

深度伪造与国际政治的新图像时代

张 锐 肖 毅

内容提要 深度伪造通过图像技术、图像本体两种形态,对国际政治产生日益显著的影响。具体而言,深度伪造技术带来国际政治图像的生产传播变革,包括多元主体的图像共创、虚实结合的画面涌现、高度依赖网络社交媒体和图像技术竞争主流化。深伪图像主要有四种类型,宣传性、娱乐性、批判性图像能够引导受众认知真实世界、传播积极情感,而欺骗性图像则以误导目标受众、煽动消极情绪为核心目的。深度伪造使图像获得更多创造国际政治叙事的机会,既能在虚实杂糅中促进国际良性互动,也能破坏国际交往氛围、催生分歧或冲突。当前,深度伪造的实施仍受制于技术门槛、社会文化环境和行为体防范等因素,它在国际信息战中的角色、所引发的图像竞赛和大众地缘政治新走向仍需跟踪观察。

关键词 非传统安全 深度伪造 图像 国际政治 大国博弈

* 张锐:全球能源互联网发展合作组织经济技术研究院研究员。(邮编:100031);肖毅:天津大学新媒体与传播学院副教授、天津大学网络空间国际治理研究基地研究员。(邮编:300072)

** 本文系国家社科基金青年项目“突发公共事件中深度伪造技术的舆论催化机制与对策研究”(项目编号:23CXW020)阶段性成果。感谢《国际政治研究》匿名评审专家的意见和建议,文责自负。

深度伪造(Deepfake,以下简称“深伪”)是人类图像史上一次重大的技术变革^①,无论人们是否喜欢这一技术,国际政治都在加速迈向新的图像时代。近期最容易被提及的案例无疑是乌克兰危机爆发之后的图像攻防,2022年3月16日,乌克兰总统泽连斯基宣布投降的深伪视频在全球社交媒体广泛传播,有黑客攻击了乌克兰新闻媒体,在该国电视直播中播放了这则视频;3月18日,西方社交网络开始流传一段俄罗斯总统普京宣布“设法与乌克兰达成和平”的深伪视频。两段视频尽管都在很短时间内被揭穿为虚假信息,但其政治意图不言而喻,都旨在“影响目标官兵的士气及网民立场,从而为物理世界的战争赢得心理优势”。^②2023年10月,新一轮巴以冲突爆发后,反映双方受害者遭遇的深伪照片或视频充斥于全球社交媒体,有观点认为“承担深度伪造的代价已经是战争的一部分,我们处于一场深刻的虚假革命的开端。”^③其实,深度伪造不仅高度介入当前的国际危机和冲突,更活跃于日常的、事关国际政治的大众舆论场,无论官方还是民间,对其的应用都存在多元的路径、多重的目的。

深度伪造是由“深度学习”(Deep learning)和“伪造”(Fake)构成的合成词,指“利用深度学习算法进行自动化的数据修改、处理和生产,实现图片、音频及视频的智能模拟和伪造”。^④目前,最受关注的技术应用就是所谓的“人工智能换脸”,可以实现人脸合成、人脸替换、唇形同步、人脸再现、动作迁移等图像篡改效果,也可以通过动态滤镜对人物形象进行实时改造,使人能够在视频通话中“换脸”;对于视频中的人物声音,该技术也可实现语音克隆、语气模仿、多语种切换等效果。近年来,该技术的发展势如破竹,在影视、教育、营销、医疗等诸多领域扮演日益重要的角色,“对升级传统内容制作方法、改善制作效果、提

① 很多观点认为,应用“深度合成”概念取代“深度伪造”,因为后者在字义上容易给相应的人工智能技术造成污名化影响,可能扼杀技术的潜在社会福利。本文仍使用“深度伪造”,原因是国际政治新闻报道、学术探讨中仍普遍使用这一概念。

② 何康等:《认知的罗生门效应制造:深度伪造在俄乌冲突中的案例分析》,《新闻界》2023年第1期,第93页。

③ Matt Lebovic, “These Israelis Are Fighting Hamas on the War’s Emerging ‘Deepfake’ Cyberfront,” *The Times of Israel*, October 18, 2023, <https://www.timesofisrael.com/these-israelis-are-fighting-hamas-on-the-wars-emerging-deepfake-cyberfront/>, 2024-06-15.

④ 吴静:《“深度伪造”技术在传媒领域的新应用和异化风险》,《传媒》2023年第3期,第51页。

高制作效率有积极作用”，但也容易被不法分子用于制造虚假信息、从事诽谤、诈骗等活动，给国家安全和社会稳定带来挑战。^①

深度伪造对国际政治造成哪些影响？既有文献主要关注其所带来的负面冲击，担心这一技术正在沦为地缘政治的斗争工具。浙江大学刘国柱教授认为：“总体来看，深度伪造给本就紧张的大国关系和国际安全形势带来了更为复杂的挑战，尤其是加剧了国际社会和现有国际体系本就存在的信任赤字。”^②清华大学的苗中鸣等认为深伪技术的异化使用面临武器化风险，“特别是在冷战和局部热战对峙的危险时期，一则深度伪造的新闻和消息，可能是导致战争爆发的‘最后一根稻草’”。^③有评论认为，“通过侵蚀真相的内涵，损害公民与国家之间、国家与国家之间的信任，深度伪造可能成为全球治理体系的最大破坏力量。”^④悲观的观点甚至认为，深度伪造所创造的后真相世界可能抵消人类“于 20 世纪在和平、稳定和繁荣方面取得的大部分进展”。^⑤另外，一些观点强调不应过度放大深度伪造的威胁，“深度伪造及其相关内容的出现只是假新闻的延续，这些挑战在世界历史上并不罕见，在 21 世纪头几年更是尤为突出”。^⑥美国媒体人詹姆斯·奥斯特罗夫斯基(James Ostrowski)认为，大众舆论在缺乏根据的情况下不断宣传深度伪造将减弱社会对现实的控制，但“为什么我们至今很少看到高调且成功的深伪攻击”，“人工智能并未像警告指出的那样，变戏法似地向后事实的荒地进军、扼杀真相”。^⑦总体来看，现有研究虽然揭示深度伪造对国际社会的威胁，但普遍存在以下不足：一是未能充分呈现深度伪造

① 赫然：《规范深度合成技术中国走在全球前列》，《人民日报》（海外版）2023 年 1 月 30 日，第 10 版。

② 刘国柱：《深度伪造与国家安全：基于总体国家安全观的视角》，《国际安全研究》2022 年第 3 期，第 15 页。

③ 苗中鸣等：《颠覆性技术异化及其治理研究：以“深度伪造”技术的典型化事实为例》，《科学学与科学技术管理》2020 年第 12 期，第 90 页。

④ Anis Bajrektarevic, “The AI-Dual Use Technology?” *Modern Diplomacy*, November 18, 2023, <https://moderndiplomacy.eu/2023/11/18/the-ai-dual-use-technology/>, 2024-06-15.

⑤ Simon Parkin, “The Rise of the Deepfake and the Threat to Democracy,” *The Guardian*, June 22, 2019, <https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy>, 2024-06-15.

⑥ Christopher Whyte, “Deepfake News: AI-enabled Disinformation as a Multi-level Public Policy Challenge,” *Journal of Cyber Policy*, Vol.5, No.2, 2020, p. 213.

⑦ James Ostrowski, “Shallowfakes: The Danger of Exaggerating the AI Disinfo Threat,” *The New Atlantis*, Spring 2023, <https://www.thenewatlantis.com/publications/shallowfakes>, 2024-06-15.

