

電影藝術



Film Art

ISSN 0257-0181

CN 11-1528/J

技术奇点与影像未来

作为后电影的 Sora 影像
姜宇辉

人工智能与灵韵重现
彭 锋

《白蛇：浮生》：传统题材的
现代动画讲述
——陈健喜、李佳锴访谈
曹小卉

《异人之下》：跨界叙事融
合与类型及技术创新
——乌尔善访谈
梁君健

从“镜头调度”到景深美学：
再论爱森斯坦的理论演进
黄小轩

中文社会科学引文索引
CSSCI 来源期刊



总第 418 期
1956 年 10 月创刊
单月 5 日出版
2024 年第 5 期

目次

特别策划 · 技术奇点 与影像未来		
003 作为后电影的 Sora 影像 姜宇辉	044 技术奇点及其限度 ——精神分析对话“后人类状况” 戴宇辰	072 《抓娃娃》：消费文化语境下的 喜剧生产与意义表达 于 帆
影艺观察		
012 人工智能与灵韵重现 彭 锋	053 新科技伦理：电影与第四次 工业革命 张建珍 吴海清	076 《从 21 世纪安全撤离》：元现 代主义科幻与非线性青春叙事 王 昕
电影批评		
019 黑栈、算法与 AIGC：行星智能 时代的电影本体论 韩晓强	061 《逆行人生》：从“迭代现实” 走向“算法现实主义” 李道新	访谈录
028 人工智能影像生成的模型诗学 朱恬骅	065 《白蛇：浮生》：越界之爱与 文明涵化 白惠元	080 《白蛇：浮生》：传统题材的现 代动画讲述 ——陈健喜、李佳锴访谈 曹小卉
035 Sora 与时间第三综合 谢周浦	069 《解密》：现实与梦境的双重生活 陆绍阳	088 《异人之下》：跨界叙事融合与 类型及技术创新 ——乌尔善访谈 梁君健
		095 《抓娃娃》：幽默感并不是公式 —— 闫非、彭大魔访谈 何 亮

理论研究

099

从“镜头调度”到景深美学：
再论爱森斯坦的理论演进

黄小轩

109

影像的“非思之思”：戈达尔的
关系思维

李 双

影史影人

116

于敏与电影：言讷显风骨，身
躬写春秋

李 镇

124

中国早期电影服装的形式探索
与观念构建

田乐乐

视与听

132

城镇的“失落”与情感表达
——电影《狗阵》的美术设计

霍廷霄 | 李 畅

国际视野

139

GenAI 影响下美国好莱坞影视
劳动力市场的变化及挑战

魏梦雪

147

大萧条与罗斯福新政：早期美
国演员工会（1933—1945）

陈建州

153

“面纱与面纱化”：1979—
2000 年伊朗电影中的视觉文
化与民族语法

周 涌 | 郭培振

中文社会科学引文索引（CSSCI）来源期刊
RCCSE 中国权威核心期刊（A+）| 全国中文核心期刊
中国人文社科期刊 AMI 综合评价核心期刊
中国科技核心期刊（社会科学卷）
2020 中国国际影响力优秀学术期刊

编委会

以姓名笔画为序

顾问 / 郑洞天、仲呈祥、黄会林、谢飞

主任 / 邓光辉

委员 / 丁亚平、王一川、王海洲、尹鸿、厉震林、石川、吕新雨、
孙立军、李洋、李道新、杨远婴、宋智勤、张真 [美国]、
张宏、张同道、张会军、张阿利、张清华、张颐武、
陆弘石、陆绍阳、陈旭光、陈犀禾、周星、周安华、
郝戎、胡智锋、钟大丰、侯光明、饶曙光、贺桂梅、聂伟、
贾磊磊、路海波、谭政、潘若简、戴锦华

