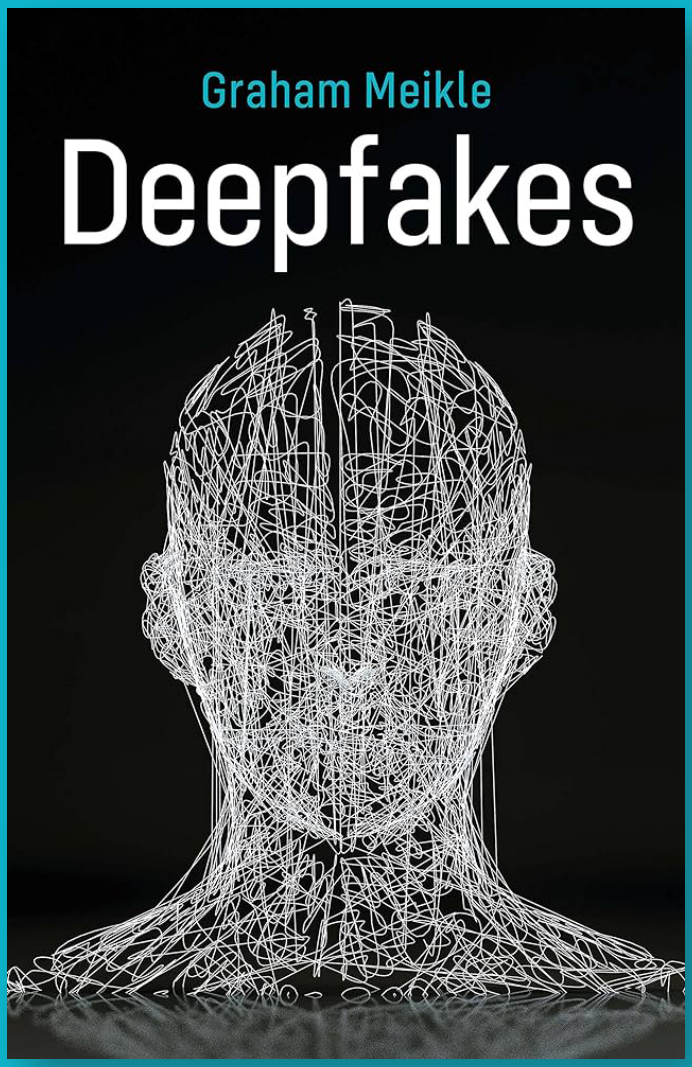

Daniel E. Levinson Review of *Deepfakes* by Graham Meikle, Polity Press, 2023

By Daniel E. Levinson

Suggested Citation

Levinson, Daniel E. Review of *Deepfakes* by Graham Meikle, *Homeland Security Affairs* 19, Article 4 (Sept 2023) www.hsaj.org/articles22483

The term "deepfake," which informs both the title and the focus of author Graham Meikle's new book, is a portmanteau, taking "deep" from the concept of "deep machine learning" with "fake," to indicate that the image created is not genuine. And as Meikle makes clear, it is the combination of these two elements which render the final product not merely insidious, but perhaps catastrophically corrosive when it comes to fundamental pillars of modern liberal democracies, including the ability to access reliable news online, hold elections without interference, and maintain trust among key stakeholders. While the author is not the first to sound the alarm regarding the potentially devastating damage that this technology may unleash, he does do so in clear and compelling language, examining the frankly existential threats that close observers have discerned in its unfettered growth. For this reason, *Deepfakes* will not only undoubtedly prove valuable to homeland security practitioners, but should also serve as a wakeup call to those who only see the upside of advances in the field of generative artificial intelligence. At the heart of this book is a convincing argument that we should be deeply worried about the rapidity with which generative artificial intelligence is becoming increasingly powerful and easy to use, and as a result, ubiquitous, without any restriction regarding how or by whom such tools may be used.

Early adopters of technology have never been comprised of those possessed solely of benign or beneficent intentions, nor have terrorists and other bad actors been slow to exploit new opportunities when it comes to propaganda or technology. In the late 19th and early 20th century anarchists embraced the concept of "Propaganda of the Deed" through which they sought to "shock" an apathetic citizenry out of their supposed political apathy by making spectacular acts of violence visible to the masses. In the early 1990's American white supremacists established a presence on online bulletin board services. The 9/11 attacks and subsequent digital propaganda distributed by groups like ISIS and Al-Qaeda in the Arabian Peninsula were clearly conceived of with spectacle in mind. Photography, the telegraph, the telephone, radio, television, personal computers and much more have each found their place in the terrorists' toolkit, but perhaps nothing which we have seen before can compare to the propagandistic appeal of generative artificial intelligence. As Meikle makes clear in his book, for good or ill, the possibilities are practically endless, and the barriers to entry for use of this technology, whether financial or skill-related, will soon be non-existent.

One of the major strengths of this book is that the author draws on an existing body of evidence relating to the production and dissemination of deepfakes in recent years in two significant areas. The first is in the production of unauthorized semi-artificial pornography, in which someone takes easily available digital images of a celebrity from social media and uses generative AI to alter an existing pornographic image or video, replacing the head or face of the original person in the graphic image with that of the celebrity. As Meikle makes clear, this is a very popular activity among a certain subset of internet trolls, and the net result has been

an untold number of pornographic films floating around online which now appear to feature famous mainstream actors. The other, less salacious use of the technology which has drawn attention is its intentional use by artists to create works that alter digital video of political leaders and other influential individuals in order to provoke discussion or alter perceptions (as much art does).

The author's examination of the use of generative AI among artists is also important because it offers a more relatable example for the majority of readers, than perhaps, that of illicit celebrity pornography. After all, the manipulation of images for artistic purposes is nothing new, and anyone who has visited a museum or gallery can probably relate to the idea that offering alternative visions of past events and imaginative glimpses into a possible future, are central to many artistic traditions and schools. Artists have been offering interpretive visions of reality for tens of thousands of years, from the caves of Indonesia and Spain, to the paintings of Salvador Dali and the work produced by the early twentieth century Dadaism. One of the contemporary, AI-driven examples that Meikle provides in his book, "In Event of Moon Disaster," raises important questions about how such technology may be used by conspiracy theorists. In the digital video installation, which combines the text of an actual presidential speech (written, but never delivered) drafted by one of Richard Nixon's speechwriters to be used in the event that the Apollo 11 mission had not landed safely on the moon in 1969, combined with generative AI to show the late president actually delivering that speech, an event which never occurred in reality.

This is a brilliant example on the part of Meikle, because it highlights the potential for creative expression using this technology, but by using an event which has also drawn decades of attention from conspiracy theorists, he highlights the potential for digital revisionists to run wild with respect to practically any critical moment in history, noting, "Social media have provided fresh opportunities for moon-landing deniers to connect, communicate, and collaborate. So in this context it's not hard to imagine ways in which 'In Event of Moon Disaster' could be used to further undermine both the historical and scientific record and public trust in science and its communication" (117). The reader is then left to wonder (in horror, one might suppose) how else this technology could be further exploited by conspiracy theorists and extremists, who could produce "proof" to back up any outlandish or hateful narrative.

Another significant contribution that this book makes to our collective discussion on AI and national security is the highlighting of the indirect manner in which the broad scale use of this technology to generate misinformation and disinformation, and then distribute it practically without restraint, is likely to have cascading effects with profound consequences for governments and societies around the globe. Practitioners in various fields where AI and security intersect are already looking at digital media with a critical eye when it comes to determining the credibility and relevance of evidence comprised of zeros and ones.

This is challenging enough, but what happens when *everyone* does the same with *everything* they see online? As Meikle notes, the critical issue here is not necessarily that the net result will be broad segments of society embracing conspiracy theories or making bad financial or political decisions based on partially or wholly false information. This scenario

is troubling enough to contemplate, but by undermining trust in the entire digital media ecosystem, opportunities will arise organically for people to get away with all kinds of actual bad behavior. All they have to do is respond to photos, videos, voice recordings, and other material showing them to have engaged in unethical or even criminal behavior, by leveraging society's crumbling faith in online media and claim that the "evidence" being shared online is itself a kind of deepfake. This concept, referred to as "The Liar's Dividend," originated with two professors of law, Danielle Citron and Robert Chesney, and seems tailor-made (in an exceedingly troubling fashion) for the twenty-first century global media marketplace. As Meikle notes, the mere *existence* of this technology may pose a more proximate threat than what bad actors can create through its use, writing, "The Liar's Dividend is a very real threat, and can certainly be exploited more easily, frequently, and quickly, than persuasive deepfakes can be made" (P.161).

Ultimately, Meikle's work is as much a meditation on the nature of trust in the digital age, as it is an exploration of the current and potential impact of synthetic media and artificial intelligence. In the examples he provides of individual artists presenting altered voice and video to provoke a reaction or explore a particular idea, the audience is essentially "in" on the joke - they are viewing the material in a context which makes no pretense that what they are about to experience is factual. This stands in sharp contrast to what bad actors, using the same tools, are able to unleash on broad segments of society without any such framing or qualifiers. The difference is that when a viewer watches "In Event of Moon Disaster" in a museum they are prepared to see it in its proper context, essentially as counterfactual digital art, but when someone takes the same material, in whole or in part, and begins to share it online as "proof" to undergird conspiracy theories relating to the space exploration, it becomes something much more dangerous.

Meikle notes toward the end of his book that he has tried not to look too far into the future, and while he has perhaps avoided laying out a concrete roadmap for the ways in which he anticipates that this technology will evolve and impact society, he has crafted a well-written warning for what is likely to lay in store. As the book makes clear, actions taken online, including the creation and dissemination of deepfake material, can and do have consequences in the real world, on an individual scale, but also critically, on a societal one. The true threat at the heart of generative AI may lie not in the erosion of our ability to determine whether any one discrete image or video is real or not, but in finding ourselves in a place in which all norms around communication and our capability to navigate digital spaces have been completely undermined. It is not a pretty future, but it is one for which we must prepare, and Meikle's book is a good place to start.

About the Reviewer

Daniel E. Levenson is Director of the Communal Security Initiative at Combined Jewish Philanthropies in Boston, Massachusetts. He holds an MLA in English and American Literature from Harvard University, an MA in Security Studies with a concentration in Homeland Defense from the University of Massachusetts at Lowell, and will begin a doctoral program in criminology in the coming year. Daniel is a member of the FBI Boston Mass Bay Threat Assessment Team, and regularly presents on violent extremism, threat assessment, and related topics for a wide variety of professional audiences. He is presently working on a series of essays examining issues at the intersection of national security and artificial intelligence. He may be reached atlevenson.daniel@gmail.com

---

Copyright