Managing the Cultural Record in the Information Warfare Era

Several rapidly emerging lines of technology development and exploitation are converging, and they are going to change the world in the next decade. They will have massive social and political impact; indeed, we are already far down that path, as I'll discuss shortly. These trajectories will create new complexities for a wide range of scholarly investigations. They will challenge us to rethink the way we define and teach information literacy. They will demand that memory institutions such as libraries and archives reconsider the documentation and contextualization of the cultural record, and they may even drive the creation of new public infrastructure supported by memory institutions and responsible content creators and distributors.

Fully exploring these developments requires a book (at least), but I will try to give a very high-level sketch here, with some limited pointers to additional information (much more can be found with a little Googling). My hope is that the reader will be able to see the broad trends.

The first development is the ability to fabricate audio and video evidence. Software that can do this is becoming readily available and doesn't require extraordinary computational resources. If you want to produce a persuasive video of someone speaking any script you'd like and if that person has a reasonable amount of available recorded video, you can synthesize that video into the fabrication software.[1] The obvious place for this is politics: pick your target politician, put words in his or her mouth, then package this into propaganda or attack ads as desired.

Fabrication is much more than talking heads, of course. In keeping with the long tradition of early technology exploitation in pornography markets, another popular application is "deepfakes," where someone (a public figure or otherwise) is substituted into a starring role in a porn video (the term "deepfakes" is used both for the overall substitution technology and for the specific porn application). This is already happening, though the technology is as yet far from perfect. Beyond the obvious uses (e.g., advertising and propaganda), there are plentiful disturbing applications that remain unexplored, particularly when these can be introduced into authoratitive contexts. Imagine, for example, being able to source fabrications such as police body-camera footage, CATV surveillance, or drone/satellite reconnaissance feeds. The nature of evidence is changing quickly.

From a purely technological perspective, machine learning is being harnessed in fascinating ways that feed the ability to fabricate. Space does not allow me to explore generative adversarial networks (GANs) in any detail, but basically the idea is that one system creates fakes, another system identifies the fakes, and then the two are connected so that both can improve their game iteratively as they interact. GANs create a continuing "arms race" between falsifiers and falsification-detection systems; each of these systems can be unleashed on the world independently. It remains to be seen whether the advantage rests with the offense or the defense.

Part of the social challenge here is that people seem to be wired to believe their eyes and ears (i.e., "seeing is believing"). Having encountered advertising, propaganda, and fiction, they are experienced with, and hence have some level of defensive skepticism about, the written word. Even though there is a century of experience with photo manipulation. video in particular still seems to be deeply persuasive, and we don't understand how the potential for personalized fabrication in the current environment, as opposed to a "publication" or "broadcast" dissemination, may change the balances. Fixing this is going to have deep implications for how we think about information literacy going forward. While there's a great deal to be learned from our experiences over the past century, what's different today is the scale, the ready availability of these tools to interested *individuals* (rather than nation-states), and the move into audio/video contexts.

In addition, a separate and important set of issues concerns how fabricated material broadly (whether old-fashioned text materials or new digital fabrications) is introduced to the public sphere and subsequently promoted and given visibility and credibility (e.g., through manipulating social media system mechanisms or by subverting what are viewed as "official" channels). Although I will not consider these issues here,[2] tools, opportunities, and strategies in this sphere could be considered a second driving development.

The third development thread is a bit more speculative. Anyone who has followed security breaches and penetrations over the past few years knows that the track record of protecting data aggregations from exfiltration and subsequent disclosure or exploitation is very poor. And there are many examples of attackers that have maintained a presence in organizational networks and systems over long periods of time once they have succeeded in an initial penetration. While a tremendous amount of data has been *stolen*, we hear very little about data that has been *compromised* or *altered*, particularly in a low-key way. I believe that in the long term, compromise is going to be much more damaging and destabilizing than disclosure or exfiltration.

As we have moved from highly distributed preservation and storage of physical materials (i.e., libraries) to centralized digital

E-Content Editor: **Joan K. Lippincott**

Photo courtesy University of Texas Libraries

By CLIFFORD LYNCH

resources (i.e., major news media sources, scholarly journals, or almost anything else in the digital world), we are facing a scenario in which a very small number of points of potential compromise exist for a great deal of our scholarly and cultural record. This new centralized world offers few checks, tripwires, or other mechanisms to prevent an attacker from rewriting pieces of the scholarly or cultural record (including legal or government records) once the central server is compromised. I want to explicitly note here the difference between the act of quietly rewriting the record and enjoying the results of the rewrites that are accepted as truth and that of deliberately destroying the confidence of the public (including the scholarly community) by creating compromise, confusion, and ambiguity to suggest that the record *cannot be trusted*. Both acts are very dangerous and damaging, but they serve different objectives. While rewriting or populating the news with fake audio-visual material is likely to have the biggest impact on the public at large, we need to think through the potential impact of a subtly corrupted scientific record and the issue of how we will develop a generation of scholars who can question, recognize, and deal with this sort of intentional corruption.

