反生成式 AI 技术群体话语实践研究——以微博账号“赛博十块鉴定所”为例

李梓瑞 黄思为 施序

浙江传媒学院，浙江杭州，310000；

摘要：生成式 AI 技术浪潮迭起，快速渗透进人类社会方方面面，在提供便利的同时也引发技术侵害危机，因此，部分反生成式 AI 技术者结成虚拟社群，通过话语实践来抵制生成式 AI 技术。研究发现，反生成式 AI 技术群体的话语实践有其积极抗争意义，能推动技术侵害事实的曝光、搭建抵抗深层侵害的话语防御阵线。然而，其话语实践背后仍然存在失范隐忧，该群体的话语征讨正逐渐显露出标靶错位和符码越轨的问题。

关键词：生成式 AI 技术；技术抵抗；网络社群；话语实践

DOI：10.64216/3080-1486.26.01.043

1 研究缘起

ChatGPT 活跃人数不断增加，生成式 AI 技术的浪潮下，同样可见“抵抗军”的身影，反生成式 AI 技术群体[1]在微博等社会化媒体平台上自发集聚并结成虚拟社群，他们拒绝并批判生成式 AI 技术的应用，并通过群体性的话语实践发起对生成式 AI 技术热潮的抗争。

2 文献回顾

2.1 技术抵抗

“技术抵抗”是指人们拒绝接触和使用特定技术的一种社会现象。有学者指出，在算法的程序化指令控制下，更好地对抗技术不可见牵引，找到难以拒绝的科技之“势”和难以逾越的人文之“忧”间的最佳均衡位置已成为当务之急。[1]对于技术抵抗的观察并未停留于抽象的理论思辨层面，也没有止步于隐私伦理的宏观纵览，其落点正逐渐下沉至对微观具体的技术抵抗群体的参与式观察和田野调查，譬如基于豆瓣话题小组所开展的对青年“数字极简主义者”生活实践的观察[2]以及“数字排毒”中的人技互动关系探究。[3]然而，学界目前有关生成式 AI 技术的技术抵抗研究还极为匮乏，尤其是缺少关于反生成式 AI 技术群体的深度透视，即本研究所针对的反对接触和使用生成式 AI 技术的群体，这是本研究的价值所在。

2.2 生成式 AI 技术的隐忧

生成式 AI 是指基于特定算法规则，扎根于大语言模型，在排除直接性的人为参与下生成文字、图片、音视频和代码等多模态内容的人工智能技术。[4]生成式 AI 技术所呈现出的不可思议的“涌现”能力，近年来正受热议[4]。成式 AI 技术的智能迭代高度依赖于数据模型本身，这就不可避免地会陷入进更深的隐私侵扰的泥淖之中，深度伪造现象愈演愈烈，也引发人们对于信息安全问题的担忧。生成式 AI 技术所催生的职业恐慌也值得警惕，部分职业或是部分工作任务，尤其是对模式化、可迭代的高稳定度职业的确形成了不小的威胁。[5]技术迭代下的艺术、审美等社会文化方面的受侵蚀境况同样堪忧，AI 绘画的本质逻辑仍然是对既有风格和作品的分解与重组，对技术的依赖将会陷入作品风格和审美取向的同质化怪圈。[6]生成式 AI 技术的隐忧是多维度的，但大部分研究都是以技术本身作为探究的起始点，而相对缺乏对反生成式 AI 技术群体的话语实践的具象观察和反思，因而本研究将从微观具体的技术抵抗过程切入，探究反生成式 AI 技术群体的话语策略和符号表征，更具象地审视生成式 AI 技术。

2.3 研究设计

本研究聚焦反生成式 AI 技术群体的话语实践，采用参与式观察和文本分析的质性研究方法，选取微博账号“赛博十块鉴定所”作为实际的研究落点，截至 2024 年 8 月，该账号已有近 40 万粉丝数，内容累计转评赞量突破 2 千万，主要发布用户投稿的反生成式 AI 技术的相关文本和话语内容。通过比对其他类似平台和账号，研究发现该账号综合体量和成员活跃度最高，代表性强，故选取其作为实际落点。研究借助八爪鱼数据工具爬取

该微博账号自 2024 年 3 月 1 日至 2024 年 12 月 10 日所发布的全部博文，经数据清洗后最终获取到 9620 条有效博文样本，通过对数据文本的深度阅读与文本分析，以开展关于反 AI 技术群体的话语实践研究。

3 抗争：对技术侵害的多维抵抗

3.1 自组织鉴别：技术侵害的话语曝光力量

媒介技术迭代在赋予人们时空感知变革的同时，其对信息环境的深度塑形也加深了人类与真实世界间的沟壑，AIGC 正在越来越多地成为一种社会常态，不断冲击行业结构，[7]部分不法分子借助生成式 AI 技术进行内容伪造，侵害用户权益，而反生成式 AI 技术群体基于微博账号“赛博十块鉴定所”开展的自组织鉴别和曝光成为了对抗此类技术侵害的有力手段。

“淘宝商家用 ai 生成图片刷单刷好评 很可怕 如果是年纪大点的人根本分辨不了 ”（12 月 6 日投稿）

基于生成式 AI 技术实现的造假侵害首先常见于网络电商平台，部分商家利用生成式 AI 技术产出商品相关的图片或文字内容，在此技术上进行商品售卖以牟利。随着生成式 AI 技术生成内容越来越逼真，此类侵犯消费者权益的技术造假和侵害行为往往很难被快速识别，而反生成式 AI 技术群体的话语实践为处理上述难题提供了新的路径，群体成员们通过投稿来自组织地呼吁鉴别和曝光侵害式，这样的话语实践方式促进了对此类技术造假侵害事件的识别和曝光效率，这样基于话语的群体互动在抵抗技术侵害中发挥了有利作用。

“我在某黄色网站发现某账号发表我们学校的女孩照片以及 ai tuo yi 照片。他发的女孩我不认识，但是照片背景是我们学校，我一眼就能看出来。求助我该怎么办?我应该上报吗?我该怎么告诉受害者?”（9 月 7 日投稿）

利用生成式 AI 技术来对照片或视频里的真人进行换脸的深度伪造侵害模式也是生成式 AI 技术侵害的常见向度。其中最为广泛和恶劣的情况就是 AI 换脸编造关于女性的黄色谣言，这样的技术侵害是对当事人名誉权和肖像权的严重侵犯，会对当事人及其家庭造成沉重的伤害。反生成式 AI 群体通过话语实践来对此类情况辨识和澄清，以一种自发的自下而上的话语合力形式，助推辟谣进程，并对造谣者施以舆论压力，既能有效遏制谣言的泛滥，还能在技术侵害浪潮中形成对受害者的群体性话语保护。

“说得极端一点，生成 AI 制造的假冒产品可能会被认知为“我们才是真品”，就算想要自卫，出现如此精巧巧妙的造假，个人的力量也无法与之抗衡。”（11 月 21 日投稿）

反生成式 AI 技术群体的话语抗争，形成了克莱舍基在《未来是湿的》一书中所提出的“无组织的组织力量”。[8]这种力量的构造过程分为三个维度：首先是信息和注意力共享维度，反生成式 AI 技术群体依托微博这一社会化媒体平台，聚集于微博账号“赛博十块鉴定所”，经由共同关注的形式实现信息和观点共享；其次是协同合作维度，反生成式 AI 技术群体成为通过评论、转发、点赞等互动方式表明态度意见并协同合作；最后是集体行动维度，反生成式 AI 技术群体的话语征讨并不止步于群体内部的话语共振，而是会将谐振波以集体行动方式向外输出。例如采取私信谴责、向平台举报等集体行动，来充分释放“无组织的组织力量”，推动具体问题解决，取得技术抵抗的实际成果。

3.2 价值性坚守：深层侵害的话语防御阵线

微博账号“赛博十块鉴定所”的置顶博文解释了这一反生成式 AI 技术社群的建立初衷，“生成类 AI 给人带来的最大灾难就是消解价值性，它会让人质疑自己的意义。”反生成式 AI 技术群体的话语实践，除了希望能通过此种方式揭示生成式 AI 技术的具体侵害案例事实外，更是希望能通过小社群“星星之火”般的话语实践，建立起抵抗生成式 AI 技术对价值性深层侵害的广泛防御阵线。这种深层次的价值性可以细分为人文主义、艺术灵韵、审美感知三个内在向度进行探究。

“总感觉 AI 只是进一步抽打工人的鞭子，幸灾乐祸的也只是没抽到他们身上呗 ”（6 月 5 日投稿）

贝尔纳·斯蒂格勒曾以“爱比米修斯的过失”来隐喻人与技术的“代具”补偿关系，爱比米修斯在分配属性时，因疏忽而遗漏了人类，普罗米修斯由此盗取火种以补偿一无所有的人类。[9]技术挤占下，人文主义的稳固性亟需被再度审视，反生成式 AI 技术群体通过话语实践来唤醒在生成式 AI 便利性中渐趋迷醉的人们，引导人们再度开始审视技术与人的关系。

“宣辩 Ai 绘图的价值就像复印一张蒙娜丽莎后宣称自己是达芬奇一样荒谬……ai 绘图的表达不是由审美主体籍借其经验完成的，和你没有关系，那他就和复印的蒙娜丽莎一样没有艺术价值，就算它看上去再怎么

美也一样。难道有人觉得自己因为复印了一张蒙娜丽莎，而够格在卢浮宫开画展吗?况且 ai 绘图的成品，有没有蒙娜丽莎那样美呢?”（4 月 11 日投稿）

本雅明认为，艺术作品有其“灵韵”，一种朦胧的、神圣不可接近的、独一无二的价值辉光，机械复制技术的发展则正迫使“灵韵”走向消逝，这种消逝也代表着 19 世纪前传统艺术的凋零。[10]

“教育，从教育开始人类就已经要完蛋了去年改革从线面改为线性拿分更高，不知道是什么意思，说这样好看，当然，好看，方便 AI 生图嘛。。很难受，很难受，热爱画画是因为想创作出属于自己的美好的东西，而不是想被锁在这昏天暗地的小厂房里重复一遍又一遍他人喜欢接近 AI 的审美。”（8 月 12 日投稿）

“灵韵”的落点是作品，而审美感知的落点在于人。反生成式 AI 群体认为，在 AI 生成的信息流过度充盈的智能社会之中，人们的审美感知能力会不断弱化和异化，在这一浪潮下，引导人类审美感知进步的内容生产者也将逐渐落寞，被迫转型为强化 AI 式审美风格的无思想的机器。

微博账号“赛博十块鉴定所”的置顶博文中写道：“这是互联网用户一同面对的灾难，我们需要一同表示反对，表示愤怒，一旦缄默闭口不谈，那将是整个社区的覆没。”反生成式 AI 技术群体的话语抗争有其积极意义，即凭借坚定的反生成式 AI 技术的态度和实际的话语活动，尝试构筑起抵抗生成式 AI 技术深层侵害的话语防御阵线，以遏制技术浪潮下人类文明价值性的消解趋势。

4 失范：话语征讨中的沉疴隐忧

4.1 错位的标靶：从捍卫到加害

反生成式 AI 技术群体内部成员的鉴别素养参差不齐，因而在识别 AI 生成内容时误判而对无辜者进行话语攻击的情况时有发生。在面对群体性的话语征讨时，被征讨者承受着巨大的污名化压力，严重的名誉损失和精神损害，作为被征讨者，虽然可以提出撤稿要求，但其遭受的名誉损失乃至经济打击却难以修复。微博用户@WhaleSink1007 就曾发博指责微博账号“赛博十块鉴定所”的鉴别错误，用充分的图文实证材料证明被鉴别海报是自己纯人工原创完成，而非 AI 技术生成，同时对投稿人拒不道歉的行径大加谴责。该证明博文下用户@用户名招募中的评论恰到好处地点明了问题的症结：“空口鉴别 AI 让投注了心血的画师寒心，让真正的 AI 画手半夜做梦都笑醒。”反生成式 AI 技术群体的话语实践的初衷本在于保护和捍卫传统内容创作者的正当权益，但在实践中竟成为了直接伤害传统内容创作者的一股加害力量，这种颠倒现象揭露出当下反生成式 AI 技术群体的话语实践中的一大隐忧。

