准确性。该项目利用 Azure 搜索和认知搜索的管道，识别艺术品描述、相似艺术品和相关信息，形成丰富的元数据库，并通过 Azure 的 Web 应用实现最终应用。通过这一处理方式，用户可以通过“开放访问集合（Open Access collection）”在线浏览藏品，并从多个维度深入研究艺术品。用户可以根据特定的画种、线条绘图、描绘对象、艺术家等多种条件进行检索，实现更加个性化和多样化的艺术品发现。这些触手可及的艺术知识以及动态导航功能，让用户能够以新的方式探索和发现触动他们心灵的艺术品。



图 3-6

这幅画作是 3D 打印的，是分析伦勃朗作品数据的结果，这是由荷兰 ING 银行发起的 The Next Rembrandt 项目，利用人工智能技术分析了雷姆布兰特（Rembrandt）的大量作品，并创建出一幅新的雷姆布兰特风格的绘画。用于分析伦勃朗画作的算法在 Microsoft Azure 上运行，肖像是通过高度详细和复杂的过程创建的，历时 18 个月以上，并使用了 150 GB 的数字渲染图形。首先使用高分辨率 3D 扫描和数字文件对伦勃朗的所有 346 幅画作进行分析，然后使用机器学习对其进行升级。可以生成典型特征，并使用检测绘画中 60 多个点的算法来确定对象面部上这些点之间的距离。这个项目旨在探索人工智能对于艺术家风格的理解和模仿。

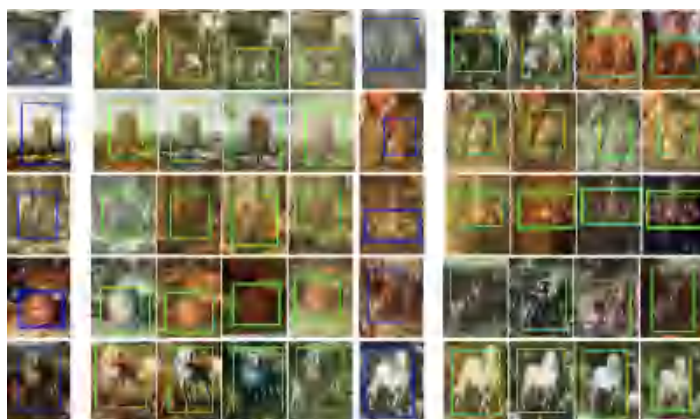


图 3-7

Elizabeth Honig 是加州大学伯克利分校的艺术史学家，她专注于利用人工智能分析文艺复兴时期艺术家的作品图像以及它们对其他艺术家创作的影响。Honig 使用了一个包含 1500 余幅勃鲁盖尔家族画作数字复制品的数据库。人工智能在这项研究中的优势在于能够从大量图像中识别出常常被人们忽视的细节。例如，对于画作中出现的风车，数据库中包含了数百张含有风车的画作，人工智能算法可以提取相似的结构和位置，进而分析作品之间的相似性和不同的绘画方式。这项研究成果于 2019 年发表在《通过空间一致的特征学习发现艺术收藏中的视觉模式》一文中详细介绍了其方法与结果。¹²⁹该项目在 Jan Brueghel 及其工作室的 1587 件艺术品的数据集中注释了 273 个接近重复的细节。

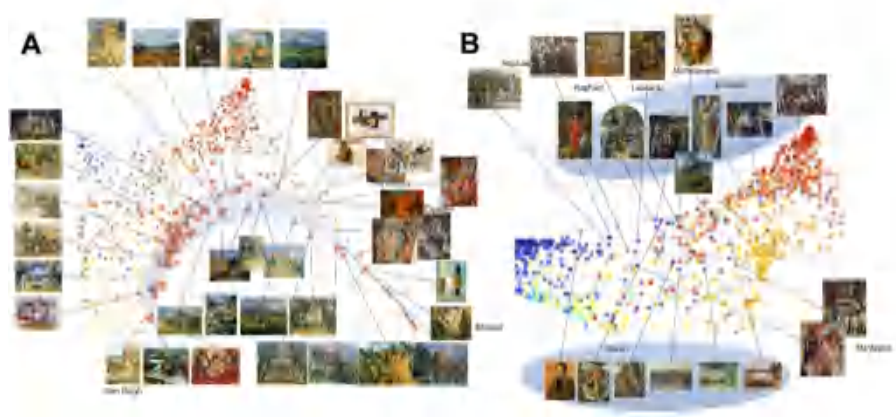


图 3-8

在新泽西州立罗格斯大学艺术与人工智能实验室，艺术史学家 Marian Mazzone 与实验室负责人 Ahmed Elgammal 合作，利用建立的数据库和模型对文艺复兴时期至流行艺术五个世纪以来的 77,000 件艺术品进行了数字分析。这项研究发现网络可以将艺术作品放置在一

¹²⁹ <https://arxiv.org/abs/1903.02678>

个平滑的时间线中，主要基于学习艺术风格标签而无需依赖任何关于创作时间、历史背景或风格之间关系的先验知识。¹³⁰该研究的成果为艺术史学家提供了一种新的探索艺术作品演变和发展的方式。通过人工智能的分析，他们可以更加准确地判断艺术家的创作时代，探索不同风格在不同时期的变化和传承，从而更全面地理解艺术历史的发展脉络。

ArtGAN 是基于艺术图片的简单生成对抗网络，使用深度学习和 **PyTorch** 实现。生成对抗网络（**GANs**）是一种生成模型的方法，利用深度学习技术，如卷积神经网络。在 `gans_training`¹³¹ 文件中，我们可以看到生成网络的改进过程。生成建模是机器学习中的一项无监督学习任务，它涉及自动发现和学习输入数据中的规律或模式，以便模型可以用于生成或输出新的案例。当然，这些新案例仍旧来自原始数据集。

在该项目中 **GANs** 是一种巧妙的训练生成模型的方法，将问题构建为带有两个子模型的监督学习问题。第一个是生成器模型，通过训练可以生成新的例子；第二个是判别器模型，通过训练可以将例子分类为真实（来自数据集）或伪造（生成的）。这两个模型一起进行训练，形成一个零和博弈的对抗训练过程，直到判别器模型大约一半的时间被愚弄，意味着生成器模型生成了逼真的例子。**GANs** 是一个令人兴奋且快速变化的领域，在生成模型方面取得了显著成就，能够在多个问题领域生成逼真的例子，尤其在图像转换任务方面，如将夏天的照片转换为冬天或白天转换为夜晚，并且能够生成以假乱真的物体、场景和人物照片，即使人类也难以分辨其真伪。

Achlioptas 和他的团队收集了一个名为 **ArtEmis** ¹³² 的新数据集，

¹³⁰ <https://arxiv.org/abs/1801.07729>

¹³¹ <https://github.com/soham2707/ARTGAN>

¹³² <https://hai.stanford.edu/news/artists-intent-ai-recognizes-emotions-visual-art>

该数据集基于 81,000 幅 WikiArt 绘画，由超过 6,500 人的 440,000 份书面回复组成，表明绘画给他们带来的感受，并包括对他们为什么选择某种情感的解释。利用这些反应，Achlioptas 和由斯坦福大学工程学教授 Leonidas Guibas 领导的团队训练了神经扬声器——以书面文字的方式做出反馈的人工智能，使计算机能够对视觉艺术产生情感反应并用语言证明这些情感的合理性。人工智能通过深度学习算法，可以快速准确地识别艺术品所属的风格。举例来说，一家名为“ArtStyle”¹³³的应用，能够通过用户上传的艺术品图像，自动识别其所属的具体风格，如抽象主义、立体派等。更为有趣的是，该应用还能进行风格对比，将两幅不同风格的艺术品融合成一幅全新的作品，让观众领略艺术家们跨越时空的创意对话。

艺术品常常承载着丰富的历史与文化背景，这使得作品在更大程度上超越了单纯的图像表现。人工智能通过大数据挖掘和知识图谱构建，可以将艺术品与历史事件、文化背景等进行关联。通过人工智能的算法，将观众带入作品创作时代的历史氛围，让观众更深入地了解作品背后的文化内涵。

可以推断出，人工智能在艺术品解读领域的应用，为观众带来了更加多样化和丰富的艺术体验。从风格识别到情感分析，从历史解读到细节观赏，这些案例展示了人工智能的巨大潜力和创新应用。随着人工智能技术的不断演进，它将继续推动艺术品解读领域的创新与发展，为我们带来更加美妙、深邃的艺术之旅。

3.5 艺术教育

在艺术教育领域内，有关 AI 的研究主要聚焦于通过多元化课程设计（如数字媒体艺术，人工智能艺术等）推广在线学习与远程教育，实现跨学科教育的融合，引入项目式学习并拓宽国际合作与交流的路

¹³³ <https://www.artstyle.ai/>

径，以达成课程改革与进步的目标。旨在为学生提供更广泛的学习选择，增强学生的跨学科技能，将理论与实践相结合，推动实践与创新并且提升学生的国际视野。

数字技术在多方面对高等教育产生了深远影响，改变了信息获取的方式，促进了在线学习和大规模开放在线课程（MOOCs）的发展，辅助了学习管理系统（LMS）的实施，提供了各类协作工具，推动了适应性学习的发展，以及支持基于数据的决策制定。此外，人工智能对教育的影响已成为教育研究领域关注的焦点。人工智能（AI）不断地改变我们的教育方式和经验，为学生提供前所未有的机会，包括访问丰富的信息资源，参与在线学习和大规模开放在线课程，利用学习管理系统管理他们的学习进度，以及使用协作工具进行团队作业。AI 和数字技术的引入，使教育实现了更多的包容性和民主化，让所有人，无论他们来自何处，都能接触到优质的教育。这些技术不仅改变了学生的学习方式，还在教师、教育者和学校之间创造了新的协作方式，提高了教育的效率和效果。借助适应性学习技术和数据分析，教育者现在能够更好地理解学生的学习需求和进度，并能针对这些需求提供个性化的学习方案。AI 的引入，特别是 AI 驱动的聊天机器人，可以提供全天候的教学支持，帮助学生解答问题，引导他们完成学习任务。

尽管人工智能（AI）在其他行业的应用广泛，但在整体教育领域其采用率相对较低。此外，关于大型语言模型（LLM）在教育中的应用研究也相对不足。尽管教师对于教育 AI 的态度普遍积极，并愿意推广 AI，但 AI 工具的整合程度和其有用性、易用性以及工具的信任度等因素对教师接受度有一定影响。最近的研究开始关注大型语言模型在学生答案自动评估，适应性反馈，以及教学内容生成等方面的应用。例如，一些人工智能系统已被用于评估化学教育学习环境中的学生答案，表明大型语言模型在评估学生答案方面有强大的潜力。此

外，基于自然语言处理（NLP）的模型还可以用于生成自动适应性反馈。

大型语言模型还可以协助生成练习题，人工智能模型可以提供各种编程任务，测试一般知识的问答对也能自动生成。语言模型也已被引入数学教育，自动生成数学应用题。值得注意的是目前 AI 较多的应用于标准化教育，在艺术教育中的应用相对较少。尽管大型语言模型能够适当地回答学生的问题，但在帮助学生解决开放性问题方面，其表现远不如人类，这凸显出需要进一步的研究。人工智能系统在美术教育用的应用较少，缺乏系统的应用策略、评估体系，以及成熟的应用案例，呼唤未来有更多的研究和实践能够针对这一领域进行探索。

3.5.1 AI 在艺术教育的应用

当考虑将人工智能（AI）应用于艺术教育时，有许多关键研究领域值得深入探讨和优化，以确保我们充分发挥这一技术的潜力，并保持教育的负责任性和道德性。以下是一些重要的研究方向：

1. 模型整合策略

首要的任务是探讨如何将大型语言模型（Large Language Model, LLM）有效地融入科技艺术的教育实践中。这包括制定切实可行的教学策略，明确定义模型的最佳使用场景，并探索如何最大程度地利用模型来优化和丰富学习体验。

2. 课程内容设计

根据教学目标，需要研究如何使用 LLM 设计和生成课程内容。这包括利用模型自动化生成艺术作品、科技创新等教学素材，或者构建基于模型的互动教育活动，以推动教学方法的创新，激发学生的学习兴趣。

3. 个性化学习体验

探索如何利用 LLM 提供个性化的学习经验。通过根据学生的学

习进度和能力差异，为每位学生提供定制的学习资源和建议，以满足他们的特殊需求。

4. 技能和素养培养

研究如何培养学生理解和使用 LLM 的能力，同时着重教育他们理解模型的局限性及其潜在的社会影响，以培养他们的批判性思维和责任感。

5. 评估和反馈机制

进行研究以评估 LLM 在教学中的效果，并根据实时反馈来调整教学策略，以确保持续改进和优化教学方法。

6. 跨文化和国际化

探索如何使用 LLM 实现科技艺术教育的国际化。这可以通过自动翻译和促进跨文化交流等方式来促进全球范围内的教育资源共享和文化交流。

7. 伦理和道德问题

反思和考察 LLM 在教学中可能遇到的伦理和道德问题，例如数据隐私、模型输出的偏见，以及模型的潜在滥用等问题，这有助于确保 AI 在教育中被健康和合规的应用。

这些研究领域将有助于我们更深入地了解 and 利用大型预言模型在科技艺术教育中的潜力，同时也能确保我们以负责任和道德的方式应用这项技术于艺术教育中。通过持续地探索和优化，我们可以更好地推动教育领域的发展，促进学生的创造力和思维能力的提升。