编辑部

主编 / 谭政

本期统筹 / 刘起

责任编辑 / 刘起、张雨蒙

特约编辑 / 徐枫、类成云、黄兆杰、孟琪、王杉

实习编辑 / 文静、方思宇、王芸、李笑颖、孙晶茹、安爽、陶诗诺、
万芊芊、郭培振、陈伊扬

装帧设计 / 忆纯轩文创、北京北青文化传播有限公司

主管 / 中国文学艺术界联合会

主办 / 中国电影家协会

编辑 / 《电影艺术》编辑部

出版 / 中国电影出版社有限公司

创刊 / 1956 年 10 月

地址 / 北京北三环东路 22 号 邮政编码 / 100029

电话 / 010-64296226 邮购部 / 010-64296443

电子信箱 / filmart1956@126.com

出版日期 / 单月 5 日

印刷 / 中国电影出版社印刷厂

国内邮发代号 / 18-280 国外订购代号 / BM33

国内发行 / 河北省廊坊市邮政局

订购零售 / 全国各地邮局 定价 / 20.00 元

国外发行 / 中国国际图书贸易集团有限公司（北京 399 信箱）

中国标准连续出版物号 /

ISSN 0257-0181

CN 11-1528/J

人工智能影像生成的模型诗学



朱恬骅 上海社会科学院文学研究所助理研究员，主要研究方向为计算机艺术与美学。

朱恬骅

【摘要】通过人工智能（AI）开展的影像生成，在艺术创作与研究面临表述危机，“模型诗学”有助于应对这一状况。模型诗学建立在对AI的历史溯源上，它以施行性为导向，从技术操作的视角审视模型的“类型”与“实例”。它结合了针对代码和文档的细读与操作者经验的采集分析两种操作方法，前者为AI模型提供“部件解释”，后者则以“接合表述”的方式呈现出AI模型承载的思想观念。模型诗学揭示出AI模型自身可以是艺术表达的载体。

【关键词】模型诗学 计算性诗学 控制论 黑箱 细读

人工智能（AI）在电影制作中的角色逐渐从幕后走向台前，成为学术界和工业界共同关注的焦点。2024年2月，“Sora”的横空出世，更让人看到“全AI”的影像生成已近在咫尺。行业分析指出，AI的规模化应用，将使“内容制作的成本显著降低、效率显著提高”，实现“个性化”的内容生产。¹一场“机器主导内容生产的全方位变革”²无疑正在迫近，一些学者因此希望，借助AI进行的影像生成，将能够“批判性地解决今天文化工业中的结构性问题和历史性困境”³。

在这些现有的论述中，进行影像生成的AI主要以技术现成品形式出

现，是可供获取的软件和服务。它的产业逻辑相对明晰，但艺术意涵上的诠释则面临多方面的困难，尤其是在艺术的研究与创作中起到关键作用的AI模型陷入了“表述危机”。⁴为应对这一状况，本文尝试提出一种“模型诗学”，以期在AI模型的技术性内容和生成影像的艺术性价值之间建立关联。首先，本文借镜AI的历史渊源，明确“模型诗学”的基本含义和立场。然后，结合“Stable Diffusion”这一常用的“文生图”模型，展示模型诗学的基本操作方式，也就是技术文本细读和对于操作者经验的重构。最后，表明技术文本的细读如何同有关

操作者经验的调查相结合，形成对AI模型的完整表述，达成模型诗学的初步目标。

模型诗学的基本立场与内涵

按照人们对“诗学”的古老定义，它是有关事物制作的知识。⁵丹尼斯·特能（Dennis Tenen）将其移植到软件研究的语境中，认为从思想到事物，穿越“表面”内容以揭示那些使其具有意义的平台和基础的过程，均属于“诗学”探究的范畴，并因而提出了“计算性诗学”（computational poetics）的解释策略。⁶按照他的解释，计算性诗学深入

文本中微小的组成部分，对具体案例展开微观分析，从而与“远读”（distant reading）和文化分析等宏观层面的考察相区别。

“计算性诗学”在软件研究的语境中重新引入了“细读”（close reading），它扎根具体细节的品格亦有利于我们立足AI模型的技术事实展开对它的解释和论述。但也正是因为“细读”所关注的是文本，特能的“计算性诗学”在很大程度上只是将文学创作意义上的“诗学”挪移到对技术文档的解释之中，探究技术人员如何自觉或不自觉地利用了各种修辞手段，构想和传达他们的发明与主张。AI影像生成依赖一系列软件环境，自然同样在“计算性诗学”的有效性范围之内。

然而，AI模型的情况要复杂得多。一方面，它是从数学上而非文本上给出定义的：一系列“超参”（hyper-parameter）确定了权重参数的规模和运算方式，这些参数和连接网络的图结构、限制模型运作条件的方程和函数等数学对象一同组成作为“类型”（class）的模型。它难以仅从技术文档等严格意义上的文本层面得到把握，遑论为其提供解释。

另一方面，模型不能“抽象地”仅凭数学定义工作；它需要数据集训练，需要读取权重参数的数据，需要一系列软、硬件环境和设施相互配套、协同工作，最终才能在实际的计算机系统上运转起来。在同样的“超参”、函数、结构等数学定义下，给予不同的权重数据，模型的行为可能会产生根本性的变化。它是人们真正能够接触到的模型“实例”（instance），然而天文数字般的参数规模，使静态的“细读”对此捉襟见肘。

