Centre for Emerging Technology and Security

EXPERT ANALYSIS

The Information Battlefield

Disinformation, declassification and deepfakes

The views expressed in this article are those of the authors, and do not represent the views of the Alan Turing Institute or any other organisