# Beyond disinformation: deep fakes and false memory implantation

**Erin Morrow**

Emory University
Neuroscience and Behavioral Biology
Undergraduate

eemorro@emory.edu

Disclosures: The author discloses their employment with the International Neuroethics Society and thus access to other contest submissions.

General Audience / Neuroethics Essay Contest, 2021

# Beyond disinformation: deep fakes and false memory implantation

Most social media users are intimately familiar with labels on their timelines and feeds for content deemed ‘manipulated media.’ Twitter rolled out new policies last year to address photos or videos found to be “significantly and deceptively altered or fabricated” (Roth & Achuthan, 2020). The first such content to be flagged on this platform showed then-presidential candidate Joe Biden appearing to state “We can only re-elect Donald Trump.” In reality, this Kansas City speech was deceptively spliced, edited, and thus taken out of context—but by the time this post received its warning label, it had already garnered over five million views (Chambers, 2020).

Yet, misleadingly cropped video is not the only form of deceptive content being popularized. Also within the jurisdiction of this new regulation is an emerging form of media: *deep fakes*. These artificial intelligence-generated creations, often in the form of videos that deceptively portray public figures, have emerged within the past decade as a force of disinformation to be reckoned with (Liv & Greenbaum, 2020; Resnick, 2018). Such videos can now manipulate the facial expressions and speech of these figures with minimal distortion (Resnick, 2018).

Deep fakes of public figures have the potential to influence discourse and decision-making in the realms of politics, public health, and more through a powerful cognitive system: memory. Deep fakes can ‘implant’ *false memories*; that is, they can cause individuals to recall entirely new events that did not actually occur (Liv & Greenbaum, 2020; Pezdek & Lam, 2007; Resnick, 2018; Wade et al., 2006). Famed memory psychologist Elizabeth Loftus (1997) has shown, with others, that certain false memories can be consistently recalled weeks after initial encoding (Garry et al., 1996). This persistence suggests long-term effects of false memories on brain structure and function, and we know intuitively that long-term memory impacts the choices we make, our socioemotional states, and our very identity.

Thus, the dissemination of deep fakes engages pressing questions: what might deep fake disinformation do to the brain, and how do these changes affect autonomy? When veracity erodes in the public sphere, what stopgaps can we employ to help ensure we are all functioning from a place of truth?

To tackle these questions, we can first step back in history to examine what we know about how susceptible our brains are to false memories. Neuroscientists and psychologists concur that memory is quite fluid—certainly not an infallible ‘tape recorder’ of passing events. Consider seminal experiments conducted by Loftus and others: in an infamous study, 29% of participants

claimed that they recalled details from a totally fabricated childhood event (being 'lost in the mall'), and a *fourth* maintained this claim in two subsequent interviews (Loftus, 1997; Loftus & Pickrell, 1995). The surprising proportion of susceptible individuals point to the potential of our memories to be intruded upon.

Further research suggests that *certain factors* may make the acquisition of false memories more likely. These include failing to provide warnings (only in certain cases; Eakin et al., 2003), status as a young child (see Ceci & Bruck, 1993) or an elderly adult (Davis & Loftus, 2005; Karpel et al., 2001), and several different personality traits (see Davis & Loftus, 2005) (Loftus, 2005). From this work, it appears that 1) memory is susceptible to suggestion, 2) a significant portion of people can acquire false memories, and 3) certain individuals and groups may be particularly at risk to this acquisition. Many of these factors play a role in the potential ethical implications of deep fakes.

Perhaps the most obvious motive for creating and spreading deep fakes in the public sphere is to influence politics. Foreign and domestic actors can use deep fakes for disinformation schemes with the intention of destabilizing a country's electoral process or benefitting their own campaign. However, the use of deep fakes is not limited to politics—in the pandemic era, one could imagine, for example, a manipulated video of a public health official giving sham advice. This threat may be magnified when the disinformation aligns with the worldview of a segment of that public, which may make implantation of false memories more likely (Bartlett, 1995; Frenda et al., 2012; Liv & Greenbaum, 2020). Regardless of the actors and of the topic, these false memories would challenge a central tenet of human experience: *autonomy*.