“计算性诗学”需要经过改造才能为进行影像生成的AI模型提供解释策略，特别是它所标举的“细读”方法，在对象及方式上都需要重新根据AI模型的情况适配。此时，追溯AI技术的历史渊源，特别是控制论时代对“智

能”的研究中以“施行性”（performative）的行动颠覆“表征性”（representational）的文本，就显得尤其具有借鉴价值。

按照科学史家安德鲁·皮克林（Andrew Pickering）的观点，控制论将世界视为一个不断“在制作中”（in the making）的场所，具有动态性和不可预知性。⁷在这样的世界中行动，我们无法依靠某些事前存在的规则、定律进行预测，而是必须在不断感知周围环境的基础上灵活适应。从科学哲学的角度，皮克林将控制论的本体论称为施行性的，而与此相对的是现代科学对于表征的侧重。前者关注“在世界中的行动”，因此具有情境化等特点，而后者则追求普遍的知识，将世界改造为认识对象，并以现代科学的术语和方法来掌握它。⁸

施行性与表征性的区别可以通过恒温器和智能家居的比较来理解。传统恒温器是一个直接响应温度变化并控制电源的金属开关。而智能家居系统通过传感器收集数据，经过计算后，根据用户在手机应用上的设置来控制电暖器。恒温器是感知与效应的直接结合，而智能家居则需要将环境信息加以表征，并在多个中介物之间传递，如手机、传感器和服务器，它们可以分布在不同位置，甚至遍布全球。后者也具象化了现代科学中的“表征性迂回”（representational detour），即通过表征来间接感知和行动，而不是像控制论那样直接地在环境中起作用。

从“在制作中”的角度理解AI模型的立场，与“诗学”的古老意涵产生了共鸣。它将技术性论述中仅仅作为背景预设的“人”放置在问题场域的中心——人如何建构AI模型，又如何借助这些AI模型生成影像，这些“制作”过程构成了问题的关键。“制作”是动态的、过程性的、情境化的，贯彻了目的和意图。借用皮克林的说法，模型诗学同控制论一样具有“局部而非全局、特殊而非普遍、适应而非预测”的特性，它与“普遍知识”相对，是一种

1 腾讯研究院. AIGC发展趋势报告2023. (2023-01-31) [2024-06-12]. <https://research.tencent.com/report?id=AJJ>.
2 杜梁, 贾天翔. 生成式人工智能与电子游戏的生产机制革新. 艺术学研究, 2024 (2): 59.
3 赵宜. 人机共创、数据融合与多模态模型: 生成式AI的电影艺术与文化工业批判. 当代电影, 2023 (8): 20.
4 关于当代中文语境中的“表述危机”的概念, 参见: 郑元者. 美学观礼. 北京: 中国发展出版社, 2000: 174-177.

5 [古希腊] 亚里士多德. 诗学. 陈中梅, 译. 北京: 商务印书馆, 1996: 3.
6 Dennis Tenen. *Plain Text: The Poetics of Computation*. Stanford: Stanford University Press, 2017: 6.
7 Andrew Pickering. *The Cybernetic Brain: Sketches of Another Future*. Chicago: University of Chicago Press, 2010: 19.
8 Andrew Pickering. *The Cybernetic Brain: Sketches of Another Future*. Chicago: University of Chicago Press, 2010: 21.

“情境知识” (situated knowledge)。⁹

模型诗学的施行性特点为看待当今的AI模型提供了批判性的视角。OpenAI公司在有关“Sora”的技术报告中使用了“世界模拟器”(World Simulator)这一概念,¹⁰宣称它能够生成动态影像的原因在于,它“涌现”出对我们身处的现实世界进行模拟的能力。但从施行性的角度来看,“技术报告”中宣称的对“世界”普遍知识的掌握是不必要的,而且在很大程度上是不可能的。更进一步地,对模型进行实际测试,能够使我们从输入和输出之间的错位与谬误中,否定模型表征了普遍知识的说法:错误合成的影像不仅不能满足文字指令的要求,而且暴露出模型事实上并不“知道”譬如人的移动和海浪的运动应如何在视频中表现。从而,Sora生成的“瑕疵”施行性地反驳了“三维空间一致性”“长程相关性”等“世界模拟器”所需要的性质。

模型诗学关乎情境知识,意味着它始终处于人和物之间的具体关系之中,基于人在特定(有限)视角下的直接经验,而不是超脱物外的“上帝视角”。这促使我们回到罗斯·艾什比(W. Ross Ashby)提出的控制论的典型意象,即“黑箱”。按照艾什比的定义,

“黑箱”是在行动中依照情境的要求而操纵与响应的对象,¹¹而非当今流行的理解中绝对地“不可知、不透明、不公开”的现成品。这种黑箱具有相对性,因为人们完全可以拆解一个黑箱,专注于其中的某些部分,而将其他的部分继续视为新的黑箱。这种拆解和寻找操纵和相应的过程,代表了人们从已知求取未知、调用现成之物而制作尚未存在之物

的方式,也就是一种“制作”。在下文中,我们就实验性地以这种具有相对性的“黑箱”视角,从施行性的“制作”角度,具体考察一类广泛应用的AI图像生成模型。

模型类型的部件解释

隐扩散模型(Latent Diffusion Models)是AI图像生成领域重要的一类模型,是多种“文本到图像”“图像到图像”生成技术产品(如Stable Diffusion)的共同基础。按照通常的表述,它的基本原理是:对输入的训练图像进行隐空间编码,并多次迭代添加不同程度的噪声信号,使之变得彻底无序(即“噪声扩散”)。模型通过“学习”扩散过程的时序信息、约束条件和添加了噪声的输入数据,学习对输入的图像进行“去噪”,在约束条件的作用下倒推出上一迭代中噪声较少的信息,最终通过模型中的“解码器”部分,生产出可以显示的像素图像。¹²

但是,这些关于原理的陈述并不能“运行”。即便是论文中的图示、方程与陈述,也不能提供实际模型行为的因果判定。这体现了AI模型的施行性:对模型作者而言,模型的程序代码和数据先于论文中的表征性描述而产生,他

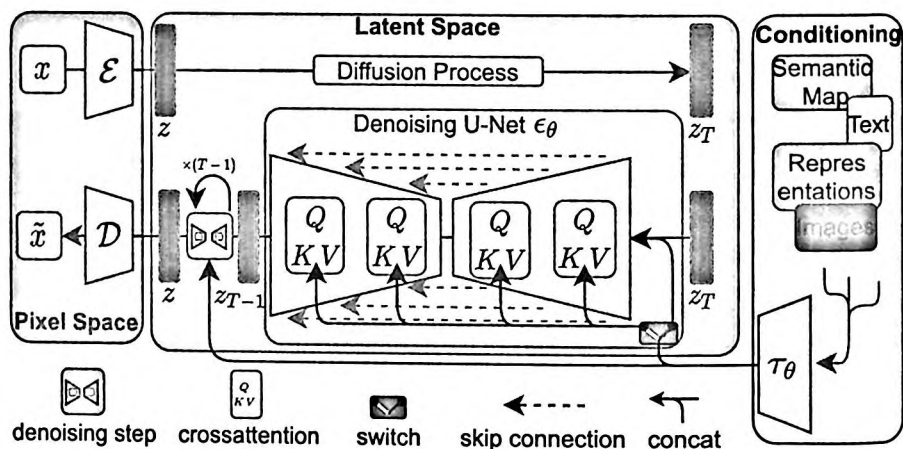


图1 “隐扩散模型”论文中关于模型结构的图示¹³

9 Andrew Pickering. *The Cybernetic Brain: Sketches of Another Future*. Chicago: University of Chicago Press, 2010: 25-26.

10 Tim Brooks, et al. Video Generation Models as World Simulators (2024-02-15)[2024-06-12]. <https://openai.com/index/video-generation-models-as-world-simulators/>.

11 参见: [英] W. R. 艾什比. 控制论导论. 张理京, 译. 北京: 科学出版社, 1965: 86.

12 Robin Rombach, et al. High-Resolution Image Synthesis with Latent Diffusion

Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 10674-10685. [代码来源于: Computer Vision and Learning LMU Munich. Latent Diffusion. (2022-07-26) [2024-06-12]. <https://github.com/CompVis/latent-diffusion>.]

13 参见: Robin Rombach, et al. High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 10674-10685.