的影响路径和深度,而且不少观点只将其作为人工智能时代的“洪水猛兽”,无视其“社会善用”的可能性;二是没有区分深伪技术和深伪图像,将两者混为一谈;三是经常将深伪图像等同于虚假信息,忽略现实世界存在大量明确标注“使用深伪技术”或为了传播真实信息的深伪图像,它们并不承载主观上的欺骗意图,现有研究几乎没有涉及这类图像。

本文旨在更全面地阐释深度伪造对国际政治的影响。研究立足于大众地缘政治学的理论视域,结合近年来深度伪造的技术应用场景和图像传播实践,运用视觉话语分析、数字民族志等质性研究方法,探讨深度伪造如何对国际政治形成多维影响,进而揭示国际政治新图像时代的具体内涵。同时,研究将分析限制深度伪造发挥作用的要素,以反驳“深度伪造必将制造世界大乱”的悲观论调。

一、图像如何影响国际政治

图像对国际政治的影响并非国际关系研究的“显学”,但随着近年来大众地缘政治学的兴起,国内外积累了一定研究基础,本文充分挖掘现有学术成果,试图为深度伪造的影响搭建一个可行的分析框架。

(一) 大众地缘政治学的基本观点

大众地缘政治学是批判地缘政治学的一个分支,兴起于20世纪90年代的欧美学界。大众地缘政治指“地缘政治事件在电视、音乐、漫画、互联网和广播等媒体中的大量表现”。^①这一概念蕴含的基本假设是地缘政治是一个“基于符号、原型和想象的话语系统”,国际关系行为体通过在大众媒体上宣扬清晰的战略叙事,能够在权力领域内追求各自所希望的结果。^②“虽然地缘政治的关键时刻一般体现为国家领导人引人瞩目的实践(如发动战争或侵略、展示军事力量等),但这些举动和构成国际政治行为的日常实践是由各种文化中丰富

① 胡志丁等:《西方流行地缘政治研究进展与展望》,《地理研究》2020年第7期,第1468页。

② Laura Roselle, Alister Miskimmon, and Ben O'loughlin, "Strategic Narrative: A New Means to Understand Soft Power," *Media, War & Conflict*, Vol.7, No.1, 2014, pp. 70-84.

的表征实践所建构、维持和赋予意义的。”^①除政治精英外,大众可以在“各种流行文本中拥有‘表征世界’,并以此‘建构’现实的权力”。^②大众地缘政治学的价值在于揭示文化传播与国际政治的互动、大众媒体影响下的地缘政治想象与行动。

大众地缘政治学的实证研究长期关注代表性或争议性的国际政治图像,国际关系史上这样的显著案例屡见不鲜。19世纪末,美国主流报纸热衷为新闻搭配极度夸张或捏造情节的插画、照片,这类新闻被称为“黄色新闻”,当时为美国发动美西战争进行舆论造势、奠定民意基础,构成美国帝国主义在“意识形态上被质询的主体”。^③冷战期间,美国、苏联频繁使用各种标志物进行图像建构、开展软实力对抗,美国将摩天大楼作为“象征美国民族主义崛起的一个视觉符号”,而苏联则在电影、动画片中将摩天大楼描绘成“贪婪、不人道、丑陋和剥削的美国形象的象征”。^④回顾历史上的大国博弈,利用精心设计的图像向竞争对手释放错误信息是大国信息战的常用手段;尤其西方大国长期通过流行图像等文化工业产品(如好莱坞电影、电视剧等)对外传递“非西方世界有赖于西方世界改造”的错误观点。大众地缘政治学长期关注电视、电影、报刊上的政治图像,但近年来,研究视角日益转向“来自新的社交媒体和其他基于互联网平台”的、具有“互动元素”的图像。^⑤

(二) 图像对国际政治的影响框架

根据大众地缘政治学的观点,本文尝试搭建图像对国际政治的影响框架。如图1所示,将图像解构为技术和本体两种形态,它们分别对国际政治的秩序

① Gearóid Ó Tuathail and Simon Dalby, “Introduction: Rethinking Geopolitics: Towards a Critical Geopolitics,” in Gearóid Ó Tuathail and Simon Dalby, *Rethinking Geopolitics*, London: Routledge, 1998, p. 5.

② 葛静深:《大众地缘政治想象中的身份政治与“他者”形象:以美俄(苏)之间的大众文化形象建构为个案》,《外交评论》2022年第1期,第96页。

③ [美]约瑟夫·格雷戈里·马奥尼:《东方主义、新帝国主义和美国的媒体政治》(张也译),《国外理论动态》2015年第10期,第57页。

④ 葛静深:《大众地缘政治想象中的身份政治与“他者”形象:以美俄(苏)之间的大众文化形象建构为个案》,第116页。

⑤ Steve Pickering, *Understanding Geography and War: Misperceptions, Foundations, and Prospects*, New York: Palgrave Macmillan, 2017, p. 89.

和行为产生影响。

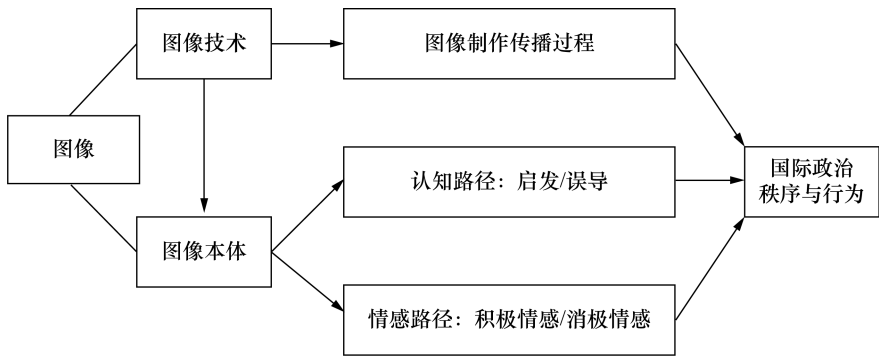


图 1 图像对国际政治的影响框架

图表来源:笔者自制。

图像技术对现实世界构成直接影响,但往往容易被人忽略,其通过“影响人类的感知器官、认知和预认知功能及情绪流露”在政治生活中发挥不可取代的施动性影响。^① 具体而言,“图像制作过程中的技术手段决定着图像类型,而不同类型的图像在所指涉内容和建构社会意义的作用上存在很大区别”^②,就模态类型上而言,动态图像描述现实、打动人心的能力经常比静态照片更胜一筹。除制作技术外,图像传播技术同样决定着图像所能实现的受众广度和接受度,这是图像在主体间构建意义、运作权力的必要环节。简言之,技术在整体上塑造了图像的制作传播过程,也决定着图像本体的形式表现,使国际政治呈现不断变化的“图像政治面貌”。

图像本体之所以能对现实世界产生影响,根源在于其施为性,即“图像本身能够做事和创造事物,从而在视觉上引发社会秩序的变化及相关实践”^③。事关政治的图像一般承载特定的政治意图,“它们要么倾向于复制主流话语形态,巩固现有的权力、权威和交流体系,要么引导受众以批判性分析的视角看

① Deborah Dixon, *Creating the Semi-living: on Politics, Aesthetics and the More-than-Human*, "Transactions of the Institute of British Geographers", Vol.34, No.4, 2009, p. 413.

② 何伟:《图像与对外政策:艾兰·库尔德和英国关于欧洲难民危机辩论》,《世界经济与政治》2015 年第 12 期,第 80 页。

③ William Callahan, *Sensible Politics: Visualizing International Relations*, New York: Oxford University Press, 2020, p. 33.

