

Rationality and Prejudice: AI as the Process of Image Generative Mechanism

理性与偏见：人工智能作为过程的图像生成机制

文/陈静

Text by Chen Jing

【摘要】 人工智能作为图像生成机制，与人类创作者一起生成了大量的数字图像。这些图像遵循了计算机的理性计算逻辑与人类艺术规则的同时，也启发了人类去重新思考关于创造性的定义。尤其是摄影作为数字图像的主要艺术形式，在快速接受新技术的同时，也为人工智能提供了大量的原始数据。这些数据所蕴含的人类社会的意识形态规训、日常生活的语言限定与偏见也进入了数字图像的再生成过程。通过图像生成来打破偏见就需要进行规则的重塑，这一方面是对于艺术创造性的重新思考，另一方面也是对数据基础设施和算法规则的重塑。

【关键词】 人工智能 图像生成 创造性 偏见 数据

图像的生成

随着基于生成大模型的人工智能技术在艺术设计方面的应用普及及其强大能力，使得“生成”也在艺术领域成为一个流行词为人所熟知。但实际上，在生成大模型之前，生成艺术（Generative arts）就已经广为流行。艺术家通过编程语言或者算法程序来生成影像或者视频，由于后台程序和生成过程的不完全可控性，作品的生成过程和结果总会有些出人意料，但这也成为生成艺术自身的特点。

生成人工智能，尤其是基于生成大模型的人工智能出现，将原本需要艺术家独立完成的后台编程过程工程化为软件操作系统，将技术门槛降低到参数的选择和试错，使得复杂的人工智能程序成为一种键盘鼠标的点击、推拉，这给很多参与操作的用户产生了一种创作是如此容易的感觉。而上一次让大家感受到创作是如此容易的，正是手机摄影。所有的数字技术都在所谓的“人性化”“所见即所得”的方面越来越娴熟，越来越贴近一种人的行为本能：只需要点击，剩下的事情都留给机器。在这个过程中，用户只需要对结果是否符合所设定的规

范做出判断，正如在李郁+刘波的作品中，人本来应该有五个手指，作品生成了四个便成为错误。而在整个创造的过程，比如深度学习的模型具体使用的是哪一个，则变得不那么重要，因为这是留给技术人员的问题。由此，图像的生成和艺术的创作是如此的轻而易举以至于其背后的算法逻辑和价值判断似乎都不那么重要了。可是，真的如此吗？

让我们从三幅画作开始分析。第一幅男子肖像画名为《下一个伦勃朗》（The Next Rembrandt），是一幅3D打印的画作。由荷兰的ING银行和微软（Microsoft）创作团队与来自代尔夫特理工大学（TU Delft）、莫瑞泰斯皇家美术馆和伦勃朗故居博物馆的咨询团队共同创作，旨在沟通艺术与算法、数据与人类设计、技术与情感之间的关系。“我们想要激发有关数据和使用数据进行创新的讨论”^[1]。他们对346幅伦勃朗的画作进行了高清像素的3D扫描，积累了15GB的数据，然后用深度学习（deep learning）对数据进行了分析，得到了一些数据特征。比如画作中49%的主题是女性，51%是男性。其次，团队就伦勃朗主要创作时期（1632—1642年）画作中的人物

年龄、头的朝向和面部毛发进行了分析，结果发现了一个最具有代表性的主题；“一位年龄在30—40岁，面部蓄须，身着黑衣、黑帽和白色领子，面向右边的高加索男子肖像”^[2]。同时，为了凸显伦勃朗画作的特色，团队特意关注了画作中明暗对比度的问题。最终这些分析过程和结果都被整合到了一套软件系统之中，这套系统通过伦勃朗对几何、构图和颜料的使用来理解伦勃朗。最后的输出则是通过3D打印来完成。