Indeed, the potential longevity and durability of these false memories undermines an individual's ability to make *free decisions* for themselves. Could this staying power allow deep fakes to influence election outcomes (Liv & Greenbaum, 2020), endanger public health, or even incite violence? The decisions that individuals affected by false memories make can impact themselves as well as their communities and the broader global collective.

Moreover, as previously mentioned, some may be especially likely to acquire false memories from deep fakes (Loftus, 2005). These individuals would, then, disproportionately experience reductions of autonomy. For instance, these could include the elderly (Davis & Loftus, 2005; Karpel et al., 2001), who make up nearly a fourth of the American electorate (ages 65 and older; Cilluffo & Fry, 2019; U.S. Census Bureau, 2017) and have poorer overall memory. Influencing the votes of even a fraction of this vulnerable population could have a significant impact on the results of a general election. The rise of deep fakes, it seems, will not affect us equally.

How can we protect these at-risk populations, as well as the wider public, from such persuasive disinformation? With the knowledge that long-term false memories may be possible, may persist after correction (Frenda et al., 2012; for competing evidence, see e.g., Murphy et al., 2020), and may impact groups and individuals differently, we can begin to construct a framework in which

certain strategies to tackle deep fakes are prioritized (e.g., protecting the elderly). In addition, recognizing specific challenges to autonomy—whether political, medical, or otherwise—directs us towards particular areas of concern. Memory is integral to our identities, lived experiences, and arguably each decision we make. Deep fakes have the potential to challenge the integrity of these things, but with the incorporation of neuroethics considerations, we can edge closer to the pursuit of truth.

## References


1. Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*(3), 403.
2. Chambers, A. (2020, March 10). Twitter's 1st 'manipulated media' post? It came from the White House. *ABC News*. Retrieved from https://abcnews.go.com/Politics/twitters-1st-manipulated-media-warning-white-house/sto ry?id=69504259
3. Cilluffo, A., & Fry, R. (2019, January 30). An early look at the 2020 electorate. *Pew Research Center: Social & Demographic Trends.* Retrieved from https://www.pewsocialtrends.org/essay/an-early-look-at-the-2020-electorate/
4. Davis, D. and Loftus, E. F. (2005). Age and functioning in the legal system: Perception memory and judgment in victims, witnesses and jurors. In *Handbook of Forensic Human Factors and Ergonomics* (eds. I. Noy and W. Karwowski). Taylor and Francis, London.
5. Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 813.
6. Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, *49*(2), 280-286.
7. Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*(2), 208-214.
8. Karpel, M. E., Hoyer, W. J., & Toglia, M. P. (2001). Accuracy and qualities of real and suggested memories: Nonspecific age differences. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(2), P103-P110.
9. Liv, N., & Greenbaum, D. (2020). Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. *AJOB Neuroscience*, *11*(2), 96-104.
10. Loftus, E. F. (1997). Creating false memories. *Scientific American*, *277*(3), 70-75.
11. Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory*, *12*(4), 361-366.
12. Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720-725.
13. Murphy, G., Loftus, E., Grady, R. H., Levine, L. J., & Greene, C. M. (2020). Fool me twice: How effective is debriefing in false memory studies?. *Memory*, *28*(7), 938-949.
14. Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study “false memory,” and what are the implications of these choices?. *Consciousness and Cognition*, *16*(1), 2-17.
15. Resnick, B. (2018, July 24). We’re underestimating the mind-warping potential of fake video. *Vox*. Retrieved from https://www.vox.com/science-and-health/2018/4/20/17109764/deepfake-ai-false-memory -psychology-mandela-effect

16. Roth, Y., & Achuthan, A. (2020, February 4). Building rules in public: Our approach to synthetic & manipulated media. Retrieved from https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-mani pulated-media.html
17. Wade, K. A., Sharman, S. J., Garry, M., Memon, A., Mazzoni, G., Merckelbach, H., & Loftus, E. F. (2007). False claims about false memory research. *Consciousness and Cognition*, *16*(1), 18-28.
18. U.S. Census Bureau. (2017) *2017 National population projections datasets*. [Pew Research Center 2020 projections from data]. https://www.census.gov/data/datasets/2017/demo/popproj/2017-popproj.html