待图像,将人们全盘接受的事物变得不再正常,从而为某些议题的政治化提供路径。”^①深伪图像尽管是新鲜事物,但其承载的政治目的仍然无外乎以上两点。

图像本体对国际政治的影响主要通过两条路径,其中一条是认知路径。图像可以充分影响受众认知的范围与深度,“在当今的后文字时代(post-literate age),大多数人从视觉媒体上获取国际事务的信息,因此,图像通过使某些事物可见,同时使某些事物不可见来塑造我们对世界的看法”。^② 图像对受众认知的影响存在启发和误导两个截然不同方向。“启发”意味着为受众呈现真实事例、传递可靠信息,使他们有所知晓、思考乃至领悟。一个典型案例是2015年叙利亚男童在土耳其沙滩俯卧遇难的照片,这一视觉事件令欧洲民众广泛意识到难民问题的严重性,并使他们对难民身份的普遍认知从此前的“入侵者”转变为“人道主义灾难的受害者”^③,图片对民意的启发效应使欧洲多国政府一度大幅调整难民的接收政策。图像是否如实呈现现实世界与其能否发挥启发性影响并不存在必然联系,比如,2020年,外交部时任发言人赵立坚在外网发布了乌合麒麟就澳大利亚士兵在阿富汗割喉杀死两名儿童的罪行所创作的插画作品《和平之师》,以讽刺澳军在阿富汗滥杀无辜的行为;在传统报刊上刊登的政治漫画通常是对现实事态的夸张或虚构性演绎,但仍能为读者揭露常在公众视野之外的真相或深层次矛盾;欧美大量电影和电子游戏画面通过接受主流地缘政治叙事的驯化,帮助西方青少年构建“对地缘政治过去和现在情况的理解”^④。“误导”意味着有意地向受众灌输错误信息,并期望造成不良的负面影响。要发挥图像的误导作用,很多时候不需要专业的篡改或伪造技术,拍摄者有意识的局部取景、间断记录的画面都可以让受众形成错误认知。相比其他感官信息,人类更依赖视觉信息,误导性的视觉信息也比误导性

① Michael Shapiro, *The Politics of Representation: Writing Practices in Biography, Photography and Policy Analysis*, Madison: University of Wisconsin Press, 1988, p. 131.

② William Callahan, *Sensible Politics: Visualizing International Relations*, p. 1.

③ 何伟:《图像与对外政策:艾兰·库尔德和英国关于欧洲难民危机辩论》,第74—91页。

④ Matthew Benwell, “Encountering Geopolitical Pasts in the Present: Young People’s Everyday Engagements with Memory in the Falkland Islands,” *Transactions of the Institute of British Geographers*, Vol.41, No.2, 2016, p. 122.

的语言信息更容易令人信服。^①

另一条路径是情感路径。情感在政治互动中的作用不可取代,“即使在纯粹的理性决策过程中,情感依然会透过微观政治的方式影响政治决策过程,情感不是与理性相对立的个体偏见,而是与公共政治密切关联的核心因素”。^② 国际政治中的情感可以分为积极和消极两类,积极情感包括团结、友好、愉悦、信任等,消极情感包括仇恨、愤怒、悲伤、怀疑等。^③ 而图像可以化约为引发群体性情感反应的诱因,以直击人心的方式打动受众,使个体或群体都能成为情感共同体的一员。这种情感驱动既有助于国际关系行为体之间明晰彼此关系,还能够为一国对外政策争取更多合法性。大众地缘政治学研究较多的一个案例是“9·11”事件图像所产生的情感影响,有观点指出,“9·11”灾难性图像的情感逻辑与美国军事干预的道德逻辑、地缘政治逻辑不可避免地交织在一起,前者的情感强度为后者的明确必要性提供强大支撑;^④ 牛津大学教授托德·霍尔(Todd Hall)认为,“9·11”事件现场画面所产生的同情作用,使一贯反对美国军事扩张的中国、俄罗斯及时表明了对美国一系列对外反恐行动的支持。^⑤

二、深度伪造技术与国际政治图像的生产传播变革

深伪技术的应用过程一般分为三个步骤,首先是数据的收集和提取,为深度学习提供足够的数据喂养;然后是依托生成对抗网络、自动编码器、卷积神经网络、循环神经网络等工具进行模型训练,输入这些工具的数据越多,新的伪造图像就越准确;最后实现合成转换,输出伪造图像。^⑥ 与那些已经获得长

① Camille Koppen, Charles Spence, “Seeing the Light: Exploring the Colavita Visual Dominance Effect,” *Experimental Brain Research*, No.180, 2007, pp.737-754; Markus Prior, “Visual Political Knowledge: A Different Road to Competence?” *The Journal of Politics*, Vol.76, No.1, 2014, pp.41-57.

② 李鹏等:《西方情感地缘政治研究进展》,《热带地理》2021年第6期,第1169页。

③ 尹继武、王海媚:《中国国际政治心理学理论与实践研究的进展与问题:尹继武教授访谈》,《国际政治研究》2017年第6期,第145页。

④ Gearóid Tuathail, “Just Out Looking for a Fight: American Affect and the Invasion of Iraq,” *Antipode*, Vol.35, No.5, 2003, pp.856-870.

⑤ Todd Hall, “Sympathetic States: Explaining the Russian and Chinese Responses to September 11,” *Political Science Quarterly*, Vol.127, No.3, 2012, pp.369-400.

⑥ 赵国宁:《智能时代“深度合成”的技术逻辑与传播生态变革》,《新闻界》2021年第6期,第67页。

期应用的技术(如电脑合成影像、Photoshop)相比,深伪的核心优势是其能够主动实现对图像制作的自我学习,以日趋精细的方式不断提升其与伪造对象的逼真程度。同时,这一技术与互联网社交平台构成天然的技术联盟关系,使图像传播能够充分利用后者的信息级联效应来获得前所未有的曝光度与影响力。深伪技术是推动国际政治进入崭新图像时代的源动力,其全局性的影响主要体现在四个方面:

(一) 创作主体:多元主体的图像共创

过往,国际政治图像的创作主体集中于官方通讯社、权威的电视和报刊媒体,它们在一国政府授权下能够近距离拍摄国际政治的互动场景,然后选择性地向外发布。在这些少数传播主体掌控图像的时代,“复杂的国际政治进展和关切往往被压缩为可控的地缘政治情景”。^① 即使智能手机、数码相机在全球日益普及,全球媒体格局日益去中心化,国际政治图像的创作也远未实现多元化,主流媒体及其背后的各国政府在很大程度上控制着能够到达受众的画面流。

深伪技术有望打破这种集中式的图像生产格局,使国际政治图像的广大消费者转变为生产者,直接参与国际政治的视觉形塑。这样的突破主要基于深伪技术的易获得性。专业人士可以通过网上的开源软件进行伪造,其操作存在一定技术门槛,“需要使用者拥有一定数量的任务图像并在图形处理器(GPU)上进行训练,训练的稳定结果取决于使用者的专业水平”。^② 普通人可以借助商业化的手机或电脑软件,轻易制作人工智能(AI)换脸、人脸合成、语音合成等伪造内容。“随着深伪技术的传播,越来越多行为体能够以一种曾经仅限于好莱坞电影公司或资金最雄厚的情报机构的方式,制作出令人信服的视频内容。”^③ 在近期的一些政治危机中,网民利用该技术创作图像、影响舆论的兴趣已经被调动起来。例如,2023年10月,哥伦比亚广播公司在网络上筛

① Sebastian Kaempf, “Digital Media,” in Roland Bleiker, *Visual Global Politics*, Abingdon: Routledge, 2018, p. 99.