第二幅肖像画则更加出名。这幅名为《埃德蒙·德·贝拉米肖像》（Portrait of Edmond Belamy）的画在2018年10月23—24日的佳士得拍卖中以43.25万美元的价格成交，创造了第一幅人工智能画作的拍卖纪录。这幅画作的提供者是来自法国的Obvious组织。三位年轻人通过生成对抗网络（Generative Adversarial Networks，简称GAN）训练了14世纪至20世纪之间的1.5万张肖像画的数据集，并经过生成对抗网络中的生成器（Generator）和鉴别器（Discriminator）的处理，最终形成了这样一幅似乎具有格伦·布朗（Glenn Brown）风格的画作。事实上，他们生成的不仅仅是这样一幅画作，而是一个贝拉米家族的肖像画（La famille de Belamy）。与此同时，他们在这个过程中也试图探讨“什么是艺术”：他们邀请了拥有血肉之躯的人对人类创作的画作和机器创作的画作进行了辨识，然而结果却是“两类画作的差别不大，且有些人认为机器画作给了他们更多启发”^[3]。尽管对于这点，我个人持保留意见——并不是所有人都具有艺术品的鉴赏能力且每个人的审美体验也各不相同，但必须要看到的是，机器创作的画作已经具备了一般肖像画的基本特征。Obvious组织解释道：“我们尝试了裸体和风景画，我们也试图喂给算法那些著名画家的作品。但我们发现肖像画是最能证明我们的观点：算法能够仿真

创意性”。

第三幅画作较前两幅相比，则显得不那么出名了。它的作者是当年只有20岁的罗比·巴拉特（Robbie Barrat）。他也使用来自维基艺术（WikiArt）的上千幅裸体肖像画训练了一个基于GANs的神经网络来生成了多幅裸体艺术画。然而罗比本人却并不认可得到的结果。相反地，他认为这件作品所呈现出的结果证明机器在学习的过程中失败了，并没有学到裸体绘画中所有的正确的属性，相反却掉入了局部极小值（A local minimum），从而产生出超现实的肉团。因此罗比提问：“这难道是机器看人的方式吗？”^[4]

人工智能的“创造性”与算法逻辑问题

以上三幅作品尽管并非用生成大模型完成的，但很好地诠释了计算机在图像生成方面的算法逻辑，同时也能启发我们（人类）对于创造性的反思。这三幅使用了基于训练数据的算法分析在“生成”艺术图像，但得出了三种不同的结果，其体现的是不同的算法逻辑。如果借用深度学习的比喻，三幅作品恰恰体现的是从监督学习（Supervised learning）、半监督学习（Semi-supervised learning）和无监督学习（Unsupervised learning）的过程：第一幅作品是根据训练数据集去学习伦勃朗的绘画模型，并依据此模型推测一种新的实例。输入物件（向量）和预期输出之间具有比较强的因果关系，通过对回归分析中的变量的控制（伦勃朗绘画特征的提取、明暗对比以及油画画面高度扫描数据）等，使得输出结果以“合理”的方式实现一般化；第二幅作品虽然用了无监督学习的对抗神经网络技术，但因为在训练数据集的选择、数据结果的判断和输出结果的选择方面加入了人工干预的部分，因此实际上呈现的是一种有选择的、尽可能符合预期的结果；第三幅作品则因为从数据集的选择、算法实施的过程以及最终的结果选择方面都尽量地减少了人工的主动及主观干预，因此更多地呈现了一种“无监督”的过程，而结果的预期满足度也是最差的。如若从艺术创作的角度而言，第一幅作品虽然高度符合了我们对于伦勃朗画作的期待，但因为其生成的方式导致其呈现了一种均质化的平庸，也就是说，我们只能认为它符合了伦勃朗画作的特点，但很难说这在伦勃朗风格的作品中是一幅杰作；第二幅作品虽然呈现了机器的一些特性，但因为其呈现效果经过了人为的导向干预，因此介乎于了一种似是而非的中间境地，而更多地依赖其与人工作品的相似度博得与“艺术品”之间的关联性；第三幅作品尽管在相似度上是失败的，但因为它是更多地体现的是一种机器的“自主性”，因此显得具有了一种出乎意料的趣味性或者令人不适的反差性。而这种出乎意料的趣味性或令人不适的反差性更多地指向了一种创造性的可能。