们先取得了实际运作并取得良好效果的模型，再对它进行表征性的描述以使之公开。因此，程序代码和数据不是从图示等表征的模块结构中推导出来，而是相反，图示等对实际能够运行的模型（代码与数据的总和）进行了记录。我们可以粗略地在程序代码和论文作者所提供的图示中摘出一些方面，来描述和解释模型的“类型”。

依照论文中给出的图示和相关说明，“像素空间”中的编码器和解码器（即左侧的花体字E、D）为模型主体部分操作的“隐空间”（Latent Space）与图像的“像素空间”提供了相互转换的机制，而在中间的主体部分中，负责图像生成的主要是其中的“去噪U形网络”（图中的“Denoising U-Net”），其中又由若干的交叉注意力机制（crossattention）所构成，它们分别对应了程序代码文件（“openaimodel.py”）中定义的“UNetModel”和“AttentionBlock”类（行413-742、行248-324），而每一个“AttentionBlock”实际又由代码中的“QKV-Attention”或“QKVAttentionLegacy”（即前者的旧版本）类提供实现。图中多个交叉注意力机制之间的跳接（skip connection），在代码中是通过调用具有恒等连接的残差模块（在代码中被称为“ResBlock”）¹⁴、前向传播（“forward”成员函数）中定义的运算顺序和各模块间输出作为输入的关系得到体现的：前半部分交叉注意力机制的输出，将被压入栈中保存；后半部分在运算前，先从栈中依“后进先出”的顺序取得先前的运算结果，连缀到本模块所得到的输入上（行736）。实际上，在当前版本的代码中，上面所描述的过程已经处于版本过渡阶段，只有置“legacy”参数为真，即在旧版本兼容模式下才会运行（行594），暗示这一模块的实现方式已经过时，是出于兼容性的考虑被保留。

比较程序代码文件和论文中的图文叙述，可以发现前者提供的内容，实际远比后者中已经归化为现代科学语言的描述来得丰富，而代码本身的修改记录又进一步赋予了它历史的维度。同时，黑箱的相对性，也通过在模块化设计而在程序代码中得到了充分的体现。开发者可以将编程的任务进行充分的分解，以保持复杂程度在可接受的水平

之内，同时也使团队协作的编程成为可能：每一个开发者只需关心自己的任务，而可以假定其他部分已在功能上取得完善。例如，开发者无须手动计算梯度函数并为之编写进行计算的代码，就可以安全地假定深度学习框架（在这一案例中为PyTorch）的自动求导功能可以提供正确的数值解。

诚然，上文对程序代码的解读是带有任意性的，只挑选了其中框架性和概要性的部分。但是，对代码的解读可以任意跳过不感兴趣的细节，这一事实本身是有意义的。模型诗学要求研究者像使用那些程序代码的人一样对程序进行一种有所“克制”的“细读”——不必完全理解模型在计算机中进行演算的物理过程、认识操作系统与硬件设备的数据指令往来，也无需在意被跳过的细节，而只需假定它们完成了相应的功能，只要程序的实际运行表明它的确合乎设计的意图。这种“任意”的解读要求人们关心程序代码中各个模块“做什么”，而不是它“是什么”，从而快速定位到对自己有用的“接口”。

上述带有任意性（但同时也贯穿了目的性）的解读方式，可以借助理论生物学家斯图亚特·考夫曼（Stuart Kauffman）提出的“部件解释的接合”（articulation of parts explanation）得到说明。¹⁵考夫曼认为，我们可以通过分析各个组成部分（部件）的特性，来解释这些部分组合后所表现出的整体行为。分解的方式如果不同，那么得出的部件和基于此部件得出的解释也自然有所不同。不同的解释相互接合，将能够帮助人们对整个系统取得更多的认识。这些不同分解的结论之间只需相互兼容，而无须能够相互推导（相互等价）。在上文对于隐扩散模型的解读中，程序代码和论文图示这两种形式，就体现出了看待模型的两种不同方式：图示提示的是模型的初步框架，而程序代码则在框架中各模块的缝隙处填充了等待解释的大量细节，两者并不严格等价。而且，部件解释的表述也不要求某种“终极”的分解方式，即便这种分解有可能确实存在。这也呼应了程序中不同版本的代码并存的状况，因为即便从一个角度来看，二者具有相同的功能（即“做什么”）的模块，从另一个角度看，二者也可以有所不同，比

14 Kaiming He, et al. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

15 Stuart A. Kauffman. Articulation of Parts Explanation in Biology and the Rational Search for Them. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1970, 197: 257-272.