② 李旭嵘等:《深度伪造与检测技术综述》,《软件学报》2021年第2期,第499页。

③ Robert Chesney and Danielle Citron, “Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics,” *Foreign Affairs*, Vol.98, No.1, 2019, p. 150.

选了1000个据称由以色列或巴勒斯坦民众上传的冲突视频,但检查后发现仅10%的视频可以被使用,不可用视频中的一部分都是深伪技术制作。^① 需要强调的是,多元主体指的不仅仅是普罗大众,也包括传统媒体和各类官方或非官方机构。政府机构可以将深伪技术作为新型信息战或对外宣传工具,电视媒体可以利用深伪技术呈现他们未能拍摄到的新闻画面,政党能够利用深伪技术争取民意支持。^② 当然,技术的低门槛和开放性同样适用于从事叛乱、恐怖活动的群体,如恐怖主义组织利用深伪技术开展认知领域的“非对称作战”。据《华盛顿邮报》报道,极端组织“伊斯兰国”(IS)支持者开始使用图像生成和语音克隆等人工智能工具来传播极端主义和招募年轻人,其发布的视频获得了“病毒般”的扩散效果。^③

(二) 图像形态:虚实结合的画面涌现

在人工智能技术尚未普及的时代,国际政治图像的主流形态是呈现真实事件的照片、视频。“图像为我们提供了一个诱人的信念,即我们在其中看到的是世界的真实再现,图像准确地呈现了特定时刻所发生的事情。”^④

当前,深度伪造技术已经创造出大量以国际政治为主题的虚实结合图像,这类图像并非纯虚构图像,它们普遍附着于真实存在的人物(尤其各国领导人和国际名人)、具有较高关注度的政治事件,但同时又能脱离现实参照物的限制,融入创作者自行构思的画面内容。深伪技术既能还原受众较为熟知的场景(如国家领导人的公开讲话、国际危机中的当事人表态),也可以充当政治人物的印象整饰道具(如对视频中的人物进行实时美颜或增强其与观众的眼神

① Matt Lebovic, “These Israelis Are Fighting Hamas on the War’s Emerging ‘Deepfake’ Cyberfront.”

② 2023年,阿根廷总统大选被媒体称为生成式人工智能选举元年,人工智能生成的深伪图像在两个主要政党的竞选活动中发挥了关键作用。两方的竞选团队一边利用 Midjourney、Stable diffusion 等智能图像生成软件快速传播竞选者强大的形象,包括竞选者换脸为罗马皇帝、拳击手的图像和视频,同时也大量生成抹黑对手的图像,好莱坞电影中的大反派成了被换脸的大热门。

③ Pranshu Verma, “These ISIS News Anchors are AI Fakes. Their Propaganda Is Real,” *The Washington Post*, May 17, 2024, <https://www.washingtonpost.com/technology/2024/05/17/ai-isis-propaganda/>, 2024-06-15.

④ David Shim, *Visual Politics and North Korea: Seeing Is Believing*, London: Routledge, 2014, p. 26.

互动),还可以颠覆行为体的传统公共舞台,演绎一些较少公开的场景(如官员密室协商或私人谈话),打破国际政治中固化的视觉原型(visual prototype)和仪轨表象。在真伪程度上,深伪技术既可以制造尽可能贴近现实的清晰画面,追求“以假乱真”“真假难辨”的镜像效果,也可以制造超现实或反现实的画面,如逻辑天马行空的网络梗图、能被多数人一眼识破的搞笑视频等,这类画面从不追求对现实世界的精准复刻,更多时候是为开展解构性、去权威化、戏谑化的意象表达。在画面风格上,深伪技术可以提供电视新闻、视频监控或偷拍、美颜相机、历史影像等多种风格,增强图像传播的说服力。虚实结合图像的涌现不应简单定性为冲击国际政治认识论或行为体本体安全的“洪水猛兽”,这种情况是人工智能、赛博空间与国际社会共生互嵌的必然结果。

(三) 传播平台:高度依赖网络社交媒体

深伪技术的应用始终高度倚赖网络社交媒体。首先,使用该技术的行为体普遍将制造影响的希望寄托于网络社交媒体的信息级联效应,即“社交媒体平台能够使各类信息出现滚雪球般的集聚、扩散,而且社交媒体传播的信息往往还会蔓延到传统的大众媒体,使后者报道他们未曾注意的信息”。^①近年来受到全球关注的深度伪造事件基本发端于社交网络,在网络世界得到较多关注和转发后,传统媒体才纷纷跟进报道。其次,深伪技术的图像伪造逻辑契合社交媒体平台的信息筛选逻辑。平台的“算法为了回应人类对刺激的渴望而力推那些夺人眼球的事物,而能够夺人眼球的事物往往也是戏剧性的、出人意料的和有感染力的”。^②深伪图像对现实的再造或颠覆符合上述特质,容易在信息过载的平台中脱颖而出,获得平台的优先推送,这一过程本身也在反哺社交媒体的渠道霸权。再次,社交平台的“网络模因”(internet memes)特质有利于技术的推广,模因指“在用户不断模仿、完善的情况下,在互联网上快速传播

① Bobby Chesney and Danielle Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,” *California Law Review*, No.107, 2019, pp. 1765-1766.

② [美]亨利·基辛格等:《人工智能时代与人类未来》,胡丽平等译,北京:中信出版社2023年版,第239页。

的具有感染性的图像、视频和流行语”，代表着网民受社交活力驱动的文化体验。^① 深伪技术推陈出新的功能在社交平台获得展示后，很容易吸引网民“拍同款”“挑战翻拍”等行为，使技术的应用成为流行的风潮。

深伪技术与社交网络平台的结合无疑对国家的官方形象传播构成了新的挑战。深伪技术加速了各类图像生产和更新的效率，而且在社交平台上能够更匹配人们高度分散、稍纵即逝的注意力。面对这种趋势，各国政府需要以更快速度、更高质量、更强吸引力来发布反映真实情况的国际政治图像，在日益激烈的图像竞争中突出重围。

（四）大国博弈：图像技术竞争主流化

长期以来，图像技术竞争并非大国博弈的突出内容，人们更多只将其视为对外传播中的技术性环节。而深度伪造技术的普及预示着图像技术竞争的主流化，各国尤其大国都不愿在这一新兴技术领域上落后于人。

深伪技术正在塑造各国决策者对新型信息战的认知，改变他们对于虚假信息供给的成本、实施路径和风险等方面的预期。加拿大国家安全情报局的报告指出西方国家都可能利用深伪技术向敌对国家传播错误信息或窃取情报，“因为该技术具有成本效益，并且可能比其他更危险的情报收集形式或秘密活动更容易”。^② 美国海军战争学院的约翰·特拉马佐(John Tramazzo)认为，“深度伪造技术将给作战指挥部带来超越敌军的优势，并为机动部队提供保护……这一技术在本质上并非不道德或不合法”。^③ 可以试想，技术竞争最恶劣的前景是相互竞争的国家在这一灰色地带进行有针对性的长期准备，一国在重要危急关头向敌国散布极具信服力的深伪作品，并结合网络攻击面向

① 夏德元、王宇博：《模因论视域下短视频对外传播的媒介逻辑》，《新闻爱好者》2021年第1期，第36页。

② Canadian Security Intelligence Service, “Finding Signals in the Synthetic: Intelligence in the Era of Deepfakes,” Government of Canada, November 16, 2023, <https://www.canada.ca/en/security-intelligence-service/corporate/publications/the-evolution-of-disinformation-a-deepfake-future/finding-signals-in-the-synthetic-intelligence-in-the-era-of-deepfakes.html>, 2024-06-15.