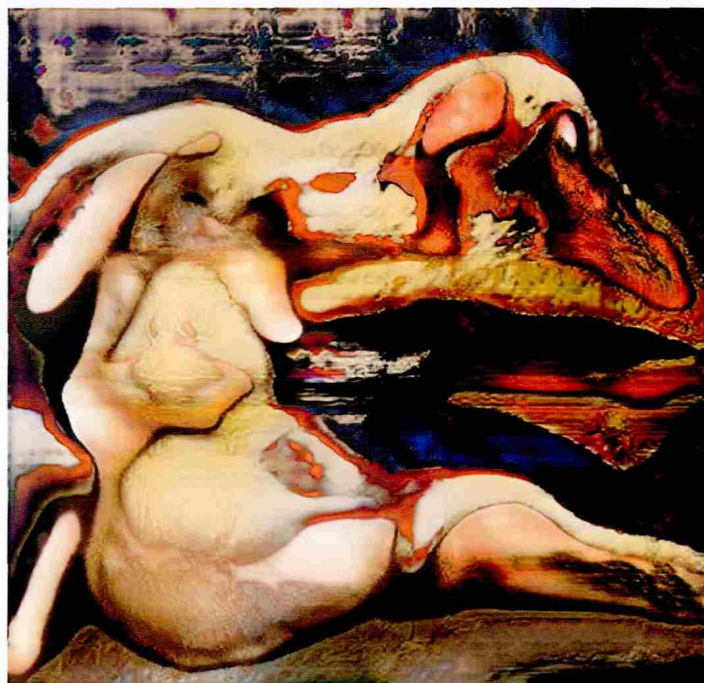
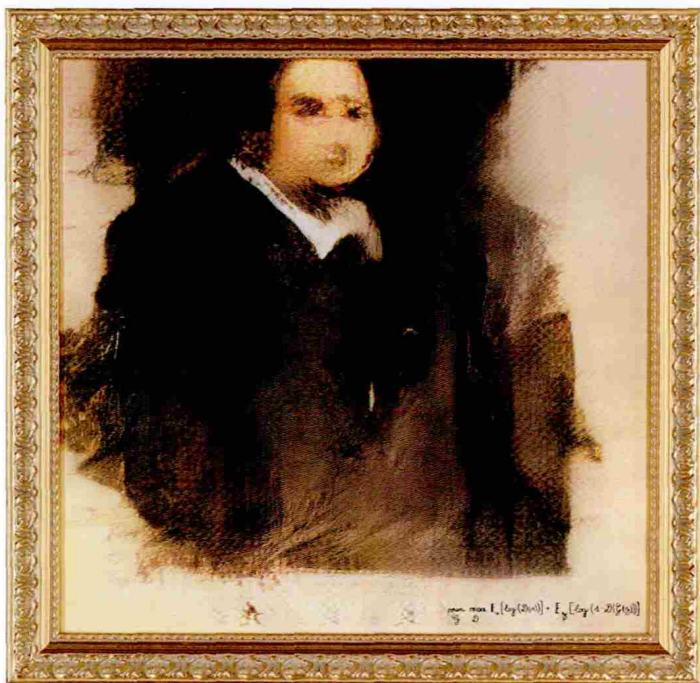
认知哲学家玛格丽特·博登（Margaret A. Boden）在《创造心灵：神话与机制》一书中以库斯勒（Arthur Koestler）的“隐没的线索”为起点，提出借助人工智能以及对人工智能的研究和理解将帮我们认知“创造性”这一人类行为。博登指出，洛夫莱斯夫人问题引发了关于创造性的四个问题：一、计算观念可以帮助我们人类的创造性是如何可能的？二、计算机（现在或者将来）能够做至少看起来像（Appear to be）创造性的事情；三、计算机看起来能够识别（Recognize）创造性；四、计算机是否能够真的（really）有创造性^[5]。她认为在前三个问题上，目前可以看到比较肯定的答案，但第四点却似乎为时过早。与此同时，她将创造性分为了“心理学的”创造性（Psychological Creativity）与“历史的”创造性（Historical Creativity），前者更多的是基于个体的人类的能力，强调创造行为和活动是