3.5.2 AI 对艺术教育的影响

在艺术学科中，相较于其他学科存在着明显的差异。特别是对艺术教育的影响，AI 的应用已经显著地改变了学习的本质和学生的能力。首先，教育范式可能因引入大型人工智能模型而发生转变。人工智能可以快速解答标准化问题，这或许损害学生学习的自主性和批判性思维的形成。也可能导致学生一味地追求正确答案，而非真正的知

识理解。反之，艺术与其他学科的不同之处在于它没有标准答案。艺术教育相当于一套评价体系，其注重于创造性和个人表达。鉴于美术作品通常无单一的正确解答，评价依赖创意、表现和思想。虽然人工智能可以为艺术家提供一定创意启发，然而它无法完全取代创作者的独特创造力与表达方式。

进而，在科技艺术领域，人工智能反而成为便捷、有力的辅助工具。对于编程等技术知识，ChatGPT 等工具可为学生提供指导和示例代码。然而，对于科技艺术，深刻理解编程背后的原理及创作思维仍显得重要，故人工智能角色应作为辅助学习，而非完全替代。

不论视觉艺术或科技艺术，教育目标仍聚焦于培养学生综合能力。人工智能虽能为学生提供资源，然而培养创意思维、批判性思维、解决问题的能力 and 自主学习的技能，对于学生在实际世界中具备的重要技能仍然不可或缺。

综合而言，视觉艺术领域内人工智能应用存在独特的挑战和机遇。教育中的人工智能应用必须审慎考虑技术与学生发展需求的平衡，以促进学生全面发展和批判性思维。在艺术教育中，人工智能应起辅助、启发和激发创造力的作用，而非替代创造力和个人表达的重要性。

3.6 艺术品保存与修复

近年来，人工智能技术在艺术品保存与修复方面的作用越来越显著。在视觉艺术领域，每件艺术品在完成之后都面临着损耗，比如剥落、断裂、掉色、遗失等问题，难以确保在任何时候都得到完善保管，因此需要对艺术品进行修复和维护，使其保持最初或最佳状态。

除了从业者的专业知识，新技术同样提升了专业团队的修复能力，人工智能在其中发挥着关键角色。人工智能的作用主要体现在两个方面：一方面是采取更精确和具有保护性的修复措施，详细地检测作品，

在尽可能避免破坏的情况下进行清理和复原。另一方面是对作品进行保存和维护，通过数字转化，构建电子数据库，从而有效地挽救即将失传的珍贵作品，防止未来的损坏。下文例举的案例展现了人工智能在该领域的最新运用，讨论它如何帮助从业者更好地保存和修复艺术品。

3.6.1 艺术品修复

艺术品修复不仅是对现有文物的简单维护，还是艺术品改善、保存和研究工作的重要前提。近年来，艺术品修复越来越成为一门跨学科专业，在油画、壁画、古建筑、出土器物等领域取得了巨大进展。借助人工智能技术进行作品修复，从业者能够减少对文物的物质破坏，提升效率和可重复率，弥补传统方法的不足，为艺术品的保护工作提供了新视野。

1. “夜巡行动”

在人工智能技术的帮助下，荷兰画家伦勃朗最优秀的作品之一《夜巡》（The Night Watch）得到修复。这幅绘画在 1715 年移至阿姆斯特丹市政厅的时候遭遇严重毁坏，部分画面被裁剪和丢失。2019 年，阿姆斯特丹国家博物馆（Rijksmuseum）开启了“夜巡行动”（Operation Night Watch），其中包含一项耗时多年、耗资数百万美元的修复项目，在科学家罗布·埃德曼（Rob Erdmann）带领下，依照格里特·伦登斯（Gerrit Lundens）的摹本对《夜巡》的丢失部分进行复原。

由于尺寸和风格上的差异，摹本无法简单地复制到原作上，如何将缺失的部分复原为伦勃朗的近似画风成为关键。罗布·埃德曼为此创建了三个独立的神经网络，训练人工智能将缺失的部分可视化，算法分三个步骤进行：

第一个神经网络用于识别摹本中的元素，包括人物面部、服装、手部和武器，并按照伦勃朗原作的尺寸将其复制出来，重建伦登风格

的版本。其次是针对细节纹路的还原，伦勃朗原作的数字化版本被切分成数千块作为参照，第二个创建的神经网络可以拉伸、旋转、缩短和解压缩伦登斯的摹本，使测量结果尽可能地与伦勃朗的原作相匹配。第三个神经网络用于识别色彩和阴影的差异，如同画家在研究构图和色彩之前先给画布调色。

经过数万次的迭代，人工智能高精度地再现了伦勃朗原画的已知部分，并且将摹本的四条边缘（《夜巡》遗失的部分）转换成伦勃朗风格。作品最终的展出阶段，AI复原部分被打印到画布上，再将其放置在《夜巡》原作前，与原作的距离不到一厘米，这样就能在参观者没有实际接触作品的情况下产生一种视错觉。该项目证明了现代科技在《夜巡》研究中的关键重要性，得益于人工智能，观众才能近距离领略原作的真实面貌。



图 3-9 《夜巡》左侧的细节，画作在阿姆斯特丹博物馆的人工智能技术的帮助下被扩展¹³⁴

2. 克里姆特绘画色彩复原

¹³⁴<https://www.rijksmuseum.nl/en/stories/operation-night-watch?filter=The%20missing%20pieces>

另一个通过算法复原遗失绘画的例子，是维也纳艺术家克里姆特的作品修复。2021年，谷歌艺术与文化实验室（Google Arts & Culture Lab）和维也纳美景宫博物馆（Viennas Belvedere Museum）合作，利用机器学习技术，重现了克里姆特的三幅绘画——《哲学》（Philosophy）、《医学》（Medicine）和《法学》（Jurisprudence），该系列在二战期间被纳粹烧毁，如今仅留下黑白照片。

为了重现画面的全彩效果，谷歌机器学习研究员埃米尔·沃尔纳（Emil Wallner）开发并训练了一种由三部分组成的算法。首先，该算法从谷歌艺术与文化数据库中获取了约十万张艺术图片，用于理解自然界的事物、艺术品和构图。接下来，算法专门学习了克里姆特的全彩作品，熟悉他在该创作时期偏爱的色彩和主题。最后，人工智能获得了三件作品相关部分的色彩线索，将这些数据点输入算法，从而更准确地还原画作的面貌。

人工智能对黑白照片的着色一直伴随争议，经常被认为无法精准地还原过去。针对该问题，算法同样结合了克里姆特专家弗朗茨·斯莫拉（Franz Smola）的研究成果，他梳理了各种新闻报道、展览目录和信件，寻找关于这些画作的色彩描述。例如，《法理学》中的三位女性被蛇缠绕，与克里姆特的《贝多芬浮雕》（1902年）相似，其中的女性也被标志性的金箔染成的蛇包围着。这些相似之处帮助人工智能算法构建出部分画作使用的色彩轮廓。

最终画作的某些方面改变了研究者的预期，例如人们可能会认为《哲学》中的星空带有蓝色调，但人工智能根据算法生成的颜色充满了翡翠色的雾。尽管无法实现完美的复刻原画，但是该项目仍然以较精准的方式将克里姆特的色彩风格运用到黑白照片，帮助人们更深入地了解这些杰作的历史价值。

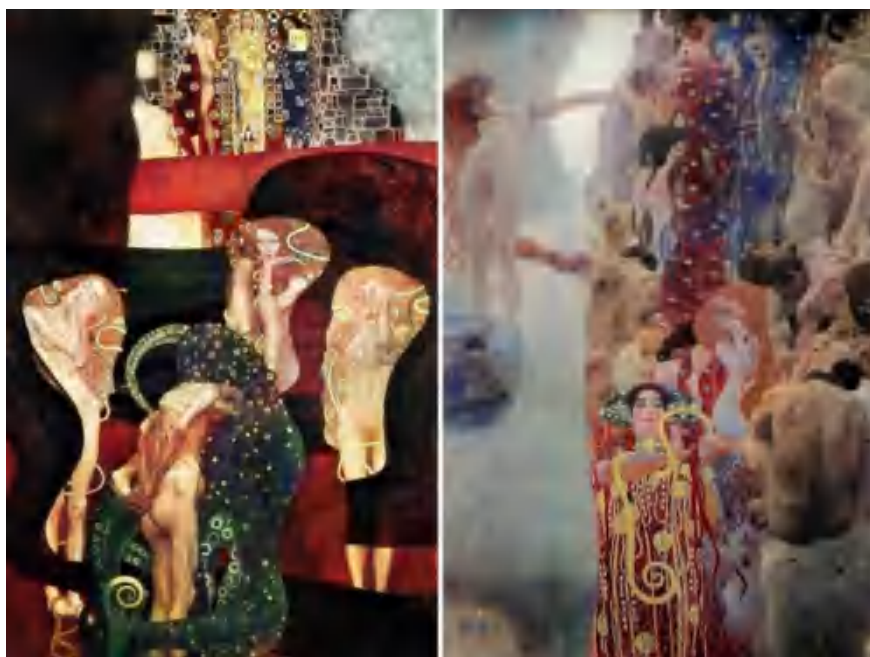


图 3-10 人工智能技术重新着色的《法理学》（左）和《医学》（右）¹³⁵

3. 箭扣长城的修缮与维护

除了绘画复原，在当前的古建筑修复中，AI 识别技术已经被用于检索损坏的部位，为修缮工作提供指导数据。以长城的修复为例，2018 年黄先锋教授代表武汉大学测绘遥感国家重点实验室携手英特尔与中国文物保护基金会，将人工智能、无人机和实景三维重建技术相结合，对长城最陡峭的通道之一——箭扣长城进行了修缮。

箭扣长城位于北京市怀柔区八道河乡境内，坐落于海拔 8 千米的悬崖峭壁，险峻地势令勘察和施工十分困难。由于年久失修，城墙的侵蚀破损严重，修护工作迫在眉睫。

英特尔（中国）研究院和武汉大学开发的深度学习算法，主要应用于破损识别和虚拟修复。在破损识别方面，针对长城的损毁和裂缝类型，研究人员首先对长城三维模型进行采样，用于训练深度学习网络，该网络采用回归和卷积相结合的架构，进行了大量数据样本训练，形成对损毁模式的识别能力。

¹³⁵ <https://www.belvedere.at/>

在虚拟修复方面，AI 在识别长城损毁部分的基础上，会对其继续数字化虚拟修复，在损毁的模型上生成 3D 的修复效果和砖墙纹理，并获得无力修缮所需的工程量的数据，作为对无力修缮的参考建议。同时，数字化修复会遵循“修旧如旧”的文物修缮原则，为长城修缮工程提供详细的位置，效果和所需工程量的估计，作为实际工程有效的参考和对照。

古建筑由于体量庞大，结构复杂，历史背景特殊，往往需要耗费大量人力和时间对建筑的类型、材料、地理环境和文化因素进行综合考虑，这为修复工作带来极大困难。借助先进技术和高质量数据，勘测工作不再需要工作人员冒生命危险实地考察，原本耗时数月的工作量在三天内完成，箭扣长城的修缮项目仅仅是一个开端，为未来人工智能技术在古建筑保护方面的应用提供了良好的技术思路。





图 3-11、图 3-12 长城缺损情况的识别与定位¹³⁶

4. 庞贝古城壁画残片修复

人工智能领域的技术进步同样在出土文物的修复方面作出贡献，很多考古出土文物，如双耳陶瓶、壁画和马赛克等等，往往以残缺不全的状态被发掘，当碎片数量非常多，人工重建几乎不可完成，因此只能长期存放在考古仓库中。

2021 年由欧盟资助的“RePAIR”（重建过去：人工智能和机器人遇见文化遗产）项目代表了该领域的最新探索。该项目由威尼斯佛斯卡里大学带领，目标是开发一种智能机器人系统，能够自主处理、匹配和物理组装大型破碎的人工制品。

最早的复原案例是对庞贝古城遗址壁画碎片的重新组装，包括“画家工作之家”（House of the Painters at Work）和“角斗士训练营”（Schola Armaturarum）的壁画残片。这些珍贵的世界文化遗产经历了公元 79 年的火山爆发、二战轰炸和建筑倒塌等破坏，目前处于残缺状态。该项目使用了名为 RePAIR 的机器人筛选壁画并将它们重新组合，极大地帮助考古工作者的重建工作。这些配有机械臂的智能机

¹³⁶ <https://v.qq.com/x/page/i0724qm9zsy.html>

机器人，能够对碎片进行扫描，提供三维数字化系统识别碎片，然后将其放入正确的位置。在扫描的同时，机械臂的精密传感器进行操作和移动，避免对文物造成轻微的损坏。

借助机器人技术、数字化和人工智能，RePAIR 项目体现了考古文物修复领域的重大突破，威尼斯大学计算机科学和人工智能教授马塞洛·佩利洛（Marcello Pelillo）表示，如果该项目成果，相关技术可以用于重建各种碎片化文物。



图 3-13 人工智能驱动的机器人试图恢复庞贝古城的壁画残片¹³⁷

5. 敦煌石窟壁画修复

在我国的艺术品保护中，利用人工智能技术对敦煌石窟壁画进行修复，成为当前图像处理和计算机视觉领域的热议话题之一。敦煌石窟作为珍贵的世界文化遗产，现保存十六国至元代的壁画共 4 万多平方米，经历千年的风吹雨打和人类活动，面临着脱落、断裂和腐坏等问题。

2021 年，腾讯与敦煌研究院进行合作，引入 AI 病害识别和沉浸

¹³⁷<http://pompeiiisites.org/en/comunicati/the-repair-project-begins-robotics-and-digitisation-at-the-service-of-archaeology/>