③ John Tramazzo, “Deepfakes and Deception: A Framework for the Ethical and Legal Use of Machine-Manipulated Media,” Modern War Institute, July 28, 2023, <https://mwi.westpoint.edu/deepfakes-and-deception-a-framework-for-the-ethical-and-legal-use-of-machine-manipulated-media/>, 2024-06-15.

对方公众进行广泛散播,这很可能将危机从可控的冲突级别推升到更危险、不可控的冲突级别。

在实践层面,西方国家针对深伪技术的最新动向都在加剧竞争激度,表现出“既防范威胁、也制造威胁”的自相矛盾:一是加大对技术的研发及其应用效果的评估。美国参议院国土安全及政府事务委员会在2019年通过《深度伪造法案》,要求国土安全部每年评估深度伪造的威胁,指出“应该主动识别对手开发深伪技术和内容的工具,并制定识别和打击深伪内容的对策”。^①在防范威胁的同时,美国特种作战司令部在2022年的技术采购清单中加入了深度伪造,以“在战术优势和作战层面开展影响力行动、数字诈骗、通信中断和虚假信息等活动”。^②二是实施基于深伪技术的舆论战,如在涉疆问题上,“西方媒体制作了包括深度伪造在内的大量虚假报道,极力抹黑中国政府的治疆方略和政策”;在涉港问题上,美英媒体也曾利用深伪视频向外界传递对香港特区政府和香港警方的不满。^③三是西方国家内部的政治斗争、竞选活动已经频繁使用该技术,这些服务国内政治需求的实践也可以帮助欧美累积在国际政治环境中的技术实力,服务于未来的地缘政治应用。

三、深度伪造图像对国际政治的影响

根据目前制作和传播的深伪图像看,按照图像本身所承载的政治诉求,本文将深伪图像分为以下四种类型:

(一) 宣传性图像

深伪制作的宣传性图像正在成为全球公共外交的信息载体,能够启发受众了解具体议题和倡议,争取受众的积极认同。一个典型的应用场景是,国际

① “Deepfake Report Act of 2019,” Report of the U. S. Senate Committee on Homeland Security and Governmental Affairs, September 10, 2019, <https://www.congress.gov/116/crpt/srpt93/CRPT-116srpt93.pdf>, p. 3, 2024-06-15.

② Sam Biddle, “U. S. Special Forces Want to Use Deepfakes for Psy-Ops,” *The Intercept*, March 6, 2023, <https://theintercept.com/2023/03/06/pentagon-socom-deepfake-propaganda/>, 2024-06-15.

③ 刘国柱:《深度伪造与国家安全:基于总体国家安全观的视角》,第25—26页。

关系行为体利用名人的深伪画面,向外开展国家形象宣传或国际公关活动。例如,2019年,法国慈善机构“团结艾滋病”制作了一段时任美国总统特朗普宣称“艾滋病已经终结”的深伪视频,呼吁西方国家加大对全球抗击疾病的资金支持。同年,英国一家健康慈善机构制作了球星大卫·贝克汉姆号召全球抗击疟疾的公益广告,视频中的贝克汉姆在55秒的发言中迅速切换了九种语言,这是通过深伪技术制造的画面,他的嘴型与不同语言的分层音轨实现了高度同步,这样的表现形式吸引了全球舆论注意,达到远超预期的宣传效果。另外可以预见,越来越多国际关系行为体将借助“虚拟发言人”“虚拟主播”等深伪人物形象,与其他生成式人工智能(如ChatGPT)结合,开展全天候、多语种及可互动的信息发布,开展贴合多元文化和诉求的公共外交。

(二) 娱乐性图像

涉及国际政治题材的电影、电视剧、短视频、网络梗图、电子游戏已经大量使用深伪制作的图像。在多数情况下,这些图像可以发挥启发受众、唤起积极情感的正面价值。第一,深伪图像可以大幅提升各类娱乐产品的现实还原能力,增强受众在文化体验中的临场感、参与感,进而反作用于他们的自我观念结构和行为决策。例如,影视作品能够更好还原历史人物和关键事件,有助于观众对国际关系的历史或要素形成清晰认知。再如,在电子游戏中,用户可以换脸扮演不同身份的“政治家”“外交家”,设身处地体验“一线外交活动”。深伪图像所赋予的虚拟身份越可信、鲜活,越有利于“个体将‘可能自我’内化到真实的自我概念中”。^①当游戏者被置身于国际危机处理中,将更有可能形成对风险的具象化认知,增强维护国家安全的公民意识;当游戏者将个人形象嵌入到外交官的职业身份中,并经历成就情景,更可能培养他们参与国际事务的热情。第二,再造现实的图像可以在影视作品、游戏中演绎丰富的国际政治叙事,使创作者和受众都能在弱规定性、强思辨性的图像世界中释放对世界秩序的印象。例如,国内短视频平台上已有不少“外国名人说中文”的作品,这类作

^① 张铮、吴福仲:《拟身链接:基于深度伪造技术的人机互动及其社会善用可能》,《苏州大学学报》(哲学社会科学版)2022年第2期,第172页。

品蕴含了网友们对中国文化“走出去”、扩大国家软实力的期许,也呈现了未来国际交流具有较高实现概率的新图景。第三,此类图像能够促进国际政治议题上的大众沟通。各国网友们以“娱乐”为名、在社交平台上发布和传播图像,有利于突破西方媒体产业控制下的信息失衡格局,增强不同区域和国家之间的民意对话和共识塑造。不可否认,深伪制作的娱乐性图像也会延续此前业已存在的问题,包括抹黑特定群体、恶意诠释某些议题、为霸权国对外政策辩护等。

(三) 批判性图像

批判性图像是国际政治领域的显著文化现象,包括人们早已习以为常的政治漫画、模仿秀、行为艺术、涂鸦艺术等。总体来看,深伪制作的批判性图像可以发挥启发认知的作用,“提供探索事件因果关系的竞争性叙事的机会,批评(国际政治领域)仓促做出的借口和遭受扭曲的结论,并促进基于好奇心、谦逊和情之所至的非暴力安全范式”。^①

流行于社交网络平台的深伪短视频在政治批判上存在显著优势,“相比漫画、梗图和动态图(GIF),深伪视频可以超越前者简单和静态的描述来吸引受众,并扩大叙事能力;相比电影和电视剧,深伪视频会表现它们不具备的简洁优势”。^② 在油管(YouTube)平台上,目前存在不少专门发布政治类深伪短视频的账号,如巴西艺术家布鲁诺·萨托里(Bruno Sartori)制作了不少批评时任巴西总统博索纳罗内外政策的深伪视频,在网络上引起较大反响;在一段法国网友创作的视频中,法国极右翼政治家玛丽娜·勒庞(Marine Le Pen)穿戴穆斯林妇女头巾,用阿拉伯语讲述自己的仇外政策。

批判性的深伪短视频已经成为国家间相互指责的工具。例如,2020年9月,今日俄罗斯电视台的网站制作了一段以特朗普为主角的深伪短视频,伪造的特朗普演绎“收受俄罗斯各种好处、未来前往俄罗斯工作”等情节,反讽美国政坛对“俄罗斯介入美国大选”的无端指责,在社交网络上引起热议,甚至有美

^① Bryan Taylor, “Defending the State from Digital Deceit: The Reflexive Securitization of Deepfake,” *Critical Studies in Media Communication*, Vol.38, No.1, 2021, p. 10.

^② Bryan Taylor, “Defending the State from Digital Deceit: The Reflexive Securitization of Deepfake,” p. 11.