如它们耗费的时间或对最终输出的影响。程序代码的不同版本，来自对某一部分模块功能从不同角度做出的审视。

但是，上文针对代码的实验性“细读”缺失了“接合”的步骤：我们无法看到任何一种视觉成分在模型代码中直接的体现，相反，我们发现的事实是，图像的信息处于“像素空间”，并不直接参与“去噪”的过程。因此，目前呈现的“部件解释”仍未能对AI模型的“整体”行为，也就是它的影像生成提供表述，而只是在模型的“类型”层面做出了解释。我们可以设想下载该程序代码并运行的情形。由于代码提供的只是对模型中算法部分的实现，它还需要确定配置和相应权重都以数据形式存在，才能真正进入施行的状态。因此，程序代码所具有的施行性，以及人们对待代码的施行性视角，与模型本身的施行性还有所距离。从程序代码出发做出的部件解释，提供了进一步寻求“接合”的基础，但这种“接合”发生在别处，也就是模型的“实例”。

模型实例的操作者经验解读

为从模型的“类型”过渡到“实例”，我们需要重新从可实际运行的事物开始。对于隐扩散模型而言，这样的事物包括可以直接应用的软件包，如“WebUI”，以及在此基础上整合的“绘世启动器”等软件。相比于单纯描述模型类型的代码，这些软件提供了方便地下载所需权重数据和配置文件的方式，而且还围绕模型类型这一“核心”提供了大量“边缘”程序，以支撑模型相关代码的运行，如下载所需的参数权重数据，或是帮助进行输入图像的准备和输出图像的调整。此时，这些对模型类型而言的“边缘”，反而成为模型实例得以运行的关键：它们决定了模型“类型”被实现为怎样的“实例”（如用于生成具有照片般质感的人物影像，还是生成动漫风格的影像），从而产生出可操纵的“黑箱”。

基于对Stable Diffusion使用者网络社区的参与观察，我们发现了三种操作AI模型的经验，分别对应了三种进行技术操作的人。首先，我们考察与影像生成最直接相关的经验，也就是输入文本内容（提示词）、设置参数和与AI模型相交互的那部分。由于从模型中生成影像的过程具有

随机性，人们也称这种操作为“采样”（sampling）。对于此时的操作者而言，文本和参数是他们面对的主要“部件”。典型的行为包括，选取适应于特定目标的模型实例，指定不同的“标签”以取得意图的效果。作为其工作的成果，一系列图像将得到生成，附带有生成该图的提示词、模型名称（网址）和与影像生成直接相关的参数信息。其中，由于提示词与自然语言最为接近，它往往也得到最多的关注。2022年以来，一些网民编纂的提示词选集受到了采样操作者的欢迎，因为其中包含各类动漫人物插画图像风格（且一直保持更新）。采样时，操作者无需对模型类型有所认知，而对结果的关注也使他们的经验更容易与具体的模型实例、有限的输出结果高度耦合。比如，在某个版本的提示词选集中，其中的提示词可对某一系列的模型实例有效，但对其他模型则不容易产生意图的效果。

在与这些特定模型交互的过程中，采样操作者们自发形成了一些描述性的术语。例如，术语“污染”被用来描述多个主题标签（如“花朵”和“人物”）发生“融合”的现象，如使模型生成“头戴花朵的人物”，而非使用者所意图的“花朵与人物”。为应对“污染”，采样操作者利用一组无意义单词构成占位符，以“隔离”容易发生“污染”的标签，并将“污染”现象解释为主题标签在提示词中距离过近所致。¹⁶可以看到，操作者并不寻求在模型类型的层面解释“污染”现象（如所谓的“隔离”与进行图文对齐的CLIP模型有关），而是施行性地寻求对不符合意图的模型输出的改进，从而适应了具体的模型实例。当然，这种情境知识也可能包含某些被戏称为“玄学”的内容，标签的顺序、标点符号（如将“\$”符号解释为AI的“自我”）、连接词（将单词“and”解释为“元素”的“融合”）以及表情符号（emoji）等均可成为解释的对象。

这些解释不一定具有技术上的根据，实践上的效果也存在疑问（如使用“\$”等标点符号，更有可能触发软件内部的过滤机制，而非让AI产生“自我意识”），却因符合人们对于AI的想象而流行。它们无法通过模型类型的描述推导得出，甚至存在着误解的成分。但是这并不妨碍操作者们提供的部件解释发挥施行性的作用，也就是让他们在自己所关心的模型实例上（如某个从NovelAI公司泄露的

16 元素法典策划组。元素同典：确实不完全科学的魔导书。（2022-11-12）[2024-06-12]。https://docs.qq.com/doc/DWFdSTHJtQWRzYk9k.