> “下一个伦勃朗”
项目画作，2016年

可能的；而后者则更多的是在历史尺度下的创造性评估，强调创造产出是被识别出来的。两者有着密切关系，但并无时间序列上的先后。基于这种模式，在人工智能与创造性的问题上，博登认为人工智能概念提供了从心理学的角度去讨论创造性的新路径，是因为我们可以通过人工智能去建构（或者测试）关于涉及思考（创造性）的结构和过程，换句话说，我们可以借助人工智能去理解人类心灵的创造性。这个过程则包括了合并（Combination）、探索（Exploration）、转化（Transformation）。

合并是指通过熟悉寻常的想法，进行不同寻常的合并，这种合并往往是无意识发生的，探索是指那些对某个群体已经熟悉的思考方式，最后的转化则更多的是在那些之前从未想过的事情，是其心灵中的概念空间里从来没有过的东西，甚至需要经过一些激烈变化。在这个过程中，心理学的创造性和历史的创造性在个体和群体的两个层面共同发生作用，个体心理层面的创造性放在一个群里的历史层面去考察，就会发现其实并非具有历史创造性的价值。这一点在人工智能生成艺术的过程中同样适用。当用户在利用像Midjourney这样的生成平台进行图像生成创作的时候，其实更多的是在个体的心理层面进行创造，其突破的是个人自身的局限，比如不会画画的也能生成画作，或者是只会某种风格的可以尝试多种风格，等等。而当将艺术创作放在艺术创作的群体以及艺术史的脉络中，这种个体的创造性就很难凸显出来。这是因为人工智能的“生成”实际上是一种理性算法的过程，它之所以被认为是艺术创造，只不过是因为它处理的是“艺术数据”，然后在生成的过程中遵循了人所规定的“艺术的规则”去生成，或者说选择生成符合人的审美观和艺术观的图像。在这个过程中，人工智能的生成模型不过是在寻找与人的创造性逻辑相符合的结果。这个寻找的过程体现的是理性算法对数据的处理以及人对生成结果的判断和辨识。因此，整个创作过程都是在对已有艺术作品的学习甚至模拟（Simulation）。正因如此，上文罗比所提的问题“这是机器看人的方式吗？”才具有启发意义。我们是否在利用算法系统生成图像的同时，考虑过那些被认定不符合的、废弃的、错误的图像？或许那些才是突破了人类创造性理



< 埃德蒙德·贝拉米肖像, 2018年 Obvious 小组 由 GAN 生成
> AI 生成裸像 2019年 罗比·巴拉特

解的、具有非人“创造性”的作品。如同本期专题中的艺术家刘诗园在《悬针》系列作品中提到的，“那些所谓的拍摄失败的照片为什么是失败的？为什么这张照片通过某一个角度拍，我们会觉得完美，而这张只是画面稍许不同但评价却大不一样？”也许所谓的“失败”是因为我们的观念只能接受一种规则。只不过放在人工智能生成的时代，艺术的规则与算法的规则实现了同谋。因此，要突破人类创造性的局限，重新塑造人工智能时代的艺术规则，就要回到生成过程之中，甚至回到生成之前，从数据开始。

共谋的偏见

人工智能的生成模型，无论是大模型还是在之前的深度学习神经网络，都高度依赖于训练数据库。像ImageNet这样的大型图像数据库，是基于团队自身和亚马逊众包平台来使用日常语言标签对图像进行语义标注。而看似“自然”“客观”的日常语言的逻辑遵循的也是打标人的语言习惯，受其主体意识、教育背景、表达能力和思维方式的影响，同时也会将个体的偏见带入标签中。比如本

期专题中特雷弗·帕格伦所使用的“ImageNet人类数据集中，一个穿着比基尼的女人被标注为荡妇，中年亚裔女性被标注为女仆”^[6]。但这种偏见在后期的算法训练中是看不到的，因为隐藏在算法的背后而被隐蔽起来了。而算法的生成过程是无法直接看到这种偏见和歧视的认知的，只有当人对结果进行经验性检验的时候，我们才可能发现这种偏见。然而发现的前提是人具有足够的经验去识别这种偏见。