式远程会诊技术，协助敦煌壁画的修复工作。通过深度学习敦煌壁画病害数据，腾讯形成一整套自动识别并添加图示的算法，对壁画进行“对症下药”。技术团队多次与多家文物考古所、保护所及美术所沟通，赋予 AI 进行病害诊断的能力。同时，基于多媒体实验室的高性能编解码引擎，腾讯还提供了沉浸式的远程会诊技术，利用 4K 超清画质，全方位、无死角地展示洞窟内的景象和文物的细节，让不同地域的专家都可以实时清晰查看现场情况，共同为文物保护出谋划策。

对于 AI 而言，敦煌文物的数据量极小，共通特征极少，AI 学习难度高，对于蚕食严重的壁画，AI 的诊断更为困难。现阶段通过深度学习对壁画进行诊断还处于发展阶段，腾讯 AI 病害识别技术的研发与应用仍然面临着巨大挑战。在未来，随着人工智能的深度学习技术，在结合壁画的图像语义、历史、宗教等多维度知识方面得到提升，对敦煌壁画的数字化修复保护有望取得突破性进展。



图 3-14 AI 诊断敦煌壁画的损坏类型¹³⁸

3.6.2 艺术品保存

在人工智能时代，对数字信息的有效组织成为艺术品保存的关

¹³⁸ https://www.guancha.cn/ChanJing/2021_03_24_585159.shtml

键。在构建数字档案的过程中，相关从业者利用三维建模对艺术品的图像结构进行转化，将影像、手稿、照片等珍贵资料存入数据库，通过先进的计算机管理手段，对海量信息进行监控、分类、检索和呈现。艺术品的数字化保存并不排斥“物理”形态，而是提供了增强艺术的多种方式，能够避免艺术品信息的破坏与丢失，尽可能地保留其文化价值，与此同时也可以方便用户查阅和研究人员的深入研究。

1. “百度 AI 秦始皇兵马俑复原工程”

在人工智能时代，博物馆建设打破传统博物馆建设的限制，打造了全新的艺术品展示模式。“百度 AI 秦始皇兵马俑复原工程”代表了人工智能技术首次在世界顶级博物馆的大规模应用。2017年5月18日，百度与西安兵马俑博物馆宣布联合启动该项目，实现了破损兵马俑的“复原”以及文物信息智能化展示，背后依托于图像识别、3D 构图、全景合成、增强现实在内的多种 AI 技术。

西安兵马俑博物馆的观众通过手机百度 AR 功能，可以扫描兵马俑二号坑“平面布局图”“跪射俑灯箱”“铜车马结构图”，亲眼看到“活起来”的兵马俑等文物。同时，该项目不仅强调数字化展示，还包含文物的复原、透视、解说等等。例如，AI 能够自动复原破损的兵马俑，包括形状、色彩等等，避免了兵马俑出土后因环境湿度和保存问题产生的颜料剥落问题。再比如 AI 可以进行透视，让观众看到尚未挖掘的二号俑坑，领略 2000 年前兵马俑排兵布阵的盛景。

百度与西安兵马俑博物馆的合作才刚刚开始，双方还将展开人工智能技术领域的更多合作和探索——比如未来百度语音技术可被用于博物馆自动讲解，再比如百度人脸技术可用于博物馆检票，甚至百度无人车可用于博物馆的游客接驳，这些合作也有望被复制到更多博物馆。

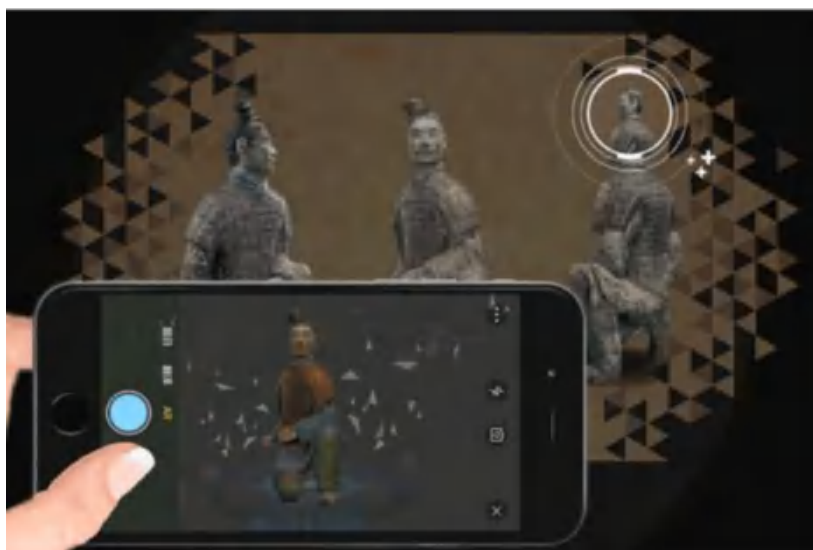


图 3-15 百度 AI 秦始皇兵马俑复原工程¹³⁹

2. “圣米歇尔山：模型的数字视角”

微软推出的“人工智能文化遗产计划”（AI for Cultural Heritage）同样旨在利用人工智能的力量，增强保护和丰富文化遗产的人员和组织的能力。通过与世界各地的非营利组织、大学和政府合作，“人工智能文化遗产计划”帮助保护我们所说的语言、我们生活的地方和我们珍视的文物。

2019 年，西雅图历史与工业博物馆（MOHAI）和微软公司联合举办的新展览“圣米歇尔山：模型的数字视角”（Mont-Saint-Michel: Digital Perspectives on the Model）是微软人工智能文化遗产计划的一部分。该展览独特地融合了 17 世纪和 21 世纪的技术，数字复原了一座坐落在法国诺曼底海岸偏远潮汐岛上的中世纪修道院。在微软人工智能和混合现实技术以及最近发布的 HoloLens 2 设备的支持下，该互动展览将游客带入风景如画的圣米歇尔山全息之旅，游客不仅可以享受物理模型的美丽，而且时会发现自己在岛上大教堂内，并近距离欣赏其哥特式设计的壮丽景色。

¹³⁹ http://home.baidu.com/home/index/news_detail/id/17691



图 3-16 2019 年“圣米歇尔山：模型的数字视角”展览现场¹⁴⁰

3. 传统服饰的虚拟转化

利用 3D 服装虚拟显示技术保存传统服饰，是近年来纺织工程领域的热门趋势，利用虚拟仿真技术对传统服饰进行虚拟仿真和还原，将二维画面转化为三维虚拟服饰，增加越剧服饰的传播方式，使人们更轻松、更细致地观察传统服饰，成为传承和保存传统服饰所传达的文化历史的一种方式。

在传统服饰方面，结合人工智能的虚拟保存已经被该领域的很多学者研究。例如学者 Liu 采用虚拟仿真技术对精选的 12 套越剧服饰进行还原。首先分析了十二种入选越剧服饰的服饰风格、面料、颜色和图案，了解越剧服饰的特点，分析了越剧风格制作结构图，利用服装 CAD 系统和虚拟仿真技术对越剧服装进行 3D 仿真恢复和展示。通过对越剧服饰的三维虚拟仿真分析，了解越剧服饰文化，提取越剧服饰风格元素，进一步进行越剧服饰元素设计，融合传统与现代，为传统戏曲服饰的保存与传播提供了替代方案，实现越剧服饰数字化展示，发展了戏曲服饰文化的保存和传播方式。

¹⁴⁰ <https://mohai.org/exhibits/mont-saint-michel/>



图 3-17 越剧服装数字分类技术线路图¹⁴¹

4. 南通蓝印花布的数据分类

在信息分类方面，基于人工智能的多层次数字化资源分类体系为艺术品的储存和管理提供了便利。以南通蓝印花布的数据分类为例，作为最早列入非物质文化遗产的传统艺术形式之一，南通蓝印花布由于缺少对蓝印花布数字化纹样的归一化编码，在传承与保存方面面临巨大的挑战。人工智能成为解决该问题的重要途径之一，2020年，于翔、沈美等研究者利用人工智能技术，对其纹样进行分类与总结，为南通蓝印花布纹样的继承和创新研究提供了新思路。

这项工作首先根据南通蓝印花布博物馆的大量资料，对南通蓝印花布博物馆的蓝印花布实物照片进行了收集，从中提取了大量的蓝印花布的原始数据资料。随后，采用数据增强技术，对采集到的纹理样本进行了扩展。最后，在深度学习的 AlexNet 模型的基础上，对基于纹理图案的文化内涵进行了分析，并将其训练集和测试集划分为 8：

¹⁴¹ Kaixuan Liu, Shunmuzi Zhou, Chun Zhu, and Zhao Lü, Virtual simulation of Yue Opera costumes and fashion design based on Yue Opera elements, Fashion and Textiles, (2022) 9:31.

2；通过对南通蓝印花布纹样数据进行对比。结果显示，在 0.002 的学习速率下，AlexNet 模型对南花布的纹样分类结果具有较高的准确度。

在蓝印花布的发展过程中，蓝印花布纹样数量、种类不同，导致分布密度不同，而且是随机变化的。对蓝印花布数字化纹样进行归一化编码处理，有效地调整了蓝印花布纹样数量分布的密度，不仅在时间效率上有显著提高，在准确性方面也取得了突破性的进展。

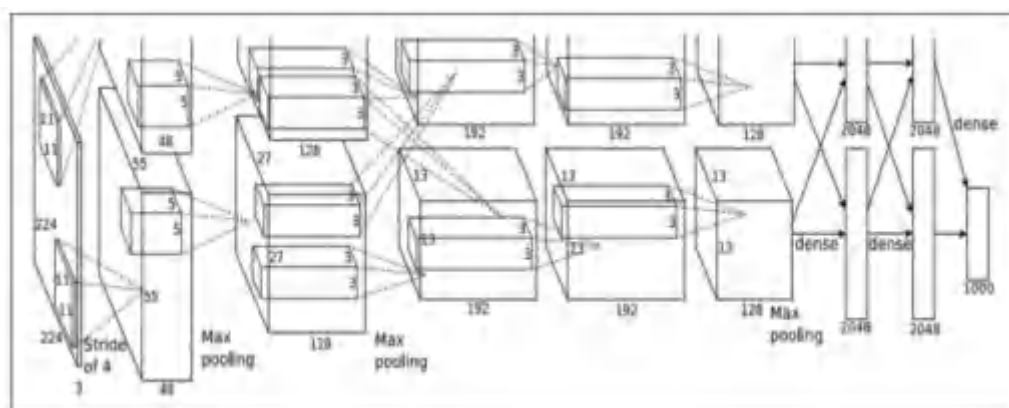


图 3-18 AlexNet 网络结构示意图

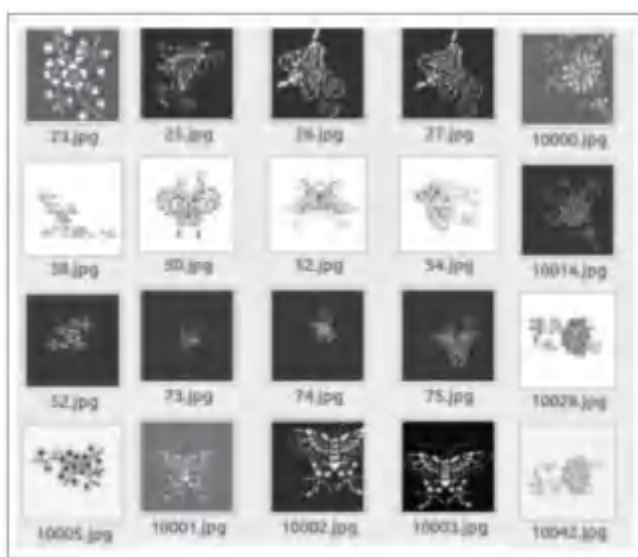


图 3-19 部分蓝印花布纹样样本数据截屏

3.7 数字人在 AI 艺术领域的意义

3.7.1 数字人

数字人（Digital Human / Meta Human），是运用数字技术创造出来的、与人类外形极为接近的数字化人物形象。狭义的数字人是信息科学与生命科学融合的产物，是利用信息科学的方法对人体在不同水平的形态和功能进行虚拟仿真。其研究过程包括四个交叉重叠的发展阶段，“可视人”，“物理人”，“生理人”，“智能人”¹⁴²，最终建立多学科和多层次的数字模型并达到对人体从微观到宏观的精确模拟。广义的数字人是指数字技术在人体解剖、物理、生理及智能各个层次，各个阶段的渗透。需要注意的是，数字人是正在发展阶段的相关领域的统称。中国在2001年和2003年两次香山科学会议后，启动了数字人领域的研究工作。¹⁴³

数字人作为一种虚拟实体，通常由计算机程序和人工智能技术创建，具备与人类相似的外貌、特征和行为。它们在人工智能和艺术生成领域扮演着重要的角色，代表着在模拟和复制人类智能和行为方面的巨大进展。数字人的创建涉及自然语言处理、计算机视觉、机器学习等领域，这些技术的发展也推动了数字人技术的进步。艺术家和创作者也利用数字人技术来创作虚拟角色、电影特效、游戏角色和虚拟世界中的角色，为创意领域带来了新的可能性。

3.7.1.1 信息与生命科学的结合

信息科学与生命科学在创造数字人类中的融合可以分为四个发

¹⁴² 钟世镇·数字人——信息与生命科学结合的新领域[J]：科技导报，2005(2)

¹⁴³ 钟世镇·数字人——信息与生命科学结合的新领域[J]：科技导报，2005(2)

展阶段。第一阶段，可见的人（Visible Human）：这一阶段涉及对人类外部特征和形态的数字化建模，包括创建具有高度逼真外貌的数字人。这通常涉及计算机图形学和三维建模技术。