国媒体信以为真、当成了一个真的新闻进行报道。对此,今日俄罗斯的网站表示美国媒体“如果真的注意到这段视频,就会看到不断出现的模仿免责声明,如果没有看到,说明他们只会看他们想看到的东西”。^①再如,2022 年 4 月,乌克兰政府与该国人人工智能企业 Reface 合作发布了一段 1 分 15 秒的深伪视频,显示“普京在战火蹂躏下的马里乌波尔街头讲述俄军罪行”,借以声讨俄罗斯的军事行动。

在情感影响上,不少深伪的批判性图像都融入了大量喜剧、夸张或恶搞的文化元素,容易唤起受众的积极情感反应。加深对图像诉求的认同。同时也有一些图像试图激发、散播气愤、悲伤等消极情感。例如,2021 年,黎巴嫩电视台在贝努特大爆炸事故发生一年后,在征得家属同意的情况下,制作一段遇难者控诉灾难处置不力、批评国际社会未能给予支持的深伪视频。

宣传、娱乐和批判三类图像虽然在多数情况下,其创作初衷可能并不包含恶意,但存在误导受众、制造国际政治矛盾或冲突的风险。第一,这三类图像面临共同的“视觉安全困境”,即图像传递的信息具有先天的模糊性,容易引发受众的误读或不满。例如,在传统的印刷媒体中,政治批判图像往往出现在报刊特定的专栏、社论、漫画等板块,拥有一国相对稳定且往往具有基本社会共识的读者群;但在跨文化、碎片化的网络空间,深伪图像所发起的批判很可能被某些群体视为“有意的冒犯”“精心设计的挑衅”。第二,尽管大量用于宣传、娱乐和批判的深伪图像都明确标记了自身性质,提醒受众“不能当真”,但视觉真实性的增强容易导致受众信以为真。在乌克兰危机爆发后,一些欧洲媒体或网民误将战争类电子游戏《武装突袭 3》的视频片段作为俄乌战场的画面;深度伪造打造的历史影视剧可能加剧重大公共事件记录的不稳定性,乃至破坏不同国家、不同民众对国际交往事实的共识。传播学的研究还显示,即使受众知道一段丑化政治人物的视频是由深伪制作的,也依然会在观看影像后之后

^① Nebojsa Malic, “‘Bizarre & Deeply Disturbing’: MSM Think that RT’s Trump Deepfake Parody Is Admission He Is ‘Putin’s Pawn’,” *Russia Today*, September 23, 2020, <https://www.rt.com/news/501527-rt-trump-deepfake-reaction/>, 2024-06-15.

产生对该政治人物的负面印象。^① 第三,网络空间大量事关国际政治的深伪图像奉行“蹭热度、引流量”的商业逻辑,并没有理性讨论公共事务的意图,它们猎奇式、狂欢式的演绎会消解某些议题的严肃性,令受众无暇顾及议题的实质。

(四) 欺骗性图像

欺骗性图像以误导目标受众为核心目的,在情感维度上往往试图引燃群体性的消极情感。欺骗性图像的国际政治影响主要可以分为以下三方面:

1. 误导政府部门的决策或行动。从现有实践和未来趋势看,各国政府部门都面临立足深伪图像的欺骗或干扰威胁。具体表现包括:一是发起欺骗性会谈。继新冠肺炎疫情之后,世界广泛应用网络平台举办线上国际会议,一些不法分子或好事者将深伪图像作为招摇撞骗的“面具”组织官方会议,表达政治诉求。例如,2022年6月,有人用深伪假扮乌克兰基辅市长克利钦科,先后与柏林、马德里和维也纳市长进行了视频通话,在与柏林市长的交流中,伪装者谈论了乌克兰难民骗取德国政府福利的问题,呼吁德国将难民送回乌克兰。2023年4月,冒充乌克兰总统泽连斯基的俄罗斯喜剧演员通过深伪画面,与美联储主席鲍威尔进行了视频通话。2023年10月,网络不法分子利用非盟主席法基的深伪形象,与一些欧洲首都官员进行了视频通话。

二是伪造政令或军事命令。在军事行动中,战场上的对手可以生成高级指挥官的深伪图像,必要时在网络攻击的配合下,向对手部队下达虚假或矛盾的指令。2023年11月7日晚间,乌克兰社交网络出现一段视频,内容是乌克兰武装部队总司令扎卢日内将泽连斯基称为叛徒,并号召军队向基辅进军。乌克兰政府的反虚假信息中心迅速做出警告,表示这是俄罗斯方面“制造乌克兰政府和军方之间分裂”的深伪作品。

三是用于窃取信息的情报工作。深伪制作的人脸图像已经具备欺骗面部识别系统的能力,不法分子可以利用深伪图像绕过生物认证系统,获得对安全设施和敏感个人信息的未经授权的访问。

^① Tom Dobber and Nadia Metoui, et al., “Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?” *The International Journal of Press/Politics*, Vol.26, No.1, 2021, pp. 69-91.

四是降低一国的态势感知能力。欺骗性深伪图像的泛滥“使决策者的战略视野变得浑浊”,他们只能相信一手的情报报告和内部简报,排除外部尤其社交媒体上的信息或反馈。^①另外,开源信息已经成为全球情报搜集的重要渠道,但在国际激烈博弈的情况下,一国可以有意散布大量深伪图像,降低开源信息池的可靠性,增加情报信息筛选和验证的难度。

2. 挑拨不同行为体之间的关系。欺骗性图像能够通过污名化、妖魔化画面,触发行为体恐惧、愤怒、焦虑等应激性情感,达到制造心理隔阂、破坏各种关系的目的。其实施路径包括:一是制造一国社会的内部分裂。美国研究者们担心以俄罗斯为代表的敌对国家利用美国的种族矛盾,专门制作和传播一些煽动对立态势的深伪视频,如“白人警察在枪杀手无寸铁的黑人时高喊种族歧视言论”。^②再如,企图破坏一国稳定的敌对分子通过捏造目标国的政府官员从事不法行为、发表不当言论的深伪图像,能将民众对政府的潜在不满转化为显性的动荡。二是恶化不同国家之间关系。欺骗性图像可以呈现事关国际政治的各种负面场景,如“一国高官私下批评他国政策”“某国驻外军队粗暴使用武力”“外交官对驻在国表达轻蔑态度”,这类画面可能对特定受众造成情感伤害,抹黑一国在海外的国家形象和声誉。欺骗性图像也可以放大国际分歧或矛盾,限制一国政府在外交中的行动选择,如在国际峰会期间,“一个精心策划的深伪骗局能够煽动公众舆论,使一方在政治上站不住脚,无法像正常情况那样推进自己议程,或使达成某项合作协议的成本过高”。^③另外,同其他类型的虚假信息一样,欺骗性图像还可以扭曲受众记忆,使他们更改对历史事件的看法甚至将未曾发生过的事件视为过去的一部分,^④这都可能制造国际社会的

① Marina Favaro and Heather Williams, “False Sense of Supremacy: Emerging Technologies, the War in Ukraine, and the Risk of Nuclear Escalation,” *Journal for Peace and Nuclear Disarmament*, Vol.6, No.1, 2023, p. 40.

② Daniel Byman, et al., *Deepfake and International Conflict*, The Brookings Institution, January 2023, https://www.brookings.edu/wp-content/uploads/2023/01/FP_20230105_deepfakes_international_conflict.pdf, p. 8.

③ Bobby Chesney and Danielle Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,” p. 1782.