模型实例)取得希望的视觉效果和审美内容。此外,操作者也形成了自己圈内的专门用语。比如,较长的“标签”组合被称为“咒语”,因为它能够如魔法故事一般“召唤”出所意图的人物形象。较短的“标签”组合则被视为“元素”,对这些“元素”和“咒语”有意识的整理和汇编,则形成了“法典”。这些专门用语的流行,标志着采样操作者开始形成松散的人群,从中甚至衍生出“魔法师”等具有明显身份意味的名称,用来指涉采样操作者的核心圈层。

第二种经验类型是操作者以模型的权重数据为主要“部件”,进行模型训练或微调,以求获得特化于自身意图的AI模型。其中特别值得注意的是以LoRA方式进行的模型适配(adaptation)。¹⁷由于所需的训练数据较少、训练时长较短、硬件要求相对较低且易于发布流通,其操作流程也在界面操作者的帮助下变得日益简便,几乎成为当下训练操作者的普遍选择。相比于大型的模型实例(被称为“基座模型”),LoRA的数据量较小,可视为对前者所做的“补丁”。在全球最大的AI图像生成模型社区Civitai上,好评度排名前1000的模型中就有80%以LoRA的方式提供,而它们往往专门针对某种图像风格、某种人物(角色)形象、某种动作或场景制作。¹⁸

由于往往为个人发布,这种模型在功能上也更加体现个人的审美偏好。但同时,公开模型的行为又意味着这种个人偏好能够激发他人的共鸣。进行训练的操作者,使AI模型在另一种意义上服务于自己的表达:他们将各自追求的图像形态赋予了通用的基座模型,从而左右了模型实例在整体上的行为模式。他们还可以为自己偏好的图像要素命名,一旦提示词中触发了相关字眼,即可调用这些训练好的模型“补丁”。

同采样的情形一样,操作者在进行训练时,也有属于他们自己的术语,其中一些来自技术社区公认的正式或非正式名词,前者如“学习率”“梯度下降”“微调”等专业概念,后者如“炼丹”(模型的训练过程)或“炸炉”(模型训练遭遇的各类失败)等戏谑性的说法。此外,这一群体也用一些专有名词来代指其所完成的某些功能,如使用少量图片即可较好进行模型微调的DreamBooth技术,其

缩略语“DB”被用来表示一般的模型微调过程,无论是否实际用到了DreamBooth本身。术语来源的高度技术化、术语内涵的功能化以及术语外延的模糊扩展,都折射出训练操作者并非专业的技术研究者。他们携带着术语,折返于技术的进展与关乎图像的审美讨论之间,调用了更多、更新的技术,服务于他人对特定审美风格和内容的需要。

位居采样与训练之间的,是以模型的运行环境等为操作对象的支持服务工作,它构成了第三种经验类型。提供支持服务的操作者们出于简化影像生成和相关准备过程的目的,设计构建人机界面等辅助设施,并帮助配置采样与训练所需的软件环境。受惠于这些提供支持服务的操作者,对AI模型的操作才能变得容易。因此,虽然服务型的操作者并不直接参与影像生成,而是以制作发行“整合包”“启动器”等技术现成品为主要工作,但他们仍然在AI影像生成中颇为重要,甚至有时可以成为操作者群体的核心成员,因为他们需要同时对两种经验有充分的认识。

这再次体现了模型诗学的基本立场,即通过“黑箱”的相对性所显示出的施行性特征。模型“类型”作为图像生成的“核心”,依赖于少数研究者的贡献,不会被轻易调整(也就是视之为黑箱)。但操纵参数的方式和对输入、输出数据的加工、修饰等“边缘”,却可以获得广泛的技术参与;它们对编写代码以扩展其功能的人而言是清晰明确的。“边缘”与“核心”一样不可或缺,也就是说,围绕“边缘”与“核心”、“类型”与“实例”的技术操作并无必然的高下之分。操作者只是根据自己的需要,将复杂的AI模型分解为若干与自己有关的部件,并在这些部件和模型输出的影像之间建立关联。

居间进行软件环境建构和界面设计的操作者,为不同“黑箱”、不同施行方式之间提供了转化迁移的过渡地带。他们在技术研究者和图像消费者之间频繁“越境”,将自身对技术部件的解释及其与图像产物的关联,固化在所制作的软件之中,也就使软件蕴含了他们自己对于模型类型的解释。他们在不同模型类型、不同技术之间穿梭,将它们编织、整合在一起,从而扩展了AI影像生成的能力。例如,将隐扩散模型与3D建模软件、渲染引擎结合,用于制

17 Edward Hu, et al. LoRA: Low-Rank Adaptation of Large Language Models. (2023-10-22)[2024-06-14]. <https://openreview.net/forum?id=nZeVKceFYf9>.