But putting the primary burden for this task on higher education is insane. Information literacy needs to start in elementary school, and students must be consistently and continuously engaged as they mature from there. New challenges in how we identify and contextualize various kinds of fabrications in libraries, archives, and museums—and in the classroom and in the learning experiences of students more broadly—will continue to arise. We *must* collect these materials: they will be essential in the future for understanding the present, if for no other reason than that they play an important role in shaping reality for today's broad public; their personalization will be an immense challenge.[3]

Obviously, we need greater capability in digital forensic technologies to detect computer-generated fabrications. The Defense Advanced Research Projects Agency (DARPA) is making a substantial investment in this area.[4] Simply claiming that we will develop machine learning systems to identify fakes is much too simple, however, and much too glib: humans need to be able to understand clues, and evidence, and to work in tandem with these machine-learning systems (much as, I think, the best medical image analysis will be done by machine-learning-based systems and human experts working together).

Actions can be taken to "harden" the system. Being able to prove that a digital object existed at a given time, and/or was captured at a specific time and place, and/or wasn't altered since it was registered, is very valuable. There's an active research area in terms of imaging devices that aim to "sign" captured imagery with GPS coordinates and timestamps,[5] but this is a difficult problem to solve (e.g., GPS receivers can be easily spoofed).

Independent of the efforts to document the capture of materials in a trustworthy way, the cultural memory sector must step up to the challenges of contextualizing media once it is created and disseminated. Registries are fairly straightforward, at least technologically. It's important to track chains of custody and provenance in ways that are transparent and secure. I want to distinguish such registry and subsequent provenance tracking carefully from third-party escrow *preservation* systems, like *Portico*, which deal with redundant custody of full-content copies; these are also critically important to the survival and resilience of content rather than just integrity. They can actually *substitute* for the original content holders under appropriate circumstances. Setting up and operating such preservation systems is a complex financial, legal, and contractual as well as technical undertaking; it's also essential, and in many sectors beyond scholarly publishing, motivating content holders to participate has thus far been intractable.

All a registry system can do is provide testimony that shows, to a very high degree of confidence, that a digital object held by someone else has existed since a given time and has not been modified. Registry is a much simpler thing than preservation and can be used by third parties as well as the content holders. We don't have such systems today as part of broadly recognized public infrastructure for digital content.[6]

A four-pronged approach to the new information warfare environment seems to be emerging. One prong is greatly improved forensics; this is a mostly technical challenge, and memory organizations will be mainly users, not developers, of these technologies. Documentation of provenance and chain of custody are already natural actions for memory organizations; the challenge here is to make this work more transparent and rigorous and to allow broad participation. Capture of materials, particularly in a world of highly targeted and not easily visible channels, will be a third challenge at both technical and intellectual levels (though we are seeing some help now from platform providers). Finally, contextualization of fakes or suspected fakes is perhaps the greatest challenge, and the one that is least amenable to technological solutions. ■

Notes

1. See, for example, Michael Zollhöfer, "Deep Video Portraits," 2018.
2. These issues have received a great deal of recent attention through explorations of developments such as the Russian manipulation of social media platforms in US elections and the UK Brexit referendum.
3. Clifford Lynch, "Stewardship in the 'Age of Algorithms,'" *First Monday* 22, no. 12 (December 4, 2017).
4. For example, see Matt Turek, "Media Forensics (MediFor)," U.S. Department of Defense website, sponsored by the Defense Advanced Research Projects Agency (DARPA), n.d.
5. See "CameraV: Secure Verifiable Photo and Video Camera," *Guardian Project*, n.d., and "Combating 'Fake News' with a Smartphone 'Proof Mode,'" *Guardian Project*, February 24, 2017.
6. Blockchain is very fashionable and getting a great deal of press; really what's relevant here is the broader idea of distributed ledgers. See Hanna Halaburda, "Blockchain Revolution without the Blockchain?" *Communications of the ACM* 61, no. 7 (July 2018).

Clifford Lynch (cliff@cni.org) is the Director of the Coalition for Networked Information (CNI).