2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Seeing Through the Fog: The Impact of Information Operations on War Crimes Investigations in Ukraine

Lindsay Freeman
Director of Technology, Law & Policy
Human Rights Center
UC Berkeley School of Law
Berkeley, California, USA
lfreeman@berkeley.edu

Abstract: As Russian forces closed in on Kyiv, a MiG-29 Fulcrum swooped in and took down six Russian jets. The next day, the same MiG shot down ten more. Stories of the hero fighter pilot spread like wildfire throughout Ukraine and across the internet, turning the "Ghost of Kyiv" into a living legend.

But he was not living, or even real. The pilot and his exploits were a total fiction created as part of an influence campaign spread via social media to strike terror into Russian forces, fortify the resolve of Ukrainian citizens, and amaze the world with Ukraine's unexpected strength and courage.

The strategic use of the online information environment is only one facet of intangible warfare between Russia and Ukraine that makes this contemporary conflict particularly unique and complex. Propaganda, disinformation, and psychological operations are as old as warfare itself, but advanced digital technologies now reshape conflicts in often unanticipated, unforeseen, and surprising ways. These changing dynamics inevitably have an impact on those tasked with investigating war crimes and establishing the truth of what occurred on the battlefield.

This paper examines the strategic use of digital information and communications technologies in the Russia–Ukraine conflict to better understand how they are changing the dynamics of war, war narratives, and war crimes investigations. The first section of the paper briefly explains how war crimes investigators and prosecutors are

increasingly relying on digital material as evidence in their cases. The second section considers how digital information operations are being deployed and how these operations impact the investigation of war crimes. Finally, the third section highlights some of the tools that can help war crimes investigators fight back against a complex and chaotic information environment.

Keywords: *information warfare, disinformation, influence operations, war crimes investigations, digital evidence, Berkeley Protocol*

1. INTRODUCTION

"The great uncertainty of all data in war is because all action, to a certain extent, planned in a mere twilight—like the effect of a fog—gives things exaggerated dimensions and unnatural appearance."

Carl von Clausewitz[1]

Throughout history, military generals and strategists have espoused the importance of information in warfare, characterizing the ways in which information operations can be used to further military objectives and gain a competitive edge over the opposition. As Napoleon Bonaparte once stated, "War is ninety percent information."[2] Information operations—a term that encompasses a range of activities from disseminating propaganda to spreading disinformation to blocking access to communication channels—can have a profound impact on the course of conduct on the battlefield, as well as the narratives that surround it. While information warfare is not a new concept, the adoption of digital information and communications technologies (ICTs) has changed the nature of wartime information operations in interesting and unforeseen ways. These new dynamics are currently playing out in Ukraine, where the largest international armed conflict in Europe since World War II is taking place.

While historical analysis of information operations can provide some insight into what the world is witnessing in Ukraine today, there are many novel elements of modern information warfare that cannot be fully understood within traditional frameworks. The teachings of Sun Tzu and Carl von Clausewitz never explained how to crowdfund weapons or troll the enemy on social media. These traditional military theorists could not have imagined the speed and scale of modern digital communications, nor could they have envisioned a world in which ordinary citizens across the globe could monitor the battlefield in near real time with high-resolution satellites and live drone

1 Carl von Clausewitz, *On War*, vol. 1 (Altenmünster: Jazzybee Verlag, 1950)
2 Napoleon I, *Military Maxims of Napoleon* (New York: Wiley and Putnam, 1845).

feeds. But this novel information universe is precisely the reality being experienced in Ukraine's fight against Russian aggression and occupation.

New technologies can shrink time and space, blur borders, and alter the ways in which information travels from the battlefield to the outside world, and from the outside world to the battlefield. Although social media, smartphones, and other digital ICTs have played critical roles in past armed conflicts over the years—from their use in documenting war crimes in Syria to their role in furthering crimes against humanity in Myanmar—their application in the Russia–Ukraine conflict is unprecedented. As policy analysts Christian Perez and Anjana Nair explain, "Throughout the ongoing conflict, social media has served as a battleground for states and non-state actors to spread competing narratives about the war and portray the ongoing conflict in their own terms."[3] In a hybrid war between state militaries with far-reaching implications for the rest of the world, the information environment has become contested, corrupted, and dizzyingly complex. The lessons of the past can only take us so far in understanding this new digital world.

Advanced digital technologies are not only changing the nature of warfare and facilitating the weaponization of information but also transforming how war crimes investigations are conducted. Over the past decade, in response to hostile governments blocking access to crime scenes, the international criminal justice community has built up the capacity to conduct remote investigations using ICTs. In recent years, satellite imagery, call data records, wire transfers, and social media content have all been recognized as admissible evidence in international criminal trials.[4] Investigators, lawyers, and judges are also becoming more accepting of the use of video conferencing platforms to conduct witness interviews or provide testimony. The ability to gain virtual access to witnesses and allow them to share their experiences with investigators thousands of miles away in The Hague is a truly groundbreaking development in international legal practice. However, relying on information coming out of a conflict zone without having been there in person raises fresh concerns about the ability of investigators to assess the credibility of witnesses and the reliability of information viewed through a digital prism. These developments raise a key question: When the information environment is itself a domain of battle, how can investigators separate fact from fiction to establish the truth?

3 Christian Perez and Anjana Nair, "Information Warfare in Russia's War in Ukraine: The Role of Social Media and Artificial Intelligence in Shaping Global Narratives," *Foreign Policy*, 22 August 2022, https://foreignpolicy.com/2022/08/22/information-warfare-in-russias-war-in-ukraine/.

4 *Prosecutor v. Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu and Narcisse Arido (Bemba et al.)*, Decision on Requests to Exclude Dutch Intercepts and Call Data Records, ICC-01/05-01/13-1855, TC VII, 26 April 2016; *Bemba et al*., Decision on Requests to Exclude Western Union Documents and other Evidence Pursuant to Article 69(7), ICC-01/05-01/13-1854, TC VII, 29 April 2016; *Prosecutor v. Ahmad Al Faqi Al Mahdi*, Judgement and Sentence, ICC-01-12-01/15-171, TC VIII, 27 September 2016; *Prosecutor v. Salim Jamil Ayyash, Hassan Habib Merhi, Hussein Hassan Oneissi, Assad Hassan Sabra (Ayyash et al.)*, Trial Judgment, STL-11-01/T/TC, 18 August 2020.

This paper begins with an overview of the role of growing importance that digital evidence is playing in war crimes investigations and prosecutions. This section is followed by an analysis of ICT-enhanced information operations in the Russia–Ukraine conflict and the various challenges that are emerging for war crimes investigators as a result. It then introduces the reader to some of the tools that are emerging to help war crimes investigators grapple with and overcome the challenges, including the *Berkeley Protocol on Digital Open Source Investigations*, which is currently being used by war crimes investigators in Ukraine.

2. THE CHALLENGE OF INVESTIGATING WAR CRIMES AND THE PROMISE OF DIGITAL EVIDENCE

Investigating war crimes has always been challenging for a variety of reasons, from the complexity of cases involving thousands of victims and witnesses to the difficulty in obtaining battlefield evidence. The first trial at the International Criminal Court (ICC), *Prosecutor v. Lubanga*, brought these challenges into stark focus.[5] Due to security issues in the locations relevant to their investigation, ICC investigators were unable to visit many of the crime scenes. Instead, they relied on intermediaries—locals from the area—who could help them find witnesses to the events.[6] At trial, it was revealed that several of the witnesses had been paid to give false testimony by the intermediaries. Without realizing it, the prosecution had brought unreliable narrators and a tainted witness pool before the court.

The prosecution was also overly dependent on witness statements taken by United Nations (UN) investigators, who had prior access to relevant individuals.[7] These statements had not been taken for the purposes of a prosecution, and investigators had been unable to find the original witnesses, validate their statements, and get their consent for use in court. This posed a problem at trial, since the prosecution did not have the authority to disclose the statements to the defense, although they were already relying on them in their case. The trial was stayed for several months as a result. Moreover, ICC investigators did not visit and forensically examine the crime scenes until they were nearing the trial phase. When they did finally visit the relevant geographic locations, they discovered that some of what their witnesses had testified to was not accurate. The very first ICC case came close to getting dismissed, which raised larger questions about the ICC Office of the Prosecutor's (OTP) ability to do fulfill its mandate. The inability of ICC investigators to visit the territory where the crimes occurred was not unique to the Democratic Republic of Congo in the *Lubanga*

5 Aliza Shatzman, "The Prosecutor v. Thomas Lubanga Dyilo: Persistent Evidentiary Challenges Facing the International Criminal Court," *George Mason International Law Journal* 12, no. 2 (2021).

6 Caroline Buisman, "Delegating Investigations: Lessons to be Learned From the Lubanga Judgment," *Northwestern University Journal of International Human Rights*11, no 3 (2013).

7 Christodoulos Kaoutzanis, "A Turbulent Adolescence Ahead: the ICC's Insistence on Disclosure in the Lubanga Trial," *Washington University Global Studies Law Review* 12, no. 2 (2013).

case. For years, investigators could not enter the territory of Sudan to investigate its head of state, Omar al Bashir, and other accused persons. Today, investigators face similar circumstances in Myanmar and Burundi.

These immense obstacles led the OTP to look for other solutions in its investigative work. Advanced digital technologies that could facilitate remote investigations, such as high-resolution satellite imagery or call data records, seemed like an answer to the problem. With the introduction of smartphones, internet connectivity, and social media, suddenly individuals within the conflict zone could capture what was happening on the ground and share it with the outside world. This was incredibly promising for the prosecutor's office, which could hire analysts to collect and review the data, thus moving the investigation forward even while their access was blocked.

The appeal of remote, technology-enabled investigation tactics was not limited to ICC investigators. About a decade ago, the international criminal justice community began to recognize the potential for utilizing user-generated content in their work.[8] This interest in digital open-source information, especially social media content, was initially driven by the conflicts in Syria, Iraq, and Libya, where smartphones and social media platforms became primary tools for war documentation. These tools allowed, and continue to allow, witnesses and first responders to record atrocities as they unfold, often in places where it is difficult, if not impossible, for international investigators to gain access.

These contemporary conflicts raised new and important questions about the possibility of investigating entities to acquire, analyze, and authenticate large volumes of digital information. The internal conflict in Myanmar, in which social media was used as a tool to incite violence against the Rohingya minority, further demonstrated how this type of digital information could provide evidentiary value by establishing the criminal intent of the perpetrators. In such cases, the propaganda and hate speech itself served as critical evidence in building a case against the Myanmar military for several crimes against humanity, including persecution.

However, this new source of potential evidence came with its own challenges. The volume of digital information online was immense, and the anonymous nature of the internet made it difficult to verify or even identify the information source. Thus, during these conflicts, conversations in the international criminal law community focused on how new technologies could improve investigative practice, leading to a series of workshops and the drafting of the *Berkeley Protocol on Digital Open Source Investigations*,[9] a UN manual co-published by the Office of the High Commissioner for Human Rights and Berkeley Law's Human Rights Center. The issues also led

8 Rebecca J. Hamilton, "User Generated Evidence," *Columbia Journal of Transnational Law* 57 (2018): 1–61.

9 The author of this paper led the drafting of the *Berkeley Protocol*.

to conversations about how new technologies, such as machine learning (ML) and artificial intelligence (AI), could be used to improve and enhance investigative practice—for example, the use of natural language processing to assist in document review and the use of object recognition technology to assist with the imagery analysis. The field was evolving, and then Russia invaded Ukraine, the digital battlefield exploded, and a whole new set of challenges began to emerge.

3. INFORMATION OPERATIONS IN THE RUSSIA–UKRAINE CONFLICT AND EMERGING CHALLENGES FOR INVESTIGATORS

As Russian forces closed in on Kyiv, launching the Kremlin's initial offensive against the capital city in late February 2022, a MiG-29 Fulcrum shot down six Russian planes. The next day, the same Ukrainian pilot shot down ten more Russian jets. Stories of this Ukrainian pilot spread and amplified on the internet, quickly turning him into a living legend. But he was not living, or even real. Rather, he was a fiction created to strike terror into Russian forces, fortify the resolve of Ukrainian citizens, and amaze the world with Ukraine's unexpected strength and courage. Now recognized as propaganda, stories of the "Ghost of Kyiv" were believed by many, until the Ukrainian Air Force ultimately admitted that this character was created as part of an influence campaign.[10] Nevertheless, his legend lives on in murals and other artwork commemorating the hero pilot.[11]

The Ghost of Kyiv is one of several examples of influence operations in the Russia–Ukraine conflict, which are the focus of this section. Since a significant amount of scholarship and commentary has already been written on the topic generally, this section focuses on a few of the emergent trends seen in the conflict that directly affect war crimes investigations. This includes the way in which military forces are engaging in content creation to influence or deceive not only enemy forces but individuals beyond the battlefield; the ability of governments to create chaos through the proliferation of competing narratives, while also controlling information flows to specific audiences; and the parallel trends of exposing closed-source information through hack-and-leak operations on the one hand and censoring information through internet shutdowns on the other.

A. Influence and Deception

With the global popularity of social media, parties to a conflict have a much larger potential sphere of influence, with a multitude of platforms through which they can

10 Lateshia Beachum, "The 'Ghost of Kyiv' Was Never Alive, Ukrainian Air Force Says," *Washington Post*, 1 May 2022, https://www.washingtonpost.com/world/2022/05/01/ghost-of-kyiv-propaganda/.

11 "Ghost of Kyiv Mural Unveiled in Ukrainian Capital in Celebration of Aviation Day," Yahoo! News, 27 August 2022, https://news.yahoo.com/ghost-kyiv-mural-unveiled-ukrainian-200800183.html.

reach a broad audience. To exploit social media successfully, however, the parties need to create interesting, engaging, and emotionally driven content to capture the public's attention. The ability to produce this type of online content requires specialized skills that are not traditionally associated with the armed forces.

From the start of the full-scale invasion in February 2022, the Ukrainian government and military have demonstrated an adeptness for conducting creative information operations, successfully using social media to win the hearts and minds of the Western world and, in so doing, gaining the political and financial support necessary to sustain their fight against Russian forces. From the start of the conflict, President Volodymyr Zelensky has used digital media to connect with the people of Ukraine. His nightly addresses have played an important role in boosting the morale of the Ukrainian people and soliciting support from the rest of the world. The Ukrainian president's videos represent only a small fraction of the videos, images, and memes generated by the Ukrainian government and military, distributed across a range of social media platforms, including Twitter, Facebook, Telegram, and TikTok. For example, the official Twitter account of Ukraine's Ministry of Defense posts a persistent stream of content that can be sincere and heart-breaking one minute and sassy and biting the next.[12] The account posts well-produced, professional-looking videos set to music that encourage sympathy for Ukrainians while antagonizing their Russian invaders. In one case, they adroitly used close-up imagery of a bombed-out playground for a campaign that generated millions of dollars from the public to buy kamikaze drones.[13] Similar imagery depicting the destruction of schools has been widely circulated online, leading some investigators to quickly conclude that these attacks were war crimes. However, such photographs can be misleading based not on what the photographer captures in the frame but on what they leave out. In some instances, panning out reveals a military base or indicators that the school was being used for military purposes. Thus, while the content from these accounts may not be manipulated or altered, the framing is far from neutral and objective. Rather, it is intended to influence the consumer. War crimes investigators are not immune from this influence.

The Ukrainian approach to propaganda, using real imagery in clever ways, stands in contrast to the favored Russian tactic, which relies on falsified narratives, fake content, and disinformation to amplify emotions, stoking fear and fueling hatred. Russia has been engaging in these tactics for many years as part of its geopolitical agenda against the West, but the resources put into it and the sophistication have grown with the use of information operations troops within the Russian military apparatus. Generally, rather than focusing on the quality of content with catchy phrases, popular songs, and humor, the Russian government and military propaganda apparatus patently produces false

12 See "Defense of Ukraine," Twitter, https://twitter.com/DefenceU.

13 Daniel Boffey, "Ukraine Crowdfunding Raises Almost $10m in 24 Hours to Buy Kamikaze Drones," *The Guardian*, 12 October 2022, https://www.theguardian.com/world/2022/oct/12/ukraine-crowdfunding-kamikaze-drones-russian-attack-cities-military.

claims and conspiracy theories intended to convey an inaccurate account of events to the consumer. Rather than sharing this information through official channels, it often disseminates it through proxies, such as fake websites made to look like traditional and reputable news sources. As with influence campaigns, even well-trained war crimes investigators are susceptible to being fooled by these deception tactics.

Both parties in the Russia–Ukraine conflict are creating social media content that is designed to go viral. The speed at which information spreads across the internet exacerbates the challenges for investigators in a variety of ways. First, false information travels faster than facts, as one Twitter-based study revealed.[14] This phenomenon means that investigators monitoring social media are likely to see the false version of events before seeing the accurate version. In addition, thoroughly fact-checked news stories take longer to be published than unverified ones based on speculation rather than hard facts. This is problematic because even trained investigators are susceptible to anchor bias, which describes "people's tendency to rely too heavily on the first piece of information they receive on a topic."[15] In addition to issues of bias, the speed at which online information is shared and the constant stream of content tend to create a sense of urgency and anxiety that may cause investigators to shorten or altogether skip the verification process. This means that international criminal investigators need a high degree of digital literacy, skepticism, and understanding of digital culture to do their job effectively. It also means that digital investigators need time to do their jobs well.

B. Chaos and Control

Russia's longtime go-to tactic for information warfare has been to create chaos and confusion by overwhelming the information space with a high volume of conflicting stories about a single event. Newer technologies, such as automated botnets paired with artificial intelligence, now generate content to inundate online platforms with conflicting narratives.[16] The ease with which digital information can be quickly created, altered, repurposed, or amplified is unique to our modern world in which Hollywood special effects are affordable and commercially available to everyone on a smartphone, botnets can be used to control thousands of accounts at once, and artificial intelligence has been optimized to generate fake videos that are indistinguishable from real ones. Fake imagery and audio recordings have advanced so much that it is often difficult to tell them apart from the real thing. The quality of fake imagery and the amplification of false narratives online is not intended to deceive but to undermine trust more generally so that people begin to believe that nothing is real and nothing

14 Larry Greenemeier, "False News Travels 6 Times Faster on Twitter than Truthful News," *Scientific American*, 9 March 2018, https://www.pbs.org/newshour/science/false-news-travels-6-times-faster-on-twitter-than-truthful-news.

15 Kassiani Nikolopoulou, "What is Anchoring Bias? Definition and Examples," *Scribbr*, 16 December 2022, https://www.scribbr.com/research-bias/anchoring-bias/#:~:text=Anchoring%20bias%20describes%20people's%20tendency,anchor%2C%20to%20make%20subsequent%20judgments.

16 Paul Szoldra, "Deepfakes Are Russia's New 'Weapon of War'," *Ruck*, 20 November 2022.

can be trusted. This lack of trust can create an even greater problem for investigators and lawyers who must convince judges to trust the evidence. Thus, while skepticism is important, investigators need to find a way to properly convey when content is reliable and when it is not. Deepfakes also raise concerns about "the liar's dividend,"[17] which might allow war criminals to evade accountability by claiming that real content is fake.

In contrast to the everything, everywhere, all the time approach to information operations, the architecture of the internet and diversification of platforms provide for very precise and selective information targeting. New digital technologies increase the ability to design messaging to target specific audiences. With traditional media, the same news or information was generally distributed to all recipients equally, but digital media operates differently. As the world learned from the Cambridge Analytica scandal involving digital consultants to Donald Trump's 2016 presidential campaign, anyone can pay social media platforms to micro-target messages to specific users. Micro-targeting is defined as "a marketing strategy that uses consumer data and demographics to identify the interests of specific individuals or very small groups of like-minded individuals and influence their thoughts or actions."[18] This ability to distribute targeted information to specific communities gives parties unprecedented control not only over the information shared but over how it is shared and who sees it. For example, Russian information operations troops share different messages with Western countries than they do with the BRICS countries (Brazil, India, China, and South Africa alongside Russia) or with the Russian people.[19] Depending on where you live—both geographically and on the internet—a person may have very different perceptions of what is happening in the Russia–Ukraine conflict.

The parties to this conflict are using parallel tracks—one that uses voluminous, fast-paced, and chaotic distribution of content to overwhelm internet users and another that uses information silos and micro-targeting to send precisely crafted messages to very specific audiences. These dual tactics are confounding war crimes investigators, who, on one hand, must sort through an unmanageable firehose of information to find the "signal in the noise" and, on the other hand, must actively go hunting in different online communities and forums to ensure they are getting a full picture of what is happening. Thus, the volume of information requires that investigator to search for evidence in an endless ocean of information, while the siloing of information necessitates that investigators search for evidence in the equivalent of a thousand rivers.

17 Kaylyn Jackson Schiff, Daniel S. Schiff, and Natalia Bueno, "The Liar's Dividend: Can Politicians Use Deepfakes and Fake News to Evade Accountability?" SocArXiv Papers, 10 May 2022, https://osf.io/preprints/socarxiv/q6mwn/.

18 Linda Tucci, "Microtargeting," *TechTarget*, February 2013, https://www.techtarget.com/searchcio/definition/microtargeting.

19 "The GRU's Galaxy of Russian-Speaking Websites," *Open Facto*, 27 January 2022, https://openfacto.fr/2022/01/27/the-grus-galaxy-of-russian-speaking-websites/.

C. Exposure and Concealment

Two other notable trends in information operations in the Russia–Ukraine conflict are the hacking, leaking, and exposure of private information versus the censoring of information by blocking websites or internet connectivity. Hack-and-leak operations are defined as operations in which "malicious actors use cyber tools to gain access to sensitive or secret material and then release it in the public domain."[20] Internet shutdowns are understood as "state-enforced disruptions of internet access aimed at controlling the flow of information."[21]

Since the February 2022 invasion, on an almost daily basis, there have been new online leaks of documents and datasets alleged to be from Russian government agencies and private businesses. One month into the invasion, the Secret Service of Ukraine published the names of 620 alleged agents of Russia's Federal Security Service, presumably obtained through hacking.[22] Similarly, a website called Distributed Denial of Secrets (DDoSecrets) started releasing regular document dumps to their email subscriber list. In less than two months, two million emails from Russian government and private entities were leaked, making them accessible to any member of the public. These online leaks are high volume and unlikely to have been reviewed in full by anyone before their publication. In some cases, the parties to the conflict have openly leaked private documents themselves, while at other times they have used proxies. There has also been a significant amount of leaking coming from anonymous sources and third parties. In addition to the questions around the legality of acquisition, which could lead to the exclusion of evidence in court, online leaks are extremely difficult to authenticate and can be laced with malware or contain strategically placed false information.[23]

If leaked documents are, in fact, authentic, they could serve as a fruitful source of evidence for war crimes investigators. However, like everything else on the internet, leaked documents must be handled with caution and viewed with skepticism. These document dumps could easily contain false information designed to mislead. As DDoSecret explains, datasets released during war have "an increased chance of malware, ulterior motives and altered or implanted data, or false flags / fake personas." There have already been examples of tainted leaks, in which hackers manipulated the

20 James Shires, "Hack-and-Leak Operations and U.S. Cyber Policy," *War on the Rocks*, 24 August 2022, https://warontherocks.com/2020/08/the-simulation-of-scandal/.

21 "The Impact of Internet Shutdowns on Human Rights Defenders in India," *American Bar Association*, 14 November 2022, https://www.americanbar.org/groups/human_rights/reports/india-internet-shutdowns/#:~:text=Internet%20shutdowns%20are%20state%2Denforced,controlling%20the%20flow%20of%20information.

22 "Ukraine Intelligence Publishes Names of 620 Alleged Russian Agents," *Reuters*, 28 March 2022, https://www.reuters.com/world/europe/ukraine-intelligence-publishes-names-620-alleged-russian-agents-2022-03-28/.

23 Lindsay Freeman "Hacked and Leaked: Legal Issues Arising from the Use of Unlawfully Obtained Digital Evidence in International Criminal Cases," *UCLA Journal of International Law and Foreign Affairs* 25, no. 2 (2021): 45.

documents before sharing them publicly online.[24] In addition to the very real risks of embedded malware and implanted false information, these documents come without any of the contextual information or metadata needed to authenticate them for use in legal proceedings. With all these new types of digital evidence that have not been tested in court, war crimes investigators face a mammoth challenge in determining what information will be required to authenticate the evidence and whether it will be found admissible by a future court.

In contrast to the approach of openly sharing information for strategic advantage, governments can also do the opposite. As an authoritarian regime, the Kremlin has used its monopoly over the media in Russia and Russian-occupied parts of Ukraine as its main propaganda tool. By controlling the content distributors and regulating what they share, the government can manipulate its audience. This carefully curated information is strengthened by the elimination of competing views, which can be achieved by buying out or shutting down independent news sources, blocking access to certain websites, and causing internet blackouts at opportune times. Russia uses its control over the information infrastructure, including radio, television, and internet access, to tactically deprive people of access to competing views and ensure its propaganda is the only information available to its intended audience.

The censorship of information, particularly through government control of the internet, which can be shut down relatively easily, creates an issue for war crimes investigators since these shutdowns can cut off the distribution of real-time information sharing from inside the conflict to the outside world. If a government shuts down the internet at the same time as its military forces are overtaking a village and killing civilians, then witnesses and journalists are unable to share photographs, videos, and accounts of what is unfolding, leaving a dearth of evidence for events that have occurred during digital blackouts. Since investigators are led by the evidence, due to internet shutdowns they may focus too heavily on big events with lots of documentation and ignore atrocities that are not captured digitally.

The use of the internet as a domain of battle illuminates the potential pitfalls and digital tripwires that can ensnare and confound modern war crimes investigators. These traps include the problem of investigators getting caught in information silos and failing to account for cognitive or algorithmic biases when sorting through and analyzing digital content. War crimes investigators are not immune from entrapment in these information silos, a hazard that is especially dangerous if they lack self-awareness. Therefore, while investigators should recognize the value of digital open-source information for intelligence, lead information or even evidence, it is necessary to temper enthusiasm for this supply of data with a healthy skepticism and an active awareness of the potential pitfalls of relying too much on digital information sources.

24 Adam Hulcoop et al., "Tainted Leaks: Disinformation and Phishing with a Russian Nexus," Citizen Lab, 25 May 2017, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

4. DEVELOPING TOOLS FOR WAR CRIMES INVESTIGATORS TO SEE THROUGH THE DIGITAL FOG OF WAR

While there is no easy solution to the above-described issues, several tools could help war crimes investigators in their fight against these growing challenges. The tools include both the creation of standards and guidelines, along with training for investigators and experimentation and application of technological solutions. This section focuses on the *Berkeley Protocol*, the first international standard and guidance for using open-source digital information in the investigation of war crimes,[25] and the use of machine learning and artificial intelligence in investigations.

A. International Investigative Standards

The English-language version of the *Berkeley Protocol* was published in December 2020, and many international organizations and civil society groups received training on it the following year. While there were several ongoing non-international armed conflicts during this period, it was not until Russia's full-scale invasion of Ukraine on 24 February 2022 that the protocol's dissemination and adoption picked up steam.

As the first major conflict to break out since the advance publication of the *Berkeley Protocol* in December 2020 (it will not be officially launched until it is available in all six UN languages, which will occur in mid-2023), the Russia–Ukraine conflict serves as a primary test case as to whether such standards can help address the investigation and legal challenges of a contested information battlefield.

As soon as Russia invaded, the Office of the Prosecutor General of Ukraine took the initiative to translate the protocol's text into Ukrainian. The translated document was distributed to others engaging in the documentation and investigation of what was unfolding in Ukraine. Two weeks into the conflict, the prosecutor general announced on Twitter that her office was using the *Berkeley Protocol* in their investigative work.[26] Soon after, the National Police of Ukraine and, separately, a consortium of Ukrainian civil society groups called the 5 AM Coalition received training on digital

25 United Nations Office of the High Commissioner for Human Rights and Human Rights Center, UC Berkeley School of Law. *Berkeley Protocol on Digital Open Source Information*, (December 2020); Alexa Koenig, *The New Forensics: Using Open Source Information to Investigate Grave Crimes* (Berkely, CA: Human Rights Center, UC Berkeley School of Law, 2018); Stefano Trevisan, "Open-Source Information in Criminal Proceedings: Lessons from the International Criminal Court and the Berkeley Protocol," *Giurizprudenza Penale* 4 (2021): 9–10; Sam Dubberley, Alexa Koenig, and Daragh Murray, eds. *Digital witness: using open source information for human rights investigation, documentation, and accountability* (New York, NY, Oxford University Press, 2020); Daragh Murray, Yvonne McDermott, and Alexa Koenig, "Mapping the Use of Open Source Research in UN Human Rights Investigations," *Journal of Human Rights Practice* 14, no. 2 (2022): 554-81; https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source.

26 Edward Lempinen, "In Ukraine, Berkeley Experts Are Shaping the Legal Fight Against War Crimes," *Berkeley News*, 21 February 2023, https://news.berkeley.edu/2023/02/21/in-ukraine-berkeley-experts-are-shaping-the-legal-fight-against-war-crimes/.

open-source investigations based on the protocol's methodology. In September, the ICC Prosecutor and Eurojust launched "practical guidelines for documenting and preserving information on international crimes," which endorsed the *Berkeley Protocol*.[27]

The speed of the *Berkeley Protocol*'s adoption and the near-unanimous and immediate consensus around its use is a success story in and of itself, aligning a diverse and complex ecosystem of actors who traditionally have not always worked well together.[28] Common standards and definitions are an important way for different groups with different approaches and goals to communicate successfully. Therefore, rather than engaging in crosstalk or remaining in silos, international investigative entities—from civil society documenters and human rights researchers to police and prosecutors—are now, with the guidance of the protocol, getting on the same page. The availability of the document to civil society organizations, which increasingly want to support prosecutors in their pursuit of justice and accountability for war crimes, was also an important watershed in the professionalizing of their work and getting the work on civil society organizations recognized by prosecutors.

In terms of the protocol's substantive guidance, it is too early to assess definitively whether it has succeeded in improving the quality and accuracy of digital investigations in Ukraine. That test will come when the evidence collected today is introduced into court in future trials.

While it has been helpful in this regard, the protocol provides a broad framework that must be adapted to specific operational contexts. To reach a diverse audience in different jurisdictions, the protocol was written as high-level guidance and, in order to future-proof the document, it was intentionally designed to be technology agnostic. Therefore, to be fully effective, the protocol needs to be supplemented with standard operations procedures that are context-specific and technology systems and tools to support the process. Digital evidence collection, preservation, and analysis processes perform best when calibrated to the unique requirements of specific environments and circumstances.

B. Advanced Digital Technologies

The mass adoption of the *Berkeley Protocol* has launched a new dialogue about the most appropriate and effective digital tools to assist prosecutors with these challenges. In particular, there has been a growing desire for technology solutions like the use of

27 "ICC Prosecutor and Eurojust Launch Practical Guidelines for Documenting and Preserving Information on International Crimes, *International Criminal Court*, 21 September 2022, https://www.icc-cpi.int/news/icc-prosecutor-and-eurojust-launch-practical-guidelines-documenting-and-preserving-information.

28 Stephen J. Rapp, "Bridging The Hague - Geneva Divide." *The Hague Institute for Global Justice*, 13 January 2017, https://thehagueinstituteforglobaljustice.org/nding-accountability_0bovycq8ok8pjy3pcggx3v/.

artificial intelligence—mainly natural language processing, object recognition, and facial recognition—to sort through the vast quantities of material.

While the application of natural language processing, object recognition, and facial recognition have been experimented with in criminal investigations for some time now, it has taken a while to develop the technology for the context of war crimes investigations. While natural language processing, a branch of artificial intelligence "concerned with giving computers the ability to understand text and spoken words in much the same way human beings can,"[29] has worked well in the more widely spoken languages for some time, it still struggles with rarer languages, localized dialects, and languages written in other scripts like Cyrillic. Similarly, object recognition works well for everyday objects for which there is a lot of training data, like cars, but it is less reliable when it comes to tanks, drones, and weapons in the field. While facial recognition on CCTV works well, it is far less effective when used on hand-held footage or video with occluded faces, which is generally the type of material handled and used by war crimes investigators. There are also many current efforts to develop deepfake detection and other technology tools that will assist in the verification process.

These technologies show promise for assisting investigators in their tasks, but they cannot and should not be seen as something that can replace the work of human investigators. Prosecutors might rely too heavily on them and trust them too readily. There can be bias in the training itself, in the collection of data, and in the fact that sometimes the technology simply gets it wrong. More importantly, many of these tools are still experimental and have not advanced to the stage in which full confidence can be placed in them when a person's life and the legitimacy of the justice system are on the line.

5. CONCLUSION

In less than a decade, the use of digital evidence in international criminal investigations and trials has evolved significantly, and so too have the challenges of making this type of evidence effective in court.[30] Digital technologies are developing at such a rapid pace that there are already complicated new issues arising in the Ukraine conflict that are not addressed in any currently accepted guidance.

29 "What is natural language processing?" IBM, https://www.ibm.com/topics/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can, accessed 8 April 2023.

30 Lindsay Freeman and Raquel Vazquez Llorente, "Finding the Signal in the Noise: International Criminal Evidence and Procedure in the Digital Age," *Journal of International Criminal Justice* 19, no. 1 (2021): 163–88.

The International Criminal Court and other international justice mechanisms are often criticized for the length of their proceedings. As a result, the use of automated tools, artificial intelligence, and other technology hacks becomes an appealing option for sorting through the unprecedented volume of potentially relevant digital material. However, the very complexities of the information environment necessitate a slow, deliberate, and thorough approach to reviewing digital evidence. War crimes investigators should view technological assistance as providing a useful support function, not as a shortcut that minimizes their effort.

While the *Berkeley Protocol* and the increasing sophistication of investigators mark positive progress in the field, the ongoing conflict in Ukraine reveals the growing need to also recognize the harms and dangers raised by the use of and reliance on new technologies. There is no one solution that will be able to address the multitude of complex ways in which digital technologies are exploited to advance the military and political agendas of Russia, Ukraine, and all the third parties with a stake in this conflict. As a result, investigators need to understand the online environment in which they work as dynamic, constantly changing, and requiring a level of flexibility from war crimes investigators.

ACKNOWLEDGMENTS

The author sincerely thanks Dr. Alexa Koenig, Taťána Jančárková, and the other CyCon reviewers for their thoughtful, candid, and constructive feedback, and the Ukraine investigation team at Berkeley Law's Human Rights Center, whose research and engaging conversations inspired and contributed to this article.