18 数据由作者采集,截至2024年6月14日。

作渲染动画视频，¹⁹ 从而让静态的“图像生成”模型可以完成一些动态的“影像生成”行为。

对三种操作者经验的考察使我们能够获得关于操作者经验的总体印象。通过标签与参数的组合，操作者获得体现了采样操作者意图的图像；通过模型的训练，操作者获得满足自己偏好的图像生成模式；通过人机界面与机器之间通信界面的设计，操作者又将对影像生成所涉及的各项技术内容的理解固定在软件中，令不同的模型类型相互协作。这三种经验并不是割裂的，因为进行训练的操作者也要通过提示词来检验自己“训练”的成果，而提供支持服务的操作者则往往同时兼顾“采样”与“训练”。在三种操作者经验共同作用下，AI影像生成真正体现并融会了不同操作者的意图。

由此，我们可以将“细读”带到对操作者经验的解读，发现操作者对于AI模型实例所形成的、依据各自有限的认知形成的“情境知识”。三种经验中不同的“部件解释”相互作用，并通过实际的操作行为而得到施行性的实现，展现出“情境知识”的面貌，为接近操作者经验提供了具有可操作性的框架。对模型“实例”的操作者经验与对模型“类型”的“代码细读”彼此呼应，“部件解释的接合”方才得以可能，而模型诗学的解释策略也才能初步达成其目的。

结论

早在20世纪80年代末，媒介理论家弗卢塞尔就已经预见到，装置自动合成的图像意味着一种诗学维度的敞开，能够成为“程序化了的诗歌和小说”，向人们提供看待和体

验世界的方式。²⁰ 他称合成图像为“概念计算的结果”，这准确预见Stable Diffusion中使用由大量标签构成提示词的惯常操作。这种“计算”同时使影像从“进行模拟的、模仿的功能中脱离出来，变得富于创造性和诗性”²¹。

基于AI技术发展历史的溯源，我们提出施行性地对待AI模型的“模型诗学”解释策略。它将AI模型视为相对的、情境中的“黑箱”，可以借助对代码的“细读”提供对模型“类型”的部件解释；而借助对操作者经验的调查，则可以重构和解释AI实例互动的三种不同经验，以完成“部件解释的接合”。在此过程中，不仅是由AI生成的影像，进行这种影像生成的AI模型本身也不断处于“在制作中”的状态。人们将自己从事不同类型技术操作的经验投射其上，通过技术上的努力以接近某种艺术效果。

因此，“模型诗学”不是从已有的诗学范畴（如“计算性诗学”倚重的修辞）看待技术如何被人感知和理解，而是从模型制作的（而非现成品及其产物的）角度，在围绕模型的技术性选择与思想观念间建立联系。在此意义上，AI模型是数据化了的思想，而且常常是传达审美偏好的思想。这意味着模型本身也可以成为某种“作品”，开辟创作空间；AI模型与围绕它的一系列“边缘”技术构件一道，成为人进行表达的方式。而模型诗学也就能兼顾狭义“诗学”的作用，继续提供有关文学艺术创作的有效知识。■

[本文系国家社科基金艺术学青年项目“计算机艺术历史生成问题的人类学美学研究”（批准号：21CA169）的阶段性成果。]

Model Poetics of Image Generation in Artificial Intelligence

Abstract: The use of artificial intelligence (AI) in image generation has posed a challenge to artistic creation and research, leading to a crisis in representation. "Model poetics" offers a solution to this situation. Drawing on the historical origins of AI in cybernetics, model poetics takes a performative approach to examining the "classes" and "instances" of AI models in the context of technical operations. It combines close reading of code and documentation with the analysis of operator experience, providing "parts explanation" for AI models and presenting the conceptual ideas they embody through "articulation", both in Kaufmann's terms. Model poetics reveals that AI models per se can serve as vehicles for artistic expression.

Keywords: model poetics, computational poetics, cybernetics, black box, close reading

19 渡邊英樹. AIを活用したアニメーション制作に注目! アニメTikToker『ツイ
ンズ ひなひま』. (2024-03-05) [2024-06-12]. <https://cgworld.jp/article/307HS-hinahima.html>.

20 Vilém Flusser. *Die Schrift: Hat Schreiben Zukunft?*. Göttingen: European

Photography, 2002: 62.

21 Vilém Flusser. *Die Schrift: Hat Schreiben Zukunft?*. Göttingen: European
Photography, 2002: 70.