 DOI: https://doi.org/10.12349/iser.v7i4.10055

On the Potential Negative Impact and Ethical Challenges of Deepfake Technology on Memory

Yuhuan Liang

Xiangtan University, Xiangtan, Hunan, 411100, China

Abstract

Leveraging deep learning algorithms, deepfake technology can generate highly realistic fake videos. This paper uses Bartlett's schema theory and dual encoding theory as the analytical framework to reveal the potential threats that deepfake poses to individual and collective memory. The paper argues that the threat of deepfake extends beyond information falsification and penetrates the memory reconstruction process through psychological mechanisms, thereby disrupting the original individual and collective memories. Based on the analysis in this paper, addressing this challenge requires the combination of real-time detection technology and ethical review to maintain the authenticity of memory and the social narrative justice.

Keywords

Deepfake; Memory Reconstruction; Personal Memory; Collective Memory; Ethical Challenges

论深度伪造技术对记忆的潜在负面影响及伦理挑战

梁郁涵

湘潭大学，中国 · 湖南 湘潭 411100

摘　要

深度伪造技术依托深度学习算法，能够生成高度逼真的虚假视频，本文以巴特莱特的图式理论与双重编码理论为分析框架，揭示深度伪造对个体记忆与集体记忆构成潜在威胁。本文主张，深度伪造的威胁并非止于信息造假，而是通过心理机制深度介入记忆重构过程，破坏原有的个体与集体的记忆。基于本文分析表明，应对这一挑战需要结合实时检测技术与伦理审查，以维护记忆的真实性与社会的叙事公正。

关键词

深度伪造；记忆重构；个人记忆；集体记忆；伦理挑战

1 引言

当前，以人工智能为主体的信息技术迅速发展，在不同领域并通过不同方式席卷了人类生活。与此同时，互联网早已成为人们获取信息和构建自我经验的重要媒介，因此网络信息的存在也成为了人们认知和生活的一部分。过去，信息的造假形式仅限于文字舆论的“造谣”或简单的修图技术。尽管这些伪造信息曾带来误导，但它们很容易通过简单的事实说明或技术鉴定被推翻，其欺骗性相对有限。然而，新兴的深度伪造技术依托于深度学习算法，能够制作出高度逼真的视频，展现人们说或做一些他们并未说或做过的事情。显然，在这里“伪造”这一术语本身我们就能得到欺骗的性质，同时又依托视频这一有力载体，人们将面对高度逼真的伪造影像，其原有的记忆重构过程将被影响，进而会引发集体记忆的扭曲。

【作者简介】梁郁涵（2005—），女，中国河北石家庄人，在读本科，从事哲学研究。

2 理论基础：记忆的可塑性与视频的证据效力

记忆本质上是动态的重构，而非静态的存储。当我们进入回忆时，难以确保得到的结果是当时发生的原本模样。大脑与计算机的存储机制并不相同，记忆无法做到准确无误地输入和输出。

2.1 记忆的可塑性与重构本质

根据弗雷德里克 · 巴特莱特提出的图式理论，记忆是基于“图式”进行构念，而非简单的再现。“图式”是活的、经久不息地发展的“安排形式”，涉及对过去的反应或经验予以一种积极的组织 [1]。这意味着个体的记忆包括了对过去的经验进行主动的、创造性重建。换言之，记忆包含一个重构的过程。记忆的这种重构性，使其具有可塑性和脆弱性，为外部信息的介入提供了可能。记忆的可塑性具有神经学基础。如果在记忆形成的时间内增加其他信息，那么这段记忆就可以被改写，这段时间是大脑在自己的生物部件和分子组成中重构记忆所需的时间 [2]。伊丽莎白 · 洛夫特斯的实

验研究进一步证明了这一点，在事件发生之后引入的误导性信息，能够系统地与原始记忆融合，改写甚至取代部分真实记忆。

2.2 视觉证据的优势

著名心理学家Allan Paivio基于认知心理学的研究成果，提出了双重编码理论，认为人类的信息加工与记忆过程中存在两个相对独立却又相互关联的系统：一个用于语言编码，另一个则负责非语言（主要是视觉）编码。我们知道视频这一载体同时提供了非语言（视觉）和语言（听觉）信息，这使得信息能够以双重方式被编码和存储在大脑中，观看者可以同时接收到“看到”的事实和“听到”的解释，这种双重编码加上感官的相互印证增加了信息的真实感和可靠性，所以视频是最具可信度的载体之一，正如人们常常认为的眼见为实。此外，双重编码比单一编码更不容易被遗忘，提高了信息的留存率和持久性。

3 作用机制：深度伪造如何干扰记忆重构

深度合成是指通过利用人工智能的深度学习算法和模型生成视频的技术，高度仿真性是其核心特征[3]。具体而言，使用者先通过网络轻易获取一个人身份的内容如声音、面部、身体，然后再通过深度伪造技术低成本而轻易地以对方身份的内容形成视频，结果展现在人们眼前的是，内容为自己或他人做和说从未做过的事情的视频[4]。