④ 在2013年发表的一项研究中,近5300名受试者被问及他们对3个真实的和5个伪造的政治事件照片的记忆,大约一半受试者都表示记得伪造图像表现的事件曾经发生过,27%的人表示曾在电视里看过伪造事件的报道。研究显示了捏造的政治事件通过图像可以成为受众普遍的错误记忆,参见Steven Frenda, et al., “False Memories of Fabricated Political Events,” *Journal of Experimental Social Psychology*, Vol.49, No.2, 2013, pp. 280-286.

共识坍塌与无谓争辩。

3. 恶化公共信息环境。欺骗性深伪图像的泛滥容易衍生不可靠的信息环境,使受众逐渐失去对真实媒体的信任。有传播学学者对乌克兰危机爆发后几个月的欧洲网络舆论进行了调查,发现欧洲网民普遍察觉到关于战事的深伪欺骗图像的盛行,不少言论表示深伪图像助长了“不健康的怀疑与恐惧”,担心“冲突中的任何信息都变得不可信,带有恶意的机构总在提供虚假叙事”。^①当人们对信息环境缺乏起码信赖时,受众的动机性偏差(motivational biases)就会随之增强,他们容易相信那些符合既有信念和预期的伪造画面,而将不喜欢的信息视为“假消息”,将自己置身于主观构造的信息茧房。基于这种情况,欺骗性图像的盛行很容易撕裂社会鸿沟和造成政治极化。

在信息环境恶化的背景下,别有用心的人还可以将反映真实信息的图像也谎称为深伪的欺骗性图像,赚取所谓的“骗子红利”。例如,2018年下半年,加蓬时任总统阿里·邦戈(Ali-Ben Bongo)被传中风,连续数月未公开露面;2019年1月,加蓬官方发布了一段邦戈发表新年致辞的视频,但该国反对党宣称这是“掩盖邦戈真实身体状况”的深伪视频,尽管该视频事后被证明并非伪造,但它最终成为引发该国爆发军事政变的导火索。再如,2023年10月巴以冲突爆发后,以色列官方一改此前不愿分享战争图像的做法,透过该国总理的社交媒体账号发布了遇难婴儿的照片,其本意是争取全球对以色列的同情和对哈马斯的愤怒。在社交媒体上,一些人在毫无根据的情况下指责以色列政府发布的照片是深伪制作的,并骗取了很多机构和网民的认同,但这些照片经专业鉴定并无数字篡改痕迹。这凸显了未来对外信息传播的一个困境,即真实的信息总是需要提供“自证真实”的证明。

总的来看,深度伪造使图像超越了传统的、主要的“反映叙事”功能,成为“创造叙事”的主体,使“政治的图像”获得更多衍生“图像的政治”的机会,图像本体有望频繁成为国际合作与冲突的发轫点。深伪图像对受众认知、情感上的影响是多维的,既要看到其在虚实杂糅中传递真实信息、建构良性国际政治

^① John Twomey, et al., “Do Deepfake Videos Undermine Our Epistemic Trust? A Thematic Analysis of Tweets That Discuss Deepfakes in the Russian Invasion of Ukraine,” *Plos One*, Vol.18, No.10, 2023, pp. 1-23.

观念结构的潜力,也不能忽视其破坏国际交往理性氛围、催生国际冲突的风险,即使深伪图像服务于善的目的,也可能存在混淆视听、损害媒介信誉、引发非预期冲突等不利影响。另外,深伪图像对群体情感的召唤潜能可能导致感性政治在国际社会大行其道,图像的创造者会深入挖掘目标受众的心理空间,人们对国际政治的思考在纷繁图像的牵引下将更多遵循“喜欢不喜欢”(而非“合理不合理”)的情绪性逻辑。

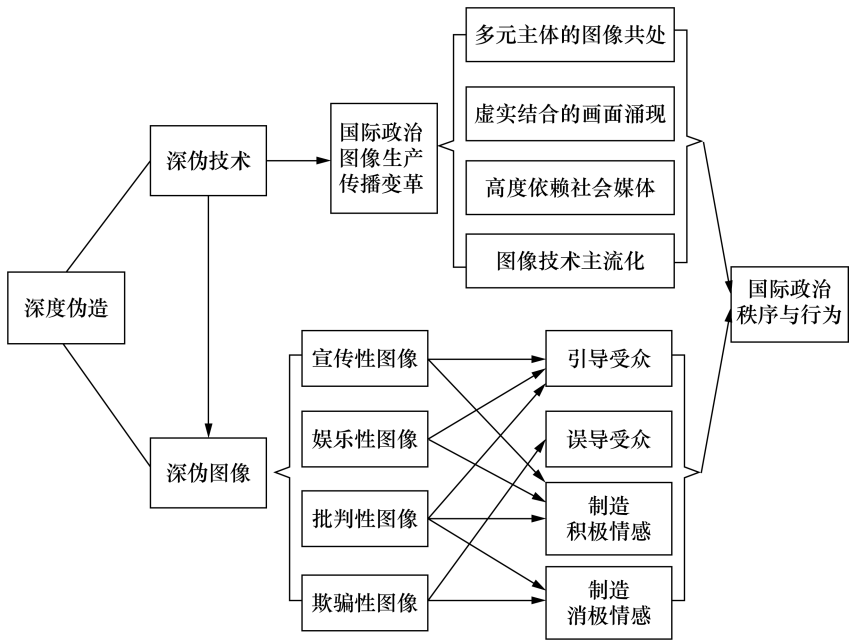


图2 深度伪造对国际政治的影响框架

图表来源:笔者自制。

四、深度伪造的制约因素

深伪技术和图像要在国际政治中发挥影响,无论正面还是负面的影响,都面临一系列的制约因素。

(一) 高质量深伪存在较高技术门槛

尽管深度伪造图像的制作已经实现平民化,但高质量、高仿真的深伪图像

仍需要较高的技术条件,包括高成本的专业设备、大量训练数据、具有丰富经验的技术人员等。即使具备上述要素,一个逼真的深伪视频制作往往也需要数月甚至更长时间。而且,现有多数的深伪技术仍难实现高分辨率下的人脸深度篡改,“轻微的样本扰动或参数设定变化就可能导致合成的样本效果变差,同时该技术还存在易产生异常斑点、背景出错、人脸属性不对称等弱点”。^①另外,网络空间的信息战不能被理解为静态时空环境下的单一事件,一个特定网络行动的实施有赖于其他网络技术、其他领域技术的配合和支持。深伪图像如要应用于规模化的信息战,必须依托高超的网络入侵技术,令图像能够在对象国的公共媒体或重要网络账号上得以广泛发布。

(二) 社会文化环境限制传播效果

在国际政治中,图像从来无法独立发挥作用,它们能否发挥影响、发挥多大影响受制于一国文化环境。受众对画面要素的认知程度差异、对内涵的跨文化和情境化解读,很多时候都会消解深伪图像创作者们精心设计的意义体系,使引导或误导的行为收效甚微。另外,欺骗性图像在不同的政治制度、政治文化下有着差异较大的效能,据现有研究显示,政治两极分化的社会比分化程度较低的社会更容易受到虚假信息的操纵;相比多党议会内阁制国家,采取“赢者通吃”总统选举制度的国家更容易受到外国虚假信息的干扰;拥有大量散居少数民族的国家比族群上较为同质的国家对欺骗性信息更缺乏鉴别力。^②

在国际冲突或争端的环境下,深伪图像未必是发动舆论战、信息战的优先选择。无论是乌克兰危机还是巴以冲突,尽管网络世界涌现出大量深伪图像,但真正误导决策者或者引起国际舆论高度关注的少之又少。一个重要原因是冲突现场的真实图像已经足够震撼,冲突方政府、持不同立场的民众仅凭“无需加工”的图像,就足以支撑他们的诉求和主张,甚至现实画面的残酷超越了

① 蔺琛皓等:《虚假数字人脸内容生成与检测技术》,《计算机学报》2023年第3期,第483页。

② Carme Colomina, “Words as Weapons: from Disinformation to the Global Battle for the Narrative,” Barcelona Centre for International Affairs, September 2022, https://www.cidob.org/en/articulos/cidob_report/n_8/words_as_weapons_from_disinformation_to_the_global_battle_for_the_narrative, 2024-06-15.