有意思的是，现在流行的通用型大模型用其背后的大规模语料和耗费巨大的算力使得用户在对其能力信服膜拜的同时，放弃了对其背后数据的追问和偏见的质疑。而提示词的有效选择实际上是在对社会规则化术语的使用训练。比如，如何选择能够让模型发挥作用，生成符合预期效果的图，就要反复去选择可以被模型“识别”的提示词。而这种提示词如果并没有被预先作为语义标签被模型学习到，那结果往往是不对的。又或者当提示词这样的技术过程被遮蔽或者隐藏起来时，我们从技术角度对图像的判断就失效了，图像本身也就回归了传统艺术评价的范围之内。例如在李郁+刘波《2012 新

闻肖像》的这组作品里，2006—2011年的新闻文字成为提示词，生成了一系列“新闻图片”。这些图片看似真实，但因为取消了历史语境和时间性，也就丧失了新闻摄影对内容真实性的要求。但有意思的是，艺术家从艺术创作的角度提出：“对我们来说，或许用AI生成的图像和用相机拍下的图像并没有区别，AI生成工具就像一台超级相机，只是唯一的问题是：同样是稍纵即逝，随机生成的角色，和我们在街头偶遇的路人到底有什么不同？”确实，在艺术真实的层面，现实真实性就不再那么重要，而重要的是技术理性逻辑所实现的艺术真实的内在自洽。这些作品在图像的层面能够达到艺术评价的标准即可，而其究竟是通过相机拍摄还是AI生成，其实就不重要了。那么应该如何评价这样的作品？换句话说，提示词是否也应该作为作品的一部分出现？当遮蔽了艺术生成的技术环节回归传统评价系统的图像，技术生成的意义在哪里？本次专题中特雷弗·帕格伦也谈到了人工智能生成中语言的有限性以及机器训练对人的意

中国摄影
Chinese
Photography

识训练的规定限制性。尤其是如果尝试过特定风格训练的人或许都有过人工智能生成图像往往图不达意的感受。比如我们用stable diffusion这样的模型就会发现想要生成具有中国风格的版画就目前的实践操作来看往往是不可能的。

在这里，最根本的原因就是其前期训练语料中中国风格的版画比较稀缺，因此在没有使用特定数据集进行定制训练的情况下，直接生成特定风格的图片非常困难。而特定数据集的数量和质量又会对算法训练效果产生极大影响。目前，互联网上最大数量的图片是照片，无论是基于现实场景拍摄的，还是后期合成或者生成的，基于现实场景的图像已经成为图像训练的最大也是最基础的来源。因此，基于训练数据集的人工智能算法也是在显示场景图像上更为有效。而这也造成了在其他图像类型上，通用性算法的能力大大下降。摄影也因此成为人工智能生成图像最广为流传、接受度最高并且冲击最大的艺术领域之一。甚至哪怕其在内容或者风格上不准确，我们会因为其秒速生成的能力和多种选择而容忍了其的不准确，并期待着后续迭代更新中的能力提升。

打破偏见与规则重塑

艺术创作是否需要在意数据以及算法的偏见？在这点上，我想回溯到图灵最早的提问。图灵在那篇著名的《计算机器与智能》（Computing Machinery and Intelligence）中就曾对机器是否能写十四行诗或者创作协奏曲的问题作出回答。他以一个模拟游戏的实验为例指出，只要机器在对话过程中能够回答出令人满意的答案且持续进行，就可以认为这是一种思维过程。稍后，他在回应洛夫莱斯夫人（Lovelace）有关“（巴比奇）分析机只能做那些我们知道怎样命令它去执行的事情”观点时，对那种“机器永远不会”使我们出乎意料”的观点进行了直截了当的反驳：“机器让我出乎意料的时候非常之多”^[7]。他随即还指出，“对某一种食物作出出乎意料的评价，同样需要创造性心理活动（Creative mental act），无论这种出乎意料的事是源自一个人、一本书、一台机器，或者是其他什么东西。”之后他还写下了这样一段话：“我相信，机器不可能让出乎意料的观点起因于哲学家和数学家们特别容易持有的谬见。它假定：一旦事实呈现于心灵，这个事实的全部后果就会与之同时涌入人的心灵之中。”在这个意义上，数据和算法的偏见实际上也是阻碍了新的创造性结果的出现。我们过于强调和依赖算法的理性和客观，同时又以人类固有的偏见和经验去忽略可能出现的“出乎意料”（想想我们在训练过程中为了追求“像”所放弃的那些“残次品”）。