第二阶段，物理的人（Physical Human）：这一阶段关注数字人的物理特性，包括模拟人体的运动、力学和生理过程。这需要生物力学和仿真技术的应用。

第三阶段，生理的人（Physiological Human）：在这一阶段，数字人的模拟将扩展到生理层面，模拟人体的器官、组织和生理功能。这需要融合信息科学和生命科学的知识，如生物医学建模和生理学。

第四阶段，智能的人（Intelligent Human）：最终阶段涉及赋予数字人人工智能和认知能力，使其能够思考、学习和适应环境。这需要深度学习、自然语言处理和智能体技术的集成。

总之，数字人代表了信息科学和生命科学领域的交叉点，展示了人工智能、计算机图形学和仿真技术的巨大潜力，同时也引发了伦理和哲学层面的问题，涉及人性、自主性和社会互动的许多方面。随着技术的不断发展，数字人的应用和研究将继续推动我们对于人类本质和数字世界的理解。

3.7.2 数字人在 AI 艺术中的意义

数字人在当今艺术生态中扮演着重要的角色，尤其是在 AI 艺术领域。它代表了人工智能技术与创意艺术的卓越融合，具有多重意义。不仅为艺术家提供了新的创作灵感和工具，还推动了艺术创作的数字化和虚拟化趋势。数字人可以作为创作的主体、模型或互动元素，丰富了艺术作品的多样性和互动性。此外，数字人的创造和使用也引发了有关人工智能、虚拟现实和伦理等方面的探讨，深化了对科技与艺术融合的思考，使艺术生态更加丰富和多元化。

首先，数字人为艺术家提供了一个独特而灵活的创作媒介。艺术

家可以通过数字人模型来探索无限的创意可能性，修改外貌、动作和环境，以创作出极具创新性的作品。这种灵感的源泉不仅推动了艺术创作的前沿，还使艺术更具多样性。其次，它代表了艺术与技术的深度融合。AI 技术的发展使得数字人能够拥有更加逼真的外观和智能互动能力，这进一步推动了艺术与科技的创新。数字人的创造和使用也驱动了计算机图形学、虚拟现实和人工智能等领域的技术进步，这对整个艺术生态产生了积极影响。

此外，数字人还提供了一个独特的互动体验。观众可以与数字人互动，参与艺术作品的创作过程，这增加了艺术品的沉浸感和吸引力。这种互动性有助于打破传统艺术与观众之间的界限，促进了更深入的艺术体验。最重要的是，数字人的兴起引发了对伦理、隐私和人工智能的探讨。它们让人们思考虚拟角色是否能够拥有情感和权利，以及人工智能与人类之间的复杂关系。这些伦理问题为艺术界带来了新的思考和讨论。

综上所述，数字人在 AI 艺术中的意义不仅在于丰富了创作工具和途径，还在于推动了技术进步、提供了更多元的互动体验，并引发了对伦理和技术与艺术的深刻思考，为当今艺术生态注入了新的活力和深度。这个领域的不断发展将为未来带来更多令人兴奋的可能性，挑战着我们对于技术和艺术的认知。

3.7.3 数字人在 AI 艺术中的作用

数字人作为创作灵感的来源，激发了艺术家的创造力，提供了丰富多彩的风格、特征和情感表达的可能性。艺术家可以将数字人视为创意的起点，将其纳入艺术品的主题或元素，从而创造出独具魅力的作品。这种情况数字人作为艺术家作品中的虚拟角色而存在。艺术家陆扬的《子宫战士》和《DOKU》都把数字人的形象设计作为表现作品内核的重要环节。其次，数字人与 AI 技术的结合使自动化创作成

为现实。通过训练神经网络模型，可以自动生成以数字人为基础的艺术作品，包括绘画、雕塑、音乐等。既展示了机器创作的潜力又将艺术创作推向了全新的领域。而且，数字人的虚拟性质也为艺术家带来了表现多样性的机会。他们可以轻松地调整数字人的外貌、动作和环境，实验不同的视觉和情感效果，拓展了艺术的创作可能性。Yona是由伊朗作曲家 Ash Koosha 最初于 2016 年创作的 AI 创作型歌手，随后与数字创作者 Isabella Winthrop 合作制作出其虚拟形象，于 2018 年亮相。在 Dazed 杂志采访中作者 Ash Koosha 表示 Yona 的引擎由一系列生成软件组成，这些软件通过复杂的过程生成句子、旋律和歌曲。它可以与人类合作表演，呈现类似于我们人类偶像的表演。

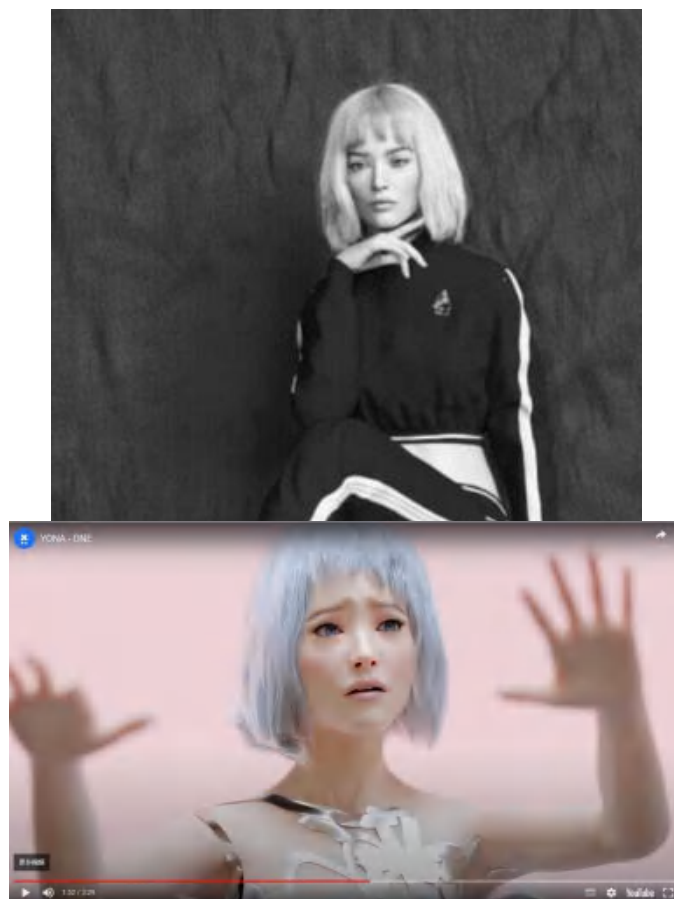


图 3-20、图 3-21 Yona.Ash Koosha.2018.¹⁴⁴

¹⁴⁴ <https://www.isntitodd.net/news/isnt-it-odd-signs-ash-koosha-yona>

此外，数字人可用于创建交互性艺术作品，增强了观众的参与感和沉浸感，使艺术更加引人入胜。当然，数字人的使用也引发了对人工智能与人类之间关系的深刻思考，其中涉及情感、创造力和自我意识等重要议题。 **LaTurbo Avedon** 是一个虚拟艺术家、策展人，其身份和存在仅限于网络和虚拟世界。她是一个由计算机生成的虚拟人物，没有真实的身体和实体。**LaTurbo Avedon** 的创造者隐藏了她的真实身份，将她视为一个虚拟化的存在，意在探索虚拟现实和数字空间对于个体和社会的影响。作为虚拟艺术家，**LaTurbo Avedon** 在虚拟世界中创作各种数字艺术作品，包括图像、视频、虚拟现实艺术等。她的作品主题涉及数字文化、身份、虚拟身体和虚拟空间的概念。

LaTurbo Avedon 的作品展示了虚拟艺术和数字艺术的潜力，以及数字化时代对于创意和身份的重新定义。她的存在和创作形式探索了虚拟现实和真实现实之间的边界，引发了对于虚拟身份、虚拟社群和虚拟世界的深入思考和讨论。在 ID 中，**LaTurbo** 探讨了社交媒体平台如何继续朝着创建认证者的方向发展。

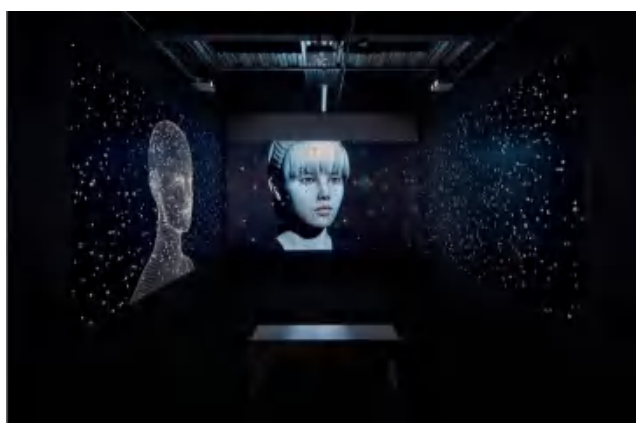


图 3-22 ID.LaTurbo Avedon.2015¹⁴⁵

AI 技术为数字人的建模、仿真、数据情感化提供了支持。表面

¹⁴⁵ <https://www.laturboavedon.com/portfolio>

上，人工智能的出现只是一种新的技术手段的显现，然而这种技术给人际关系带来了巨大冲击，逐步瓦解现有的社会体系和情感表达。人类经过数万年的演化都是与周围真实的人进行交流沟通。虽然也一直在寻找一些新的交流对象，例如动物、外星人、超自然力量等，但在大众传播中我们所面对的交流对象一直都是人。人工智能带来了传播模型的革新，数字人成为人们新的交流对象甚至像初音未来（Hatsune Miku）成为虚拟偶像被二次元爱好者所拥护。

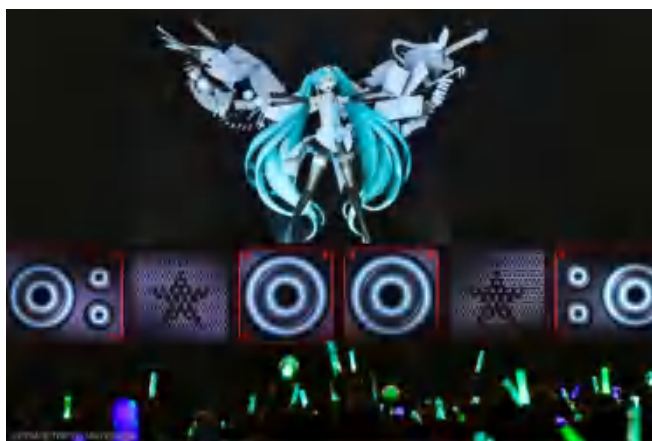


图 3-23¹⁴⁶

初音未来（Hatsune Miku）是一个虚拟偶像角色，她是由 Crypton Future Media 开发的音乐合成软件 Vocaloid 的一个声音库的代表角色。初音未来的声音由一位日本女性声优录制，然后通过 Vocaloid 软件合成音乐。她的虚拟形象通常是一个年轻的女性，蓝绿色双马尾发型是她的标志之一。她首次亮相于 2007 年，并很快成为日本和全球的流行文化现象之一。初音未来可以通过 3D 投影的方式呈现实际舞台，这种打破次元壁的演唱会，跨越了虚拟数字人与现实人类之间的互动极限。

¹⁴⁶ <https://m.weibo.cn/2462905490/4942172472213657>

韩国作为全球偶像产业最发达的国家，已经试图将韩流偶像数字化。2023年1月，由游戏公司 Netmarble 和 Kakao 联手打造的虚拟偶像团体 MAVE 在韩国出道。团队由四名成员组成，分别来自美国加州、法国巴黎、印尼雅加达和韩国济州岛，他们是 SIU、ZENA、TYRA 和 MARTY。MAVE 的首支单曲《Pandora》在 Youtube 上已经获得了超过 2000 万次播放和 44 万次点赞。

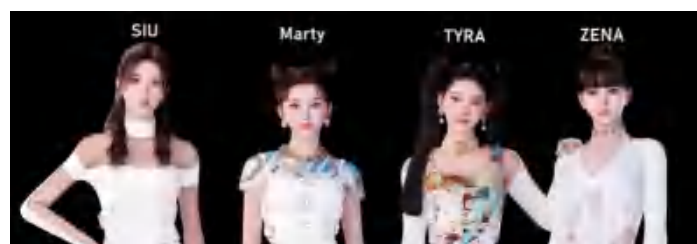


图 3-24 图 3-25¹⁴⁷

2022 年腾讯研究院发布的《数字人产业发展趋势报告》指出，Z 世代，作为互联网原生一代，已经发展成为拥有 3.42 亿用户的庞大群体。这一代的年轻人对数字人产业表现出了浓厚的兴趣和热情，可以说是其最忠诚的拥趸。数字人在这个群体中备受关注，因为它们能够与 Z 世代用户建立情感联系引发情感共鸣。当然，目前业界存在一种广泛的共识，即人工智能尚未真正具备情感和智能，其表现仅仅是基于算法和数据的处理，缺乏真正的情感体验和智能思考。虽然它借

¹⁴⁷ https://view.inews.qq.com/k/20230317A01X7F00?no-redirect=1&web_channel=wap&openApp=false

助模拟人类情感和智能行为能够实现与人类的交互，但这种交互往往是基于预设规则和模式而不是真实的情感和智能。然而这种情况确实容易导致人工智能与人类之间的假性亲密关系。由于人工智能能够回应人类的语言、模拟情感等，一些人可能会将其误认为真实的朋友、老师甚至伙伴。这种情感的投射是基于人类对互动的情感需求，但它并不代表人工智能真正具备情感和智能。类似的情况包括一些二次元爱好者过度沉迷于虚拟偶像，把她当作真实的偶像来崇拜。这种误解和假性亲密关系在一定程度上也引发了一些讨论和担忧。一方面，人工智能作为辅助工具，确实可以提供某种程度的帮助和信息，但人类应该保持对其能力和局限性的清醒认知。另一方面，过于依赖人工智能可能会减少人类之间真实的情感互动和智慧交流，影响人类社会的健康发展。在未来，随着人工智能技术的不断发展，我们可能会看到更复杂的情感模拟和智能表现，但真正的情感和智能是否能够被完全模拟仍然是一个有待深入研究和讨论的问题。同时，教育和公众意识的提升也很重要，以便人们更加清楚地理解人工智能的本质，避免过度投射情感和认知到虚拟实体上。