深度伪造创作者们的想象。

(三) 行为体具有防范深度伪造风险的能动性

长期以来,人类社会都在适应不同形式的欺骗,国际关系行为体至少会从以下方面增强对深度伪造负面影响的防范。

其一,即使没有深度伪造的出现,公众对互联网信息的可靠性已经抱有审慎态度、不会给予无条件信任。随着公众对深度伪造的了解和接触增多,不少人对那些与现实或常识存在较大反差的国际政治图像将保持警惕,在信息获取上更加倚赖权威的信息渠道。有研究指出伴随互联网成长起来的Z世代更加偏爱有公信力的专业新闻机构,尽管“这些数字原住民使用社交媒体作为他们的主要在线联系方式和新闻来源,但他们以怀疑态度对待内容,不太愿意接受流于表面的信息,也热衷参与揭穿谎言,更倾向于核实他们遇到的信息”^①。

其二,多国政府对深度伪造已经启动必要的治理,划定该技术合法合理使用的范畴,明确创作者的标识义务,加强与网络社交平台的共治共管,提高执法、情报、军事等机构对深伪相关威胁的认识及应对。随着应对技术的成熟,各国政府势必也将推广人工智能自动化检测、基于数字水印或区块链的溯源技术等。此外,深度伪造所创造的、令人不悦或感到冒犯的国家形象可能激发不同层次的国际关系行为体强化形象建设,甚至在一种对比中更加凸显自身形象的稳定性和鲜明性。

其三,在关键时刻,一国可以对深度伪造的威胁采取常态化反击。例如,在乌克兰危机爆发后,乌克兰国家安全和国防委员会下属“反虚假信息中心”(CCD)对深伪图像威胁保持高度戒备,除了对乌国政治家的深伪图像及时澄清外,也对流传于网络空间的各类欺骗性图像进行大面积的曝光和清除,既提升了乌国在信息战中的全民抵御能力,也在国际舆论上占据主动。

可以展望的是,国际关系行为体的适应可能衍生新的问题。为了规避深度伪造的不利影响,各方在国际政治互动中将更加依赖多形态的实时记录,确

^① Benjamin Jensen and Divya Ramjee, “Beyond Bullets and Bombs: The Rising Tide of Information War in International Affairs,” CSIS, December 20, 2023, <https://www.csis.org/analysis/beyond-bullets-and-bombs-rising-tide-information-war-international-affairs>, 2024-06-15.

保未来能对事件过程做到全面复原、有效追溯,进而产生国际政治的“凡事留痕”文化。这种情况除了增加工作量外,也会导致变相的过度监管,使国际政治中的一线谈判者保守谨慎、行事缺乏灵活性。各国政府在开展对外行动(尤其重大行动)时加强对图像的主动披露,也与外交活动中必要的保密需求和低调行事风格相悖,可能导致更多交往变数或人为制造的风险。

结 语

本文揭示了深度伪造正在开辟国际政治的新图像时代,其技术和图像会增加国际政治活动的交互性、复杂性,对国家间关系、国家安全和形象管理制造诸多不稳定因素,霸权国势必会将深伪技术作为其发动“叙事战”、维持影响力的战略传播工具,但从现有实践上看,并不会出现“摧毁真实世界、紊乱国际体系”的灾难性后果。

深度伪造的国际政治影响远未充分显现,随着技术不断成熟、应用持续增多。笔者认为有以下三方面趋势需要重点关注:一是各国尤其大国将如何定位深度伪造在信息战中的角色,各国是否会掀起以深度伪造为中心的信息“军备竞赛”。从近期国际冲突看,深伪图像虽然在误导敌方决策和民众观念上并没有预期般强大,揭穿深度伪造计谋的难度也不大,但这并不意味着其破坏潜能止步于此。相比冲突时的欺骗手段,笔者认为更需关注深度伪造在非危机状态下的灰色作战实力,即国际关系行为体可以通过持续向目标受众传播深伪图像(不限于欺骗性图像),以渐进渗透的方式损害一国的舆论环境与信息安全。二是深度伪造将引发国际政治的图像竞赛,图像的数量增多与形态翻新意味着受众的关注成为“稀缺资源”,这场竞赛既包括不同深伪图像之间的竞争,也包括深伪图像与纯写实图像之间的竞争,图像竞赛最终将指向国际关系行为体对视觉政体主导权的争夺,即能否在国际社会建构“一套以视觉性为标准的认知制度甚至价值秩序”^①。三是大众可以借助深度伪造与地缘政治发

^① 王雪晔:《图像与情感:情感动员实践中的图像框架及其视觉修辞分析》,《南京社会科学》2019年第5期,第122页。

生更多勾连与耦合,同时引发官方叙事与民间叙事的角力、不同民间叙事之间的碰撞。大众关于国际政治的超现实呈现虽然难以撼动国际秩序的既有物质和权力结构,但势必会给主流的政治视觉创作带来冲击,为国际社会想象共同体的构建提供新的进路。

与其他人工智能技术一样,深度伪造需要团结、高效、讲求科学与传播规律的全球治理。2024 年 6 月,习近平主席提出:“坚持以人为本、智能向善,在联合国框架内加强人工智能规则治理。”^①当前,相关的治理进程尚处于起步阶段,各国政府、国际组织、数字平台科技公司等纷纷强调国际合作的必要性,开展了明确治理重点、提出技术方案、搭建多边机制等行动。2023 年 6 月,联合国发布专门报告,指出深度伪造造成的虚假信息风险“正在加剧”,呼吁人工智能的利益相关方采取“紧急和立即”的行动;^②古特雷斯秘书长多次呼吁组建类似于国际原子能机构的国际人工智能监管机构,推进各国的紧密协作。2024 年 2 月,包括抖音、脸书母公司 Meta、ChatGPT 母公司 OpenAI 在内的 20 多家全球领先科技公司在慕尼黑安全会议上做出共同承诺,包括努力阻止不良行为者制造有害的深度伪造作品、共享人工智能生成图像的自动标记技术等。同时,需要注意的是,西方国家企图跳过中国“抱团治理”、主导议程的趋势十分明显,如“七国集团”于 2023 年 5 月组建了应对虚假信息的快速响应机制,欧美不少政治人物都在呼吁优先制定“民主阵营”的国际行为准则。在这种背景下,中国需要克服更多阻力,坚持“以人为本”、平等互利等原则,把握人工智能全球治理处在初级阶段的契机,从技术应对、平台管制、公众舆论素养等多维度出发开展国际合作,将治理重点引导到严峻的国际公害上(包括运用深伪技术的跨国犯罪、恐怖主义宣传、色情视频、网络诈骗、虚假新闻等),主动参与国际规则制定,抵制深度伪造武器化的趋势,促进其能够“智能向善”,确保人工智能始终朝着有利于人类文明进步的方向发展。

① 新华社:《习近平向联合国贸易和发展会议成立 60 周年庆祝活动开幕式发表视频致辞》,《人民日报》2024 年 6 月 13 日,第 1 版。

② 联合国:《〈我们共同的议程〉政策简报 8:数字平台上的信息完整性》,2023 年 6 月,<https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-zh.pdf>, 2024-06-15。