因此，人工智能艺术创作的可能或许就在于以艺术真实的方式去创作未知的真实可能。而这种未知的真实可能或许是技术本身所具有的，比如技术真实性所带来的不符合人类规则的非人类创造产物，或者是我们依据某种算法所产生的未来真实、历史真实。比如宋兮在《图像引擎》中的“拉法拜人”族群“繁衍”和贺肖华在《失踪者》中根据老照片进行的模拟影像实验。这些都是在既定预判之内的“出人意料”，同时又又在一定程度上符合了艺术真实和技术真实的规则。

然而，考虑到艺术数据集的数量和中国特有的文化语境，我们更需要强调具有地区文化的数据公开，确保数据集的多样性和代表性，这包括涵盖不同地区、文化背景和社会群体的数据。其次是透明和负责任的算法设计，这能确保算法的决策过程能够被理解和审查。当下通用性生成模型大部分的数据来源于互联网，因此开放在互联网上的数据决定了基于这些数据训练算法的能力有效性。这也是为什么我们在训练特定历史时期、特定风格的图像时会发现算法的

能力是不够的、有限的。大部分历史图像、艺术图像，都是通过对文化遗产物理实体进行数字化得来的，而这些数字图像和这些文化遗产一起被版权保护了起来。虽然世界上像故宫博物院、大都会博物馆、大英博物馆等多家博物馆机构都已经开放大量数据供公众免费试用，但这些数据并非都是高质量，或是打标数据，数量上也远少于前面提到的照片数据。因此基于这些数据的训练成本很高，但应用场景又是有限的，就造成了过去在研究界或者产业界进行针对性数据集开放和算法训练的比较少。近年来，这个问题在学术界已经逐渐受到了重视，数字人文领域的兴起，特别是在数字遗产和数字艺术史领域出现了大量相关的成果，都可以看到是对这种数据基础设施建设需求的回应。算法公开透明化问题，特别是对于公开数据和技术开源问题方面的讨论声势日渐兴起，也许随着编程在青年一代的普及，AI平权时代可以更快到来。

本文为“基于数字技术的艺术方法论研究”，2021年度国家社科基金艺术学一般项目（项目编号21BA26）。

陈静，南京大学艺术学院副教授、硕士生导师。主要研究兴趣为数字人文、数字艺术与数字遗产。

注释：

- [1] The Next Rembrandt. (2016 - 04 - 13)[2024 - 05 - 23]<https://news.microsoft.com/europe/features/next-rembrandt/>.
- [2] <https://www.nextrembrandt.com/>.Accessed:2024.5.14.
- [3] Is artificial intelligence set to become art's next medium? (2018 - 12 - 12) [2024 - 05 - 23]
<https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>.
- [4] AI Generated Nude Portraits,<https://robbiebarrat.github.io/archive/oth/nude.html>. Accessed:2024.5.26.
- [5] Margaret A. Boden, The Creative Mind: Myths and Mechanisms[M].Routledge, 2004:16-17.
- [6] AI 社会学 | 算法标注：我是“女仆”，你是什么？(2020 - 07 - 23) [2024 - 05 - 26]
https://www.thepaper.cn/newsDetail_forward_8402829.
- [7] (英)玛格丽特·A.博登 编著 刘西瑞 王汉琦 译.人工智能哲学[M].上海：上海译文出版社，2006：62-63.