4.2 越轨的符码：从正义到恶意

话语实践中为群体成员所共享的符码系统，既是群体内部信息通畅的重要元素，也是塑造群体对外所展示的良好符号形象的第一要素，当其走向畸变和扭曲时，群体的符号化形象也将随之异变。对反生成式 AI 技术群体话语实践的研究发现，其所塑造的符码系统具有显著的越轨特征，包括负面隐喻的谐音替换语词和负面情绪高昂的直接侮辱性言论。

“癌哥能不能棍，玩癌的都丝了”（8 月 15 日投稿）

“话说这些用尸块抄出来的作品，不会觉得是你们自己心血之作吧”（2 月 28 日投稿）

反生成式 AI 技术群体所塑造的这一畸变型符码系统，本质上是在将与生成式 AI 技术相关联的人与物，同带有衰败、低俗、污秽等特质的符码元素，建立起某种潜在的语义链接，由此来实现对生成式 AI 技术及其使用者的贬低，而谐音替换则是为了规避平台的潜在技术审核，以保证负面话语得以公开展示。诸如“杀/鲨”“死/三/丝”“晦气”“恶心”“祸害”等词汇频繁出现在该群体的话语实践之中，成为群体成员宣泄负面情绪和现实压力的途径，然而，此类强恶意倾向的符码元素在群体话语实践中的累积，却又反过来污染了群体内部的文化氛围，也将本就边缘化的该群体进一步偏置于难被主流所认可的社会角落。

5 结语

在迈向技术与人和谐共生的理想化社会的道路上，技术的复杂性应该被看见，反生成式 AI 技术群体通过话语实践捍卫价值性的努力，虽然沧海一粟，但在对新技术趋利避害的社会化过程中不可或缺，与此同时，人的复杂性也应被看见，反生成式 AI 技术群体的话语实践中存在的社会隐忧需要得到更加充分的审视，并借助恰当方式来予以引导和规范。

参考文献

[1]王鑫.在共生中抵抗：算法社会的技术迷思与主体

之困[J].东南学术,2023,(04):218-228.
[2]徐冠群,朱珊.媒介技术的抵抗：青年“数字极简主义者”的生活实践——基于豆瓣话题小组的田野调查[J].传媒观察,2023,(08):93-103.
[3]胡明鑫.用户如何走向抵抗？——从数字依赖到数字排毒的人技互动关系[J].新闻记者,2023,(06):86-100.
[4]喻国明,苏芳,蒋宇楼.解析生成式 AI 下的“涌现”现象——“新常人”传播格局下的知识生产逻辑[J].新闻界,2023,(10):4-11+63.
[5]邱泽奇.“ChatGPT,你怎么看?”——与 ChatGPT 探讨AIGC对人类职业的影响[J].探索与争鸣,2023,(03):13-16.
[6]赵睿智,李辉.AIGC 背景下 AI 绘画对创意端的价值、困境及对策研究[J].北京文化创意,2023,(05):42-47.
[7]谢梅,王世龙.ChatGPT 出圈后人工智能生成内容的风险类型及其治理[J]. 新闻界,2023,(08):51-60.
[8]克莱·舍基.人人时代:无组织的组织力量[M].胡泳,沈满琳,译北京:中国人民大学出版社,2012.
[9]贝尔纳·斯蒂格勒.技术与时间 1:爱比米修斯的过失[M].裴程，译.南京:译林出版社，2019：53.
[10]瓦尔特·本雅明.摄影小史一一机械复制时代的艺术作品[M].南京:江苏人民出版社,2006.

作者简介：李梓瑞，2001 年 5 月 8 日，女，汉族，山西太原，浙江传媒学院，新闻与传播专业研三学生，研究方向为广播电视与融合和新闻。
黄思为，2000 年 9 月 7 日，男，汉族，江苏苏州，浙江传媒学院，硕士研究生，数字媒体与智能传播。
施序，1997 年 9 月 7 日，男，汉族，浙江湖州，浙江传媒学院，硕士研究生，数字媒体与智能传播。
项目基金：浙江传媒学院研究生科研与实践创新计划项目“反生成式 AI 技术群体话语实践研究”