根据双重编码理论，深度伪造的视频提供了同步且一致的视觉与听觉信息，使伪造的场景通过两个独立编码通道进入大脑，并且具有极高生动性与真实感，这种新的输入在大脑中留下了印记。根据巴特莱特的图式理论，一种新的传入冲动不仅是一种线索，可以建立起新的反应，还是一种刺激，刺激过去的反应。深度伪造视频作为一种刺激，并且其内容高度逼真且完整，大脑容易把它加入到本就不断变化发展的“图式”之中。巴特莱特认为往事不完全作为一种团块形式起作用，建立“图式”的最后传入成分具有支配性影响[1]。当伪造信息与原始图式发生冲突时，由于伪造的视频具有极高的信源权威性，甚至从时间上也相对更新，在记忆中更加明确清楚，大脑会通过主动修改并更新“图式”以容纳新的、更生动的信息，从而将伪造内容作为填补记忆空白的合理细节。“我们会受到过去某一时刻发生的反应和经验的影响，和某种类型的所有传入冲动结合在一起，进而建立一种积极、有组织的场景[1]264。”伪造的内容就这样被直接整合入图式中，最终在记忆重构的过程中作为信息源参与形成新的记忆，同时记忆和图式以被修改后的形式再次巩固，错误信息在后续提取和重构时以我们更加相信其为真实地呈现。

4 负面影响的具体表现

4.1 对个体记忆的篡改与情感破坏

因为我们没有无限的时间和精力进行亲自接触和了解他人与世界，但是我们又和外界社会存在着无法断截的联系，所以拥有那些我们能够信任的信息来源是十分必要的。其中，直接的视觉感知就是一个重要来源。但我们不可能总是在正确的地点和应该的时间“看”事情，进而产生真实的了解。在这种情况下，视频往往是最直接的选择[4]。正如我们所坚信的，直接的视觉感知会提供一个事件发生的可靠证据，其中视频便可以提供虽然不在场但是几乎和亲身经历或与旁观一样令人信服的证据。

在深度伪造出现前，通过不停的设想与引导甚至反问和逼问，已经会对被问者（案件相关者或目击者）造成干扰，他们会开始怀疑自己原本所持有的记忆，以及有些情况由于在案件过程中，相关者确实存在错误（但并非导致结果的主要原因），而因受干扰便开始衡量记忆中的错误程度究竟是否如同对方所说。总之在怀疑中可能造成记忆的篡改，最终导致误认。而今，应用深度伪造的虚假视频，通过对设定合理情节回忆的伪造，由于记忆的可塑性，相关者或目击者因外界视频信息的干扰将重建“现场”或“情形”，导致了不准确的记忆，进而造成无辜的人自己承认或被指认而发生冤假错案[5]。在这里，相关者和目击者并不是在有意说谎，因为他们甚至在讲述他们被篡改的记忆内容时都是说服了自己后带有信誓旦旦的，由此可见这些虚假视频对于篡改人记忆的影响。如果说对于这些视频的存在的形式或来源有疑问，或许可以设想一下更改视频记录的某些细节，毕竟随处可见的凑热闹人群拍摄一段视频上传到网络中的“网民习惯”早已不是一件新鲜事。或者直接伪造记录的镜头画面，甚至侵入监控系统替换新的内容，在利用深度伪造这一技术下，实现这些目的也已不再是一件难事。

篡改记忆还会对记忆有关的情绪或情感产生不利影响。在我们的印象中，记忆不仅包括了我们一直持有的过去的存在和意义，也包括对过去的情感。无论是我们的关于他人的记忆，还是以自身为中心的记忆，都或多或少承载着不同的情感。当我们记忆被伪造视频篡改后，往事的感觉和情感的将消失或与现在的感受出现冲突。此外，记忆是源自过去事件的知识，但不一定是内容为过去的知识。当一件未来的事存在于我的记忆中后，即我知道某事会发生，带有一种“相信”，我便会有期待或者紧张等情绪波动。而如果因为深度伪造的视频，与我们已知的内容出现了矛盾冲突，甚至记忆被篡改了，那之前所谓的波澜便都成了空虚的存在，同时也浪费了精力。例如，我记得 2028 年将举行奥运会，我是在过去知道这个事情的，它在未来才发生，但是人们会因为它的存在而饱含激动和期待，而如果这一记忆被伪造的类似官方通知性信息损害，将令人感到难受，无论是期望落空，还是产生疑惑哪段记忆为真实的矛盾心理。

除此之外，心理创伤会一直存在，一旦源自过去的伤害再次被激起，心理受伤害的人就会产生不相称的反应。它把不相干的物体与产生创伤的过去的物体联系起来。这是两种再现过去的病理学表现形式[6]。倘若人们通过深度伪造制

作出当时产生创伤的场景，这样生动直观的视频将会比那些物体更有冲击力，有心灵创伤的人会被迫回忆起过去的那个场面，以至于重新恢复那种让人紧张或不安的情绪，造成更新更持久的二次伤害。过去的创伤，无论是淫秽还是施暴，比起身体上的痛苦，更为痛苦的是在记忆中的精神上的折磨，从神经学和心理学上讲，短暂的身体上的疼痛是一时的，但是造成对人格的侮辱和耻辱感的阴影将伴随一生。

4.2 对集体记忆的扭曲与政治操纵

哈布瓦赫认为存在着一个所谓的集体记忆和记忆的社会框架。[7] 的确，每个人都仅仅亲历了历史中有限的一小段，并且是以一种极为局限的方式经验着，那么有关于大部分无法直接经历的历史与世界，人们的记忆则依赖于信任的人所传播的外界知识。当一群聚集性的人接触和获取到相同的外界知识，将形成记忆共同体，即产生“集体记忆”。此外，人是符号性动物，人与人之间可以在不必面对面接触的情况下形成象征性的团结关系 [6]。那么就意味着通过操作共享记忆的符号便可以决定“集体记忆”，由此深度伪造便可以通过创造“视频”这一符号，“准确地”再现（其实是编造）历史事件，介入集体记忆，直接阻碍形成真实集体记忆或者篡改集体记忆。例如在视频中和我们说话的人看起来像一个我们知道的值得信赖的人（比如某些公众人物），但是它可能是一个别人用来欺骗大众的深度伪造形象。

在传统意义上，个人的问题导向社会的问题，个人和社会的共同演进，即通过形成共同的集体记忆去解决问题和实现发展。但是在深度伪造技术对人们获取共同记忆进行了冲击之后，这种集体的历史记忆本身不再是真正意义上的“可相信”的集体记忆。就这样，深度伪造对个人层面的记忆和社会层面的信息造成了多方面的破坏，集体记忆和信息将不再具有真实性，更深远的则会导致社会历史无法发挥原有的正面意义，由此将引发更深层的伦理危机。

深度伪造或将扭曲国家和历史的情怀与认同。人们通过深度伪造可以制作出容易令人信服的各种虚假视频，将虚构知识植入特定历史，有意地偏离公众对这段历史的集体记忆，从而使得原有的历史记忆被割裂并重新建构。其目的是以他们所需要的利益方向重新建构与真实的历史相异的集体记忆。集体记忆的价值在于形成民族的共同意识，但是如果以这样一种错误却令人信服的手段伪造更改社会集体记忆，真实的历史将被隐藏，这样形成的共同意识将背弃集体记忆道德，扭曲国家和历史的情怀与认同。

切入到政治博弈上的，即被国际政局的现实所操控。例如当小国面临政治被深度伪造技术直接逆转的困境，但就连证明“视频是伪造的”这一点，都是要去借助大国的技术和话语权的，这个意义上，它居然在破坏没有技术优势的民族、社会和共同体的生存的基本条件。另外，很多在媒体流传的视频或许本来是弱小的民众在呼唤正当的解决，但现在面对深度伪造导致民众认知怀疑增加，将对视频的真实性进行否定和怀疑，便可以故意或许含有无意地将他们的需求以正当的理由屏蔽，甚至想将他们从历史上抹去。也在利用有这种技术后，感到霸权在我们谈话的桌面上布下了一层铁幕。

5 结语

本文探讨了深度伪造技术如何利用心理学机制，对个体记忆产生负面影响，进而影响集体记忆。深度伪造技术并非仅仅创造和生成了虚假信息，并且通过利用记忆的重构性本质以及视频媒介的高权威性，给个体记忆植入虚假记忆，从而破坏了个人记忆的连贯性与真实性，产生更多方面的负面影响。同时，这种对个体记忆的破坏延伸至社会，引发了集体叙事危机和政治操纵。然而本文主要基于心理学理论和伦理学框架进行理论推演，缺乏大规模的认知实验数据来实证深度伪造视频对记忆篡改的可行性和持久性。鉴于当前的技术发展趋势，未来可以深入研究深度伪造实时检测技术，在它不断向真实完美的结果靠近过程中加入伦理审查机制。

参考文献

[1] 英)弗雷德里克·C.巴特莱特著;黎炜译：记忆[M].浙江教育出版社，1998，263-265

[2] 斯蒂芬·霍尔，Jacob Escobedo，修改可怕的记忆[J].科技创业，2013，(08):42-48

[3] 赵国宁：智能时代“深度合成”的技术逻辑与传播生态变革[J].新闻界，2021(06):65-76

[4] Philosophy & Technology (2021) 34:623–643

[5] 白洁著：记忆哲学[M]：中央编译出版社，2014:108.

[6] （以）阿维夏伊·玛格利特著；贺海仁译. 记忆的伦理[M].北京：清华大学出版社，2015.01：117

[7] （法）莫里斯·哈布瓦赫著；毕然，郭金华译. 论集体记忆[M].上海：15