艺术家宋婷以自身探讨开源文化和信息科技伦理的科幻小说《算力：幻想几何学》语料为数据集，应用 **rct Chaos Box** 技术训练了 AI 虚拟人物 **Nüwa**（女娲），赋予其“意志”。值得关注的是，**Nüwa** 作为 AI 数字人与 NFT 结合的案例对二者未来的深化发展提供了范式。NFT 封存的权益使得藏家可在硬件中实时与女娲进行关于生命、意义、爱和死亡的对话。铸造为 NFT 的女娲不仅是一个动态的数字形象，也是一个可以和人类进行日常交流与分享情绪的虚拟生命——能够根据环境改变主动调整自身行为，在实现自身“意志”的同时积极地改变虚拟世界本身。**Nüwa** 被训练为具有繁育后代的能力，当然它的后代也使 NFT 可以被转移到藏家的数字钱包。NFT 的特性使得 **Nüwa** 任何层面上的交互无法脱离智能合约。这样某种程度上，艺术

家创作的 AI 数字人作为数据对象享有了版权。这可以看作 AI 艺术在捍卫版权问题上的初步尝试。



图 3-26、图 3-27 女媧：Metaverse 之母.宋婷.2021.148

数字人还在艺术教育和实验中发挥着关键作用，为艺术家和学生提供了学习和实践的平台，无需依赖外部资源或模特。最重要的是，数字人的创建推动了相关技术领域的进步，如计算机图形学、人工智能和虚拟现实，这不仅有益于艺术，还有助于其他领域的发展，如医疗、游戏开发和模拟。因此，数字人在当今艺术生态中不仅是创作的

¹⁴⁸ <https://weibo.com/ttarticle/p/show?id=2309404677048139644963>

灵感源泉，还推动了技术创新，丰富了艺术体验，引发了深刻的哲学讨论，成为艺术领域的不可或缺的一部分。

3.7.3.1 创作中的虚拟角色

数字人是艺术家在创作中虚构出的角色，主旨是服务于艺术作品。例如，陆揚作品中的“子宫战士”和“DOKU”。陆揚的作品虽然看似光怪陆离其内在所要传达的信息非常艰涩深奥，因此需要数字人的形象作为一层“糖衣”帮助观众快速进入作品的语境当中。DOKU取自《大无量寿经》：“人在世间爱欲之中，独生独死，独去独来”，是艺术家关于“数字转生”的深刻探讨，这也是DOKU面部来自陆揚面部的3D建模的原因。艺术家与FACEGOOD公司合作，利用3D面部扫描技术，将自己的面部形态与面部肌肉运动扫描收录成数据从而完成3D建模。这种形象上的设定服务于陆揚企图打破时空的尝试，DOKU作为虚拟角色是他对自己意识穿梭到不同的虚拟角色上进行数字降灵的艺术构想。

3D动画作品《子宫战士》这件艺术品探讨了性别、身份和女性身体力量的主题。展现了雌雄同体的超级英雄“子宫战士”的各种超能力。其外形看似是一名男性，但能量的源头却是子宫——女性独有的孕育生命的器官。为了切合作品对自然世界中生物繁衍的法则的质疑，数字人被塑造成一个二次元无性别的战士。形象来源于女性子宫的形状和人类展开双臂、双腿并拢站立的形态。如同文中有提及的初音未来、Yona等作为虚拟歌手，数字人具象的人形“皮”使艺术接受者能更容易对作品产生共鸣，增强作品的可读性。在艺术传播角度来说，也有利于艺术作品的二次传播。



图 3-28 DOKU.2020.陆揚¹⁴⁹



图 3-29 DOKU.2020.陆揚¹⁵⁰



图 3-30 DOKU 与李宁合作视觉影片《运动的艺术》.2021.陆揚¹⁵¹

¹⁴⁹ <http://www.modernweekly.com/star/36707>

¹⁵⁰ http://mini.vogue.com.cn/voguemini/news_111gdf1282014189.html

¹⁵¹ <https://m.weibo.cn/1782445853/4554830209354562>



图 3-31 子宫战士 1920 × 1080 (HD) 高清 3D 动画.2013.陆扬¹⁵²

艺术家经常采用拟人化的方式进行创作，将人类特征归因于人工智能。通过设计具有类人特征的数字人，艺术家企图弥合人与机器之间的差距，培养同理心和联系。人工智能的人类外形引申出了关于人类和非人类实体之间界限的讨论。它挑战了作为人意味着什么的传统观念，并触发了关于模糊这些界限的伦理和含义的讨论。

艺术家对数字人的探索大多集中在外观的设计，并且还有个有趣的现象：几乎所有数字人的形象都与人类真实的形象有所偏差。不禁让人联想到恐怖谷效应¹⁵³(Uncanny Valley)，随着数字人类图像接近高水平的真实感，有一个阈值，超过这个阈值，与真实人类外表的微小偏差可能会引发不安或反感。艺术家和设计师经常通过故意偏离真实的人类外表来应对这种现象，以避免落入恐怖谷。他们可能会将某些特征风格化或夸大，融入非人类元素，或采用艺术诠释来创造一种独特的美学，保持一定程度的熟悉感，同时避免与超现实但不完美的表现相关的不适。

通过偏离真实的人类外表，艺术家可以探索新的可能性，挑战传统的美观念，并唤起情感反应。这种有意偏离现实主义的做法允许数

¹⁵² <https://artlabgallery.com/zh-hans/exhibition/uterusman/>

¹⁵³ 是指人类在遇到与真人非常相似但不完全相同的物体或实体时所经历的不适或不安。

字人类进行创造性的表达和构建独特的身份，同时也减轻了恐怖谷效应的潜在负面影响。与此同时，数字人的探索是一个动态的、不断发展的过程。随着技术的进步和我们对人类感知的理解的加深，艺术家和研究人员继续突破数字人类设计的界限，寻求在现实主义、美学和避免恐怖谷效应之间取得平衡。

譬如 AI 歌手夏语冰是由人工智能小冰框架创造的虚拟数字人，能够绘画、作诗、唱歌。在 2022 年，曾登上央视《对话》栏目演唱《路过人间》，其与真人无异的外貌曾引起过广泛的讨论。小冰框架由微软（亚洲）互联网工程院于 2014 年 5 月正式推出，目前，它是全球承载交互量最大的完备人工智能框架之一，技术覆盖自然语言处理、计算机语音、计算机视觉及人工智能内容生成。其中，在开放域对话、多模态交互、超级自然语音及内容生成领域居于全球领先。并且该框架注重人工智能在拟合人类情商维度的发展，强调人工智能情商。虚拟人夏语冰相关技术披露，小冰神经网络渲染引擎为 X Neural Rendering, XNR，人工智能歌声生成模型是 DNN-V/X Studio 2.0 beta，使用第九代开放域对话引擎 X Conversational AI V9，人工智能绘画模型 AI Creation / Chinese Painting Model。2019 年，在邱志杰教授的训练下夏语冰顺利完成研究生毕业创作。于同年在中央美术学院美术馆举行个展《或然世界》，并陆续在中国美院和杭州万科大屋顶美术馆举办的“小冰”的跨界艺术展《小冰，“绘”有期——微软小冰&艺术跨界展》。



图 3-32 夏语冰.2022¹⁵⁴



图 3-33 或然世界.夏语冰.2019.¹⁵⁵

¹⁵⁴https://www.bilibili.com/video/BV1J44y1H7kV/?spm_id_from=333.999.0.0&vd_source=b652f5ebdda5b44ad999e0cffd4912f6

¹⁵⁵https://mp.weixin.qq.com/s?__biz=MjM5MTUzMjg2Mw==&mid=2650620862&idx=1&sn=92ac6a5540fcfd2e2457cc14b36cc898&chksm=bebdflce89ca78f84799e7588f8ba23d720eba59421b53cb260ab5cdcb612bc4dff92a65d4c1&scene=27

3.7.3.2 应用场景

数字人在艺术生态中的主要应用场景是公共领域的数字导览。许多博物馆和美术馆已经开始使用数字人艺术导览来提供更丰富和互动的观展体验。借助虚拟现实或增强现实技术进行视觉呈现，使观众可以与其在展览中进行互动。观众通过触摸屏幕或语音指令与数字人进行对话向其提问有关展品的问题，数字人根据观众的提问提供详细的解答和背景知识。这样，观众能够获取艺术作品、文物的历史、创作背景等信息，进而深化对它们的理解和欣赏。其次，数字人通过动作和表情有效地传达情感和 information。例如，当数字人介绍一件艺术品时，它可以利用肢体语言和面部表情来表达对该艺术品的喜爱或赞赏，从而增强观众的情感共鸣。再者，作为传统导览员的替代品数字人有能力给观众提供个性化的导览体验。每个数字人都有自己独特的风格和个性，观众根据自己的兴趣和偏好来选择合适的导览伙伴。往往传统导览的传播媒介具有一定局限性，通常是通过文字、图片和音频的形式来介绍展品。此时，数字人的可塑性便为导览的多维性提供了全新渠道，通过拟人化的形象成为虚拟推荐官、AI 讲解员以及参与宣传片的拍摄。

一些文化遗产和历史场所也开始使用数字人艺术导览来介绍其历史和文化，虚拟数字人作为传统文化的载体让文物和历史文化重获新生。另外，数字人艺术导览还被应用于一些公共艺术项目中，比方城市雕塑和街头艺术。并且数字人的导览功能在艺术教育和培训领域

也发挥着作用。学生通过与数字人的互动来学习艺术作品的历史、风格和技巧，培养对艺术的兴趣和理解。

那么根据以上内容可以得出，数字人在导览中的应用不仅可以优化观展体验甚至打开全新的 AI 艺术产业发展图景。推动我国公共领域的数字化趋势，为保护历史文化提供全新渠道。同时，为艺术与大众之间搭建了桥梁。

以下是一些具体的案例信息，展示了数字人艺术导览在我国公共场景中的应用：

1. 中国国家博物馆

采用了名为“艾雯雯”的数字人作为博物馆推荐官。借助三维与实景拍摄，通过虚拟现实技术与观众进行互动。他串讲上下五千年，向观众介绍博物馆的展品和历史，回答观众的问题，并提供个性化的导览体验。据了解，“艾雯雯”的形象设计依托于山东工艺美术学院对中国女性容貌审美的领悟和中国国家博物馆馆藏古代服饰的研究。依托于腾讯新一代多模态人机交互技术“艾雯雯”拥有近似真人的形象以及逼真的表情动作。她的诞生根植于中华民族文化基因和 AI 科技前沿的融合；将是广大观众在中国国家博物馆跨越时空的朋友。



图 3-34 艾雯雯.2022¹⁵⁶



图 3-35 艾雯雯.2022¹⁵⁷

2. 敦煌莫高窟

敦煌研究院打造的数字人“伽瑶”取材于敦煌壁画，原型是声音婉转如歌的神鸟“迦陵频伽”。以互动小剧场的直播形式在线上与观众普及敦煌文化与艺术作品。“伽瑶”通过人格化的内容传播敦煌文化让传统文化在与新媒体新技术的碰撞中擦出火花。

¹⁵⁶ https://www.sohu.com/a/570303204_114988

¹⁵⁷ <https://www.bilibili.com/video/BV1vN4y1j7fc/>



图 3-36 伽瑶.2022¹⁵⁸

3. 湾区博物

在宣传片《探宝觅踪》中湾区博物馆首个虚拟数字人“岭梅香”作为重要角色登场。通过实景拍摄和数字三维建模场景相结合的形式让观众了解岭南历史文化。



图 3-37 岭梅香.2022¹⁵⁹

¹⁵⁸ <http://gs.ifeng.com/c/8G5v1nEJpOP>

¹⁵⁹ <https://www.bilibili.com/video/BV1Ug4y1K7qG/>

这些案例展示了数字艺术导览在我国公共领域特别是博物馆和美术馆方面应用的可能性。它们通过虚拟现实、增强现实和人工智能等技术，为观众提供更丰富、互动和个性化的观展体验。随着技术的不断发展，数字艺术导览在公共场景中的应用还将继续扩展和创新。AI 能够为博物馆提供更为多元的应用场景，也能使博物馆、美术馆等公共文化场所以数字化的形态永久存在。

数字人的应用对艺术生态系统带来了多方面的积极影响。首先，它丰富了观展体验，通过与数字人互动，观众可以深入了解艺术品，增强了对艺术的欣赏和理解。其次，数字人作为教育工具提升了艺术教育的效果，为学生提供了更生动、互动和个性化的学习体验。此外，数字人吸引了更多观众参与艺术活动，尤其是那些无法亲临现场的观众，通过虚拟现实或增强现实技术，他们也能欣赏艺术作品。同时，数字人激发了艺术家创作的创新力，为他们提供了新的表达方式和创作工具。最后，数字人应用促进了艺术与科技的融合，鼓励艺术家和科技专家共同合作，创造更具创意和前沿性的艺术作品。综合而言，数字人的应用为艺术生态系统带来了更多活力和生机，更广泛地加强艺术与大众的连接以及促进艺术与科技的融合。

纵使数字人在公共导览中提供了便利和优势，但仍旧存在一些弊端和隐患。具体而言，使用数字人导览可能使人们过度依赖技术，降低了他们对艺术品的真实感受和欣赏能力。通过手机或平板电脑上的数字人导览，游客可能会过于专注于屏幕上的信息而忽视了艺术品本

身的细节和情感传达。其次，数字人导览可能会削弱与实际导游或志愿者互动的机会。与有生命力的导游互动可以带来更丰富的文化体验，因为他们可以回答问题、分享故事和提供个性化的解释。数字人的应用可能会削减这种互动，使参观者失去了与人交流的机会，从而减弱了社交和教育方面的体验。甚至引发就业危机——传统导览员或导游失去工作机会。

再者，数字人导览的应用将导致游客过度拍照和自拍，而不是专注于艺术品本身。这可能会干扰其他游客的体验，同时也降低了参观者的专注度，使他们错过了与艺术品亲近的机会。最后，数字人导览的应用还可能引发隐私和安全问题。在一些情况下，这些应用可能需要访问游客的个人信息，引发了隐私担忧。此外，数字人应用可能会受到黑客攻击，从而导致敏感信息泄露或导览过程中的干扰。

综上所述，数字人的应用为艺术生态系统带来了更多活力和生机，更广泛地加强艺术与大众的链接以及促进艺术与科技的融合。但也带来了一些消极影响和弊端，如降低真实体验、减弱与导游互动、分散注意力和引发隐私问题。因此，在采用这些技术时，需要仔细权衡利弊，以确保艺术生态仍然能够保持其原有的价值和魅力。与此同时，采取适当的措施来减轻这些问题，譬如关注隐私保护、提供培训以减少技术依赖性、确保安全性等方面的举措。

3.8 AI 艺术面临的问题与挑战

3.8.1 人工智能创作的发展趋势

现有关于 AI 技术的研究为数字艺术提供了基础。2014 年对抗性生成网络（GAN，2014）和可变自编码器（VAE）的提出将 AI 生成

任务的研究推进到发展高峰，英伟达的 StyleGAN 一代和二代（2018、2022）是曾经的标杆。当下正处在大模型与基于自回归模型的扩散模型迅猛发展的时代。随着大模型的参数不断提升，涌现 GPT4、Dall-E、Midjourney 等优秀的生成工具，又如 Google 的 PaLM-E、微软的 Visual-GPT 已朝着多模态（非单一生成类）的方向发展。诸多科技会议开始下设了 AI+艺术的板块，这也标志着 AI 开始更多地影响艺术与文化的发展。AI 领域多采用量化评价机制，如图像识别有 ImageNet 等各类标准化方法来衡量算法的能力，但是同为计算机视觉类任务子集的生成任务领域却由于结果评价方式的主观性而迟迟没有建立相应的评价标准。与此同时，这也是本课题试图解决的问题。

针对文化产业数字化转型的研究也较多，均指出国内外数字文化产业呈现出发展战略化、深度融合化和产业互联化的特征，表现出了强劲的增长态势和广泛渗透性。具体的研究也从针对数字出版，逐步转向对网络空间、数字经济促进文化方面的研究。随着 AI 的发展，评述其宏观社会影响和针对国内外文化产业的研究也较充分。

反思 AI 社会影响、意义和论证其伦理如数据偏见、隐私问题的论文也较多，如提示了利用 AI 进行数字艺术创作时应注意和规避的问题，同时也可以作为创作主题。近期与数字艺术结合讨论较多的是美国研究人员在探讨由 AI 生成的艺术作品的归属权问题，即使是完全由算法生成的艺术作品，也离不开人类的创造力、辛勤工作和决策。当 AI 系统取得重大成就或引发严重问题时，如何将责任归属于周围的人类是一个复杂的问题。在大模型领域，一些声音指责相关公司训练大模型数据集的版权问题，美国著作权局认定 AI 生成图片不受版权保护。这些将直接影响 AI 参与数字艺术的成果界定和转化。

在探索 AI 与艺术结合的意义方面，提出 AI 艺术的审美不足可能源于技术偏颇和缺乏可计算的“艺术真实”的基本法则。未来 AI 艺术需要具备的特点是人类主体的指导和引领。Measuring the intrinsic

dimension of objective landscapes 论述了 AI 技术对视觉艺术的影响和潜力，包括艺术分析和创作方面的任务，并强调其创新性和探索性。使用 AI 进行艺术创作可快速生成大量新颖图片、为非专业人士提供创意平台，但作品缺乏情感和主观性可能导致同质化。本课题认为这两个缺点可能会被大模型的高参数和提示词工程（**Prompt Engineering**）重定义任务所解决，这也是本节关注的重点之一。

大模型介入数字艺术创作的应用和插件在开源社区内逐步发展，与大模型的学术研究形成互补。如中国研究者提出了 **ControlNet** 结构，用于控制预训练的大型扩散模型以支持额外的输入条件。化名 **AUTOMATIC1111** 的开源社区研究者提供了可视化浏览器界面。**CIVITAI** 是提供共享预训练模型的平台，但都缺乏正确的内容引导。在大模型时代，需要建立更完整的内容评价标准，正确引导数字艺术创作者。

综上，现有研究虽然针对 AI 在数字艺术中可能出现的挑战提供了一些解决方案，但缺乏全局视野，不能作为 AI 介入数字艺术的应对策略；同时指出针对利用大模型进行数字艺术创作的研究停留在应用类，需要展开更具体的研究探讨。

3.8.2 版权与道德

3.8.2.1 AI 艺术的版权归属

AI 生成的艺术作品在版权问题上一一直是个争议焦点。由于这些作品是由机器而非传统意义上的“创作者”生成的，所以它们的版权归属成为一个复杂和模糊的问题。在美国，根据现行版权法（**United States Copyright Act**），只有“原创性作品”的“创作者”才能拥有作品的版权。因此，由于 AI 并非一个法律主体，美国版权局通常不会承认 AI 生成的艺术作品具有版权。不同国家和地区对 AI 生成作品的版权问题有不同的解释和应用。例如，一些欧洲国家正在探讨修改相关法律的可能性，以适应 AI 和其他自动化技术在艺术创作中的应用。另一种观点是，即使艺术作品是由 AI 生成的，由于背后有艺术家的参与和创意的输入，例如选择数据集、调整参数等。因此应当认为这是艺术家与 AI 的共同创作，版权应归艺术家所有。需要注意的是，当 AI 算法或模型属于某一企业或组织时，其生成的艺术作品的版权可能会更加复杂。企业可能会主张拥有所有由其 AI 系统生成的艺术作品的版权，这可能引发一系列商业和伦理纠纷。

AI 生成的艺术作品的版权归属是一个尚未解决的复杂问题，涉及多个层面的考量，包括法律、伦理和商业。不同的国家和文化背景也可能导致不同的解释和应用。因此，这一问题需要全球范围内的深入讨论和研究。在这一部分中，我们专门针对 AI 生成的艺术作品在版权方面的问题进行了深入的探讨。这些问题不仅影响着艺术家，也影响着整个艺术生态系统，包括艺术机构、收藏家，以及与艺术有关的法律和商业实体。希望这一分析能为相关的研究和讨论提供有益的思考。

3.8.2.2 从传统艺术到 AI 艺术的版权转变

随着人工智能和数字技术在艺术领域的广泛应用，我们正在见证从传统艺术到 AI 艺术在版权方面的明显转变。这种转变主要体现在

以下几个方面：第一，原创性的重新定义。在传统艺术中，原创性通常与艺术家的个人风格、技巧和观念紧密相关。但在 AI 艺术中，原创性的定义变得模糊。算法生成的艺术作品是否能被认为是“原创”的，以及在何种程度上具有原创性，成了一个争议点。第二，创作过程的复杂性。在传统艺术中，创作过程主要依赖于艺术家的手工技艺和个人直观。而在 AI 艺术中，创作过程可能包括数据收集、算法选择、模型训练等多个步骤，这增加了版权问题的复杂性。第三，多方参与。传统艺术作品通常只涉及单一的创作者，但 AI 艺术可能涉及多个参与者，包括数据科学家、软件工程师，以及使用 AI 工具的艺术师等。这种多方参与让版权归属更加复杂。第四，数据与版权。AI 艺术作品通常基于大量数据生成，这些数据可能来源于公共领域或其他艺术作品。如果使用了受版权保护的数据或元素，将涉及第三方版权问题。

从传统艺术到 AI 艺术的版权转变正在引发一系列复杂和挑战性的问题。这不仅影响艺术创作本身，也触及法律、伦理和社会经济等多个维度。因此，适应这一转变需要跨学科和多方面的合作和研究。通过对从传统艺术到 AI 艺术在版权方面的转变进行分析，我们可以看出这是一个动态和复杂的过程，需要各方共同努力以达成一种公平、合理的版权解决方案。希望这些观点能进一步推动相关的学术和实践讨论。

3.8.2.3 数据获取和隐私问题

在人工智能领域，数据获取和隐私问题是至关重要的议题。数据在训练人工智能模型中扮演重要角色，高质量的数据可以提升模型性能，但缺乏或低质量的数据可能导致模型偏见和不准确性，因此数据获取成为研究和应用的核心环节。与此同时，隐私问题涉及保护个人信息，需要遵循隐私法规和伦理准则，特别是在医疗、金融等敏感领域的人工智能应用中更为重要。

数据伦理在数据使用中起到关键作用，避免滥用数据如歧视性算法、侵犯隐私等，确保数据使用的公平性与透明性。数据收集的透明性是建立信任的基础，应告知数据所有者数据用途并取得合法授权，从而保障数据的合法合理使用。保障数据安全至关重要，防止数据泄露、黑客攻击等危害，需要采取适当的安全措施。

在全球范围内，差异化隐私和文化差异需要得到重视，以确保跨国数据处理的合规性。数据匿名化和去识别化是保护用户隐私的方法之一，降低敏感信息的风险。遵守合法合规原则，遵循国际、国家和地区的法律法规，对于维护法律责任和信誉至关重要。

综上所述，数据获取和隐私问题是人工智能领域的核心挑战，需要在技术、伦理和法律层面进行综合考量，以确保数据的安全、合法性和隐私保护，同时促进人工智能的可持续发展。

3.8.2.4 开源与专利：资源的共享与保护

开源和专利是资源共享和保护的两个方面，对于艺术创作具有重要影响。开源软件和算法的共享可以降低艺术创作的门槛，使更多人能够参与到这一领域。开源的好处在于它促进了技术的民主化和共享，使得创作者可以共同分享和改进技术，从而推动艺术创作的发展和 innovation。然而，与开源相对应的是专利和知识产权的保护。专利保护可以激励创新，鼓励创作者投入更多的时间和资源来开发新的技术和算法。专利保护可以确保创作者的权益和利益，使他们能够获得应有的回报和认可。

在开放和保护之间找到平衡是一个需要权衡的问题。我国知识产权法律体系的一项重要原则是利益平衡。过度的专利保护将会限制技术的共享和创新，阻碍了艺术创作的发展。相反，过度的开源可能导致技术的滥用和不公平竞争。因此，需要在保护知识产权的同时，也要鼓励开放和共享，以促进艺术创作的多样性和创新。

一种可能的解决方案是通过开源许可证来平衡开源和保护的需求。这些许可证可以确保技术的共享和改进，同时也保护了创作者的权益和利益。此外，公共和私人部门可以合作，制定相关政策和法规，以促进技术的共享和保护，为艺术创作提供更好的环境和支持。

开放授权机制¹⁶⁰通过许可证约定使用者的权利和义务，目的是打破一对一的低效授权。在遵循许可证条件的情况下，使用者可以自由、免费地使用和修改作品。这种开放授权机制后来扩展到文档、图片、音视频领域。知识共享许可协议（Creative Commons license，简称CC协议）为传统作品的版权授权开辟了一条新路。

开放授权的共同点包含以下内容：（1）承认著作权，要求署上原作者或著作权持有人的姓名，这是授权最基本的限制；（2）允许免费的私人使用和商业使用；（3）允许使用者修改及修改后再发布；（4）免责声明：原作者或著作权持有人不承担作品使用后的风险及产生的后果；（5）终止授权：一旦违反开放授权条件时，终止一切授权，回归传统的知识产权保护。

以主流的深度学习框架为例，目前，人工智能软件技术开发都使用开源许可证，主要有MIT许可证、BSD许可证、Apache许可证等。训练数据库也应多使用开源资源，并推动更多科学作品的开放授权，推动生成式人工智能输出数据的开放许可。

总之，开源和专利是资源共享和保护的两个方面，对于艺术创作具有重要影响。在开放和保护之间找到平衡是一个需要权衡的问题，通过开放许可证和政策合作等方式，可以促进技术的共享和保护，推动艺术创作的发展和创新。

在这一部分中，我们详细探讨了技术和资源对于数字和AI艺术的多维影响。这些问题不仅影响着艺术创作本身，也进一步复杂化了

¹⁶⁰ 源于计算机软件领域的开源许可证，是一种对世的著作权开放授权声明。

整个艺术生态系统。希望这些分析能为未来的研究和讨论提供有益的参考。

3.8.2.5 数据知识产权合法性的挑战

在人工智能领域我国已采取多项措施来保护和促进相关的知识产权发展。主要包括知识产权法律法规完善，目前已经制定了一系列法律法规来保护人工智能知识产权，包括专利法、著作权法、商标法等。以下法案和法规均涉及生成式人工智能数据训练知识产权合法性的问题。2023 年国家互联网信息办公室发布的《生成式人工智能服务管理暂行办法》，主要规定了生成式人工智能服务提供者应当依法开展预训练、优化训练等训练数据处理活动，使用具有合法来源的数据和基础模型，不得侵害他人依法享有的知识产权，涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形。《办法》旨在促进生成式人工智能健康发展和规范应用，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益。《办法》根据《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国科学技术进步法》等法律、行政法规制定。¹⁶¹除此之外，《网络安全法》用于规定网络安全的各种要求和规则。《数据安全法》主要涉及数据的安全和管理。《个人信息保护法》关注个人信息的收集、使用和保护。

¹⁶¹ 《生成式人工智能服务管理暂行办法》（国家互联网信息办公室 中华人民共和国国家发展和改革委员会 中华人民共和国教育部 中华人民共和国科学技术部 中华人民共和国工业和信息化部 中华人民共和国公安部 国家广播电视总局令 第 15 号）中央网络安全和信息化委员会办公室 中华人民共和国国家互联网信息办公室。

这些法律为人工智能相关的发明、创作和创新提供了法律保护框架。既是促进生成式人工智能健康发展的重要要求，也是防范生成式人工智能服务风险的现实需要。¹⁶²

鉴于当前我国人工智能技术尚处于初级阶段，因此在制定法律法规的同时又需要给予科技创新足够的发展空间。“美国对人工智能研发过程中的知识产权问题采取‘先产业发展后立法’模式，即在 AI 研发过程中不提及版权，而是依赖于社会舆论、企业自觉、司法和行政系统的事后回应来解决知识产权争议。”¹⁶³虽然美国的举措出于推动产业创新、防止法律与技术进步脱节等考量，但未免有些过犹不及。

这项挑战也适用于人工智能艺术领域，特别是生成式人工智能在艺术创作中的应用。与全球相比，我国的 AI 艺术发展尚不成熟。为了加速创新和发展，需要建立健康的艺术生态系统，包括培养人才、提供创新支持、鼓励投资、采用开放包容的规范原则等方面的举措。例如生成式人工智能数据来源合法性是生成式 AI 艺术合规的必要条件，但是突破与创新离不开相对宽松的法律政策环境和技术生态建设。所以如何保持二者的平衡成为构建和谐艺术生态的挑战之一。

AI 艺术中使用的模型依赖于生成式人工智能数据的数量、质量、多样性、多模态性、实时性、长期演进性、合规性以及多语言性等，因此，数据对于 AI 艺术至关重要。已知人工智能数据来源包括两部分：训练数据库和服务用户过程中产生的 AI 合成数据。训练数据库

¹⁶² 国家网信办等七部门联合公布《生成式人工智能服务管理暂行办法》 中央网络安全和信息化委员会办公室中华人民共和国国家互联网信息办公室。

¹⁶³ 张平：《生成式人工智能数据训练知识产权合法性问题探讨》，https://mp.weixin.qq.com/s/8N4-ojQvfmRnhP_HfMinVQ

涵盖广泛的数字化信息，如公共数据、网络信息、图书、对话数据集、报刊、科学论文等。目前，这部分的数据会涉及传统意义上的知识产权纠纷。在未来，当 AI 艺术受到法律保护时，相互数据学习将会引发不同的生成式 AI 艺术作品之间的知识产权问题。对于 AI 艺术家而言，这无疑增加了创作的复杂度，需要在创作前期做到周全的准备工作。

训练数据来源的法律风险分为三种类型：第一，公有领域内容，不存在著作权问题，但可能涉及人身权问题；第二，通过授权获得的内容，难以完全准确授权；第三，使用爬虫技术爬取的网络信息，受网络爬虫协议条款限制，商业产品开发可能受限。政府在制定法律政策和规范时，应考虑数据来源的多样性和法律风险，同时平衡创新发展和知识产权保护之间的关系。所以 AI 艺术的良性发展需要更明确的法律指导，以确保数据的合法使用和维护创作者的权益。当 AI 艺术家对 AI 模型进行数据训练时则需要根据著作权法预先评估每个阶段的合法性，避免在训练数据库的输入阶段和输出阶段侵犯著作权。这种谨慎的创作方式，一方面保护了知识产权的合法性，另一方面也成为艺术家创作时的限制。总之，AI 艺术创作者需要考虑如何避免侵权风险以确保他们的行为在法律框架内合法合规。同时，又要在有限的创作空间中发挥出更多的创造力。由此可见，掌握创新突破和知识产权保护之间的“度”，无论对于政府还是 AI 艺术家都并不简单。

以上所谈及的“版权与道德”问题是通常意义上讨论人工智能时必然涉及的内容。然而，版权、隐私、数据获取渠道这些并非艺术领域独有的视角。其实，在艺术界并没有对 AI 产生非常大的抵触情绪。以美国版权总局不认可人工智能生成物有版权为例，这仅仅显示出 AI 创作的独特性并不意味着全盘否定。在视觉艺术中，作品的物理实体通常是售卖的重点，而不是版权。只有当作品用于衍生品创作时，版权问题才会涌现。但一般情况下专业、正规的画廊、美术馆会事先

与艺术家签订相应的合同。相比之下，区块链技术在传统艺术中的应用则更为明确，其本质核心就是开发传统艺术品中版权的权益。可以推断出，长久以来视觉艺术并不涉及版权这件事情。那现在 AI 艺术要不要调动出艺术的这个版权意识呢？AI 艺术的发展已经引发了对创作者、技术和版权的新思考。虽然在 AI 生成的艺术中，版权和原创性的界定变得模糊。但是，在艺术领域引入版权意识仍然可以具有多重意义。艺术界随着版权意识的觉醒必然会激发出这部分能量，从而有助于保障创作者的权益，防止他们的作品被未经授权的方式使用或衍生。此外，将区块链技术等应用于 AI 艺术中，能够为创作者提供更好的版权保护和监控机制，从而构建更透明和公平的艺术生态系统。譬如，在本章第七节中所提及的艺术家宋婷的作品《女娲》便是 NFT 与 AI 艺术的初步结合。

3.8.3 艺术创作模式的转变

3.8.3.1 艺术家的角色转变

艺术家在艺术创作中的角色也发生了转变。过去，艺术家主要是创作的“生产者”，他们创造出艺术作品并将其呈现给观众。然而，现代科技的出现已经使艺术家可以更多地成为“消费者”。他们可以借助互联网和数字技术来搜集、汲取各种信息和灵感，从而在创作中融入更多元的元素。以油画创作者为例，曾经他们以“生产”一张油画为生产目的和结果。在人工智能时代，创作者则通过 midjourney 直接生成图片。这时如果单纯把生成图片作为目的与结果，那么他们将不再是生产者仅作为 midjourney 这个 AI 产品的消费者。往积极的方面想，这种转变使得艺术家能够更广泛地接触不同领域的知识和创意，从而高效地创作出更丰富多元的作品。并且数量众多的消费者在生态系统中可以看成一种“催化剂”起加快能量流动和物质循环的作用。但是需要警惕的是一个良性的艺术生态系统不能缺少生产者。生产者的角色不管是对艺术生态还是对艺术家本人职业生涯而言都至关重要。在人

工智能时代，确保创意和原创性仍然是艺术家的核心使命。虽然人工智能可以辅助创作、提供灵感，甚至进行一些自动化创作，可它并不能完全替代人类独特的情感、经验和创造力。在这个过程中，艺术家需要更多地探索如何将人工智能作为工具，而不是完全依赖它。AI 艺术家必将迎接从零开始训练一个 AI 模型的挑战。因此，在未来艺术家必须坚守生产者的身份。那些彻底沦为消费者的人，其需求和反馈将会激发生产者更多元的创意，同时成为创作和艺术产业更加繁荣的垫脚石。当然，也引申出个性化 AI 模型、AI 艺术创作门槛提高等问题，这些将在下文进行详细论述。

3.8.3.2 AI 的角色定位

AI 作为创作工具还是创作主体？AI 技术在艺术创作中的应用引发了关于它在创作过程中角色定位的讨论。一方面，AI 可以被看作是一种创作工具，协助艺术家进行创作，提供灵感、生成图像、音乐等。另一方面，一些人认为 AI 也可以成为创作的主体，即通过从训练 AI 模型，使其能够独立地生成艺术作品。但数据集的收集、算力、算法的制定与训练又离不开艺术家、技术工作者的介入。因此，关于是否存在完全独立的 AI 仍存在争议。

虽然 AI 在一些领域已经展现出令人瞩目的创作能力，但目前大多数情况下，艺术家仍然在创作过程中发挥着主导作用。AI 更多地被用作辅助工具，帮助艺术家实现创意，但其创作主体的地位仍受到诸多限制。

3.8.3.3 个性化 AI 模型在艺术创作中的价值和挑战

个性化 AI 模型是指根据用户的个性化需求和偏好进行训练和优化的人工智能模型。它能够根据用户的特定需求和上下文提供个性化的服务、建议或推荐。在艺术领域则是经过训练以模仿特定艺术家风格或创作风格的人工智能模型。它们在艺术创作中具有一定的价值，

可以帮助艺术家更快地探索不同的创作风格，从而丰富其创作的多样性。此外，个性化 AI 模型还可以用于艺术教育，帮助学生学习和理解不同艺术风格，并通过模仿实践提升他们的创作技能。

然而，个性化 AI 模型也面临一些挑战。首先，模仿艺术家的独特风格并不总是成功的，因为艺术创作涉及情感、经验和主观性，这些难以被完全捕捉和模仿。其次，过度依赖 AI 模型可能导致创作的同质化，削弱艺术家个体的创造力和创新性。最后，个性化 AI 模型的关键在于艺术家需要自己训练 AI 模型，这才能真正实现个性化定制 AI 模型。

3.8.4 技术与资源的挑战

3.8.4.1 技术门槛与资源限制

任何 AI 艺术的生产并不仅仅是一次简单的“点与画”过程，上文提到的艺术家自己训练 AI 模型。但其所面临的技术及资金挑战是相当巨大，包括数据收集和标注、算力和存储、算法和模型设计、调参和优化、持续更新和维护以及资金投入等方面。数据收集和标注需要耗费大量人力和资源，而训练过程需要高性能计算机或云计算平台的支持。设计有效的算法和模型需要深入的领域知识和专业技能。调参和优化过程耗时且需要大量计算资源。训练完成后，模型还需要持续更新和维护以适应变化的需求。此外，训练个性化 AI 模型需要大量资金投入，包括数据收集和标注、算力和存储、人力资源等方面的费用。因此，这些技术和资金挑战对于个人来说可能是一个较大的挑战。

艺术家需要熟练掌握一系列编程语言和数据科学技术，这无疑增加了进入这一领域的门槛。除此之外，高级的计算机硬件和软件也是制作这类艺术品的必需品，但其成本不菲。

总之，AI 艺术家面临的技术门槛与资源限制需要考虑到技术知识、数据收集和标注、计算资源、学习和实践、时间和精力以及资金投入等方面的因素。这些限制对艺术家的能力和资源有一定的要求，

需要综合考虑和解决。

3.8.4.2 算力与资金的影响

在艺术创作中，算力和资金的影响是显而易见的。为了运行复杂的 AI 算法和模型，需要大量的计算资源，特别是高级的 GPU 和云计算资源。这些资源的获取和使用通常需要相当的资金投入。因此，艺术家或机构需要具备足够的财力来支持这些技术需求。

这种情况将会导致一些问题的产生。首先，高昂的成本可能使得只有资金充裕的艺术家或机构能够承担这种技术的使用。这可能导致艺术界内部的贫富差距加大，资源的分配不均。同时，一些有潜力但缺乏资金的艺术家可能无法充分发挥他们的创作能力。此外，算力和资金的限制也可能限制了艺术创作的多样性和创新性。只有那些能够负担得起高昂成本的艺术家或机构才能使用这些先进的技术，而其他人可能无法进入这一领域或受到限制。这可能导致创作领域的局限性，限制了新的艺术形式和创新的出现。

因此，解决算力和资金的影响是重要的。这可以通过降低技术的成本、提供更多的资源支持、推动公共和私人部门的合作、成立 AI 艺术基金会等方式来实现。

结语与展望

1. AI 艺术赋能国家文化数字化战略

从 AI 的初步构想至今，人工智能的技术不断取得突破，特别是 2019 年后，AI 大模型的泛化求解能力迎来了巨大飞跃，使得“大模型+小模型”的结构成为 AI 产业的核心技术路径，进一步推动了 AI 产业的全球化进程。数字艺术作为一个广受认可的独立审美领域，在 AI 的崭新技术浪潮下有着更为广阔的展望。尤其是在我国党的二十大报告中，提出了建设网络强国、数字中国和实施国家文化数字化战略的指导方针，标明了 AI 与数字艺术未来发展的重要方向。

然而，现状显示，虽然 AI 技术的研究为数字艺术提供了坚实基础，但 AI 与数字艺术的深度融合研究还在初级阶段。国际对于这个领域的研究，虽然探讨了 AI 在艺术领域的伦理、权属以及审美问题，但还缺乏完整的战略性视角。同时，数字艺术的主观性使得它难以构建一个统一的评价标准，这也是我们未来需要着重研究的方向。

开源技术社区在 AI 与数字艺术结合方面有所贡献，但内容方向和评价标准缺失成了一个突出问题。未来，结合艺术学来构建 AI 时代的评价体系，将为数字艺术创作者提供更为明确的引导。跨界合作已经在多个领域证明了其价值。尤其在 AI 与艺术的结合中，我们需要更加广泛地考虑文化、社会学、心理学、美学和技术等多个学科的交融。例如，艺术家可以根据自己的创作理念引入 AI 技术，同时，技术研究者可以根据艺术家的需求进行技术改进和创新。文化价值创新也是这种跨界合作中的核心概念。我们要思考如何通过 AI 技术捕捉、理解并推广中国传统文化的核心价值。例如，通过 AI 技术深入挖掘古代文献、音乐、艺术作品中的精髓，并将这些元素与现代创作结合，从而创造出新的艺术形式和作品。

此外，对于文化和艺术机构来说，AI 技术可以被视为一个新的

窗口，通过它们可以向公众传递文化遗产和艺术价值。例如，博物馆可以利用 AI 技术进行数字化展览、虚拟导览，甚至利用虚拟现实技术为参观者带来更加沉浸式的体验。

2.人工智能赋能传统文化的探索

在 AI 与传统文化结合的过程中，还需要不断地优化和更新模型的设计与应用，以确保模型对文化的深度理解和精准创作。中国的传统文化是千年沉淀下来的，它涉及的表达、情感、意境等要求 AI 具有高度的文化敏感性。在此背景下，AI 不仅要做到工具化使用，还要深度地融入中华文化的内核。随着 AI 技术的发展，文化的传播和传承方式也将发生巨大的变革。以往，文化传承主要依靠教育、口口相传和实践的方式，但在未来，AI 可以成为文化传承的主要载体，使得传统文化在更广阔的空间和时间里得到扩散和继承。例如，AI 可以帮助将传统的民间故事、习俗和技艺转化为数字化的形式，从而让更多的人得以接触和学习。

此外，AI 也可以作为文化交流的桥梁。在全球化的背景下，文化的交流和融合已经成为一种趋势。AI 可以帮助不同文化之间的交流，使得各种文化能够更好地融合和共存。例如，AI 可以帮助将中国的传统文化介绍给其他国家，同时也可以帮助中国了解和学习其他国家的文化。AI 在文化战略中的应用将产生深远的影响。它不仅可以帮助传统文化得到更好的传承和发展，还可以促进全球文化的交流和融合。在这个过程中，人类将成为真正的受益者，可以享受到更加丰富和多元的文化生活。

3.新思路和新方法

将 AI 应用于艺术创作，亦能推动对 AI 算法的改进研究。通过在不同应用场景下对算法的表现进行研究和分析，可以逼迫人们不断反思 AI 算法的极限、原理以及它的未来发展。AI 大模型的应用将促进艺术家和科技研究人员之间的交流与合作，进一步拓宽数字化创作的

可能性。艺术创作不仅能够拓展 AI 应用场景，还能为 AI 算法的改进提供实验数据和实践基础。跨界合作能同时推动技术的发展和文化与艺术的数字化转型，进而推动中国式现代化进程。因此，与 AI 大模型的深度融合将成为文化数字化建设创新的重要方向，这也是研究 AI 在国家文化数字化建设中的重要意义所在。

我们应当进一步强调跨学科和文理贯通，将文化发展和 AI 统一在人文精神框架下，注重科技发展对人的精神和社会心理的影响进行反思，并持批判性思考。同时强调艺术思维对科技创新的激发，将横向、发散、逆向等创造性思维融入科学思考，以艺术探索的能量激发双向创新。在艺术学内重视 AI 技术发展史，通过了解技术史、科学思维和实验方法，探索未知领域。

随着时代的发展，中国正处于一个全新的历史时期，正在积极推动文化数字化建设，以适应和推进现代化进程。这一进程不仅仅是技术的革命，更是文化与技术的深度融合，注入了中国式现代化的新活力。在快速发展的数字化时代，与之同步的挑战和机遇并存。特别是在文化艺术领域，如何结合现代技术，保持传统的魅力，同时又进行创新，成了一个亟待解决的问题。为了应对这一挑战，AI 大模型技术的引入显得尤为关键。在这样的背景下，数字艺术与 AI 大模型的结合为中国的文化遗产呈现和传承提供了全新的机会。

近年来，AI 在数字艺术领域已带来革命性的变革。从传统的艺术创作方式，艺术家们依赖个人经验和技能，到现在 AI 为他们提供了强大的工具支持，这都展示了技术对文化艺术的深远影响。而这种影响不仅限于创作，也涉及文化内容的传播和消费方式，使之更为智能和个性化。但与此同时，如何确保技术与文化的和谐融合，使 AI 真正服务于文化创新和传承成了一个关键问题。书法、山水画、皮影等是中华文化的瑰宝，这些传统艺术形式承载了深厚的文化底蕴和历史记忆，但在现代社会中，它们面临着与现代生活脱节的危机。AI

技术可以为这些传统艺术注入新的活力。通过对大量文化资料的分析和学习，AI 不仅可以协助艺术家在创作上找到新的灵感，还能通过个性化推荐、智能教学等方式，使更多的人接触并欣赏到这些艺术作品，进而推动全民美育。AI 赋能传统文化的核心关键点和优势在于其深度学习和大数据处理能力。这使得 AI 可以深入挖掘和整合传统文化的丰富内容，为创作者和教育者提供智能化的工具和推荐，使艺术与文化更加生动、互动且易于传播。这种技术介入不仅可以刷新公众对传统文化的认知，提高其参与度和接受度，而且能够个性化地适应不同人群的审美和学习需求，从而更有效地推动全民美育的普及和深化。

但要实现这一目标，仅仅依赖技术是远远不够的。构建一个以大模型为核心的知识库显得至关重要。这样的知识库可以为大模型提供充足的学习资料，确保其更加深入地了解 and 掌握中华传统文化的精髓。只有当大模型真正理解了这些艺术的内涵和价值，它才能在新时代的艺术生态中起到应有的作用。建立技术研究者与文化艺术从业者之间的桥梁，确保技术与文化的深度融合，尤其是对于中国的传统文化艺术领域。此外，文化作为国家软实力的体现，利用 AI 技术整合和展示中华的传统文化，不仅可以加强这一软实力，还为 AI 技术的发展提供了丰富的素材和挑战。

此外，为了确保技术与艺术的和谐结合，我们还需要对大模型进行针对性的训练。这意味着，在模型的训练过程中，应当注重与艺术创作者的紧密合作，确保技术能够真正服务于艺术创作，而不是简单地替代它。为了在新时代的艺术生态下应用和发展，我们必须强调建设以 AI 大模型为代表的知识库，并将其应用于如书法、山水画、皮影等传统文化的传承与创新。这样的结合，不仅能够为创作者提供更多的 AI 辅助工具，帮助他们高效地创作，还能确保文化的广泛传播和普及。借助人工智能，创作者可以超越了世界各国间的文明隔阂，

谱写“各美其美、美美与共”的文明华章。社会的每一个进步，都离不开技术的推动。但更为关键的是，如何正确引导并利用这些技术，使其真正服务于民众，为社会带来持久、广泛的福祉。为此，我们将进一步积极推进与国家战略需求、国际学术前沿相呼应的美术学、设计学等学科内涵建设，致力于全面更新数字艺术人才培养模式，推进产教融合育人，并在人才培养与科研创作领域强化学科交叉融合，实现中国美术创作和教育新格局。

参考文献

- [1] 张晓欢.数字文化产业发展的趋势, 问题与对策建议.重庆理工大学学报(社会科学), 2021(02): 1-7.[期刊]
- [2] 张伟, 吴晶琦.数字文化产业新业态及发展趋势.深圳大学学报(人文社会科学版).2022(1): 60-68.[期刊]
- [3] 陈美华, 黄轩, 陈东有.我国数字出版产业的困境及对策研究.江西社会科学.2017,37(12): 88-94.[期刊]
- [4] 张显龙. 筹划网络空间战略 促进网络文化发展. 中国信息安全, 2013(8): 50-52.[期刊]
- [5] 范周. 数字经济变革中的文化产业创新与发展.《深圳大学学报》(人文社科版), 2020, 37(1): 50-56.[期刊]
- [6] 黄永林. 数字经济时代文化消费的特征与升级. 人民论坛, 2022, 9: 116-121.[期刊]
- [7] 钟义信. 人工智能: 概念·方法·机遇.科学通报, 2017, 62(22): 2473-2479.[期刊]
- [8] 陈骞. 国外“人工智能+文化”发展新看点. 上海信息化, 2021(6): 53-55.[期刊]
- [9] 付茜茜. 人工智能时代的社会文化发展: 机遇, 挑战与应对. 新疆社会科学, 2021.[期刊]
- [10] 陈永伟.超越 ChatGPT: 生成式 AI 的机遇、风险与挑战.山东大学学报(哲学社会科学版): 1-18[2023-03-13].[期刊]
- [11] 马立新, 涂少辉.AI 艺术创作机理研究.美术研究, 2022,No.204(06):82-86.DOI:10.13318/j.cnki.msyj.2022.06.016.[期刊]
- [12] 于翔.基于 AlexNet 的南通蓝印花布纹样分类.中国新通

- 信.2021(06):45-147.[期刊]
- [13] 邱志杰.讲好当代艺术的中国故事.人民论坛, 2023(14):104-106.
[期刊]
- [14] 邱志杰.科普即美育.自然科学博物馆研究, 2021,6(05):5-11[期
刊]
- [15] 邱志杰.科技艺术的概念.美术研究, 2020(06):30-32+[期刊]
- [16] 邱志杰.科技热潮与策展的责任[J].美术馆,2020(02):162-163.
[期刊]
- [17] 邱志杰.百年荏苒:实验艺术在中.美术研究,
2019(05):22-32+41-43.[期刊]
- [18] 陈抱阳,吴啸海,江上越,张文超,商亮,卢征远,叶甫纳“中
央美术学院青年艺术家国际驻地艺术实践”作品.美术研究
(06), 94-95.[期刊]
- [19] Epstein Z, Levine S, Rand D G, et al. Who gets credit for
ai-generated art? Iscience, 2020, 23(9): 101515.[期刊]
- [20] Kaixuan Liu , Shunmuzi Zhou, Chun Zhu, and Zhao Lü, Virtual
simulation of Yue Opera costumes and fashion design based on
Yue Opera elements, Fashion and Textiles, (2022) 9:31.[期刊]
- [21] Cetinic E, She J. Understanding and creating art with AI: review
and outlook. ACM Transactions on Multimedia Computing,
Communications, and Applications (TOMM), 2022, 18(2):
1-22.[期刊]
- [22] Sohl-Dickstein J., et al. Deep unsupervised learning using
nonequilibrium thermodynamics. International conference on
machine learning. PMLR, pages 2256-2265, 2015.[会议录]
- [23] Ramesh A., et al. Hierarchical text-conditional image generation
with clip latents. arXiv preprint arXiv:2204.06125, 2022.[期刊]

- [24] Rombach R., et al. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684-10695, 2022.[会议录]
- [25] Ho J., et al. Denoising diffusion probabilistic models. Advances in neural information processing systems, 2020, 33: 6840-6851.[期刊]
- [26] Blei D. M., et al. Variational inference: A review for statisticians. Journal of the American statistical Association, 2017, 112(518): 859-877.[期刊]
- [27] Dhariwal P., et al. Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 2021, 34: 8780-8794. [期刊]
- [28] Luo C. Understanding diffusion models: A unified perspective. arXiv preprint arXiv:2208.11970, 2022.[期刊]
- [29] Brown T., et al. Language models are few-shot learners. Advances in neural information processing systems, 2020, 33: 1877-1901.[期刊]
- [30] Hu J. E., et al. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.[期刊]
- [31] Liu Y., et al. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019. [期刊]
- [32] He P., et al. Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654, 2020.[期刊]
- [33] Radford A., et al. Language models are unsupervised multitask learners. OpenAI blog, 2019, 1(8): 9.[期刊]
- [34] Li C., et al. Measuring the intrinsic dimension of objective

- landscapes. arXiv preprint arXiv:1804.08838, 2018.[期刊]
- [35] Aghajanyan A., et al. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. arXiv preprint arXiv:2012.13255, 2020.[期刊]
- [36] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).[期刊]
- [37] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.[期刊]
- [38] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.[期刊]
- [39] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.[期刊]
- [40] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.[会议录]
- [41] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in neural information processing systems, 2014, 27.[期刊]
- [42] Song J., et al. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502, 2020.[期刊]
- [43] Li, J., et al. Blip: Bootstrapping language-image pre-training for

- unified vision-language understanding and generation. International Conference on Machine Learning. PMLR, 2022.[期刊]
- [44] Ruixue Liu, Baoyang Chen, Xiaoyu Guo, Yan Dai, Meng Chen, Zhijie Qiu, and Xiaodong He. From Knowledge Map to Mind Map: Artificial Imagination. Proceedings of the Second International Conference on Multimedia Information Processing and Retrieval (MIPR), pages 496–501, 2019. [会议录]
- [45] Ruixue Liu, Baoyang Chen, Meng Chen, Youzheng Wu, Zhijie Qiu, and Xiaodong He. Mappa Mundi: An Interactive Artistic Mind Map Generator with Artificial Imagination. Proceedings of The 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019. [会议录]
- [46] Ruixue Liu, Baoyang Chen, Xiaoyu Guo, Meng Chen, Zhijie Qiu, and Xiaodong He. Another AI? Artificial Imagination for Artistic Mind Map Generation. International Journal of Multimedia Data Engineering and Management (IJMDEM), Volume 10, Issue 3, Article 3, 2019. [期刊]
- [47] Shaozu Yuan, Ruixue Liu, Meng Chen, Baoyang Chen, Zhijie Qiu, and Xiaodong He. Learning to Compose Stylistic Calligraphy Artwork with Emotions. MM '21: Proceedings of the 29th ACM International Conference on Multimedia, pages 3701–3709, 2021. <https://doi.org/10.1145/3474085.3475711>. [会议录]
- [48] Shaozu Yuan, Aijun Dai, Zhiling Yan, Ruixue Liu, Meng Chen, Baoyang Chen, Zhijie Qiu, and Xiaodong He. Learning to Generate Poetic Chinese Landscape Painting with Calligraphy. Proceedings of the Thirty-First International Joint Conference on

Artificial Intelligence (IJCAI), AI and Arts - Exhibits, pages 5019-5022, 2022. <https://doi.org/10.24963/ijcai.2022/696>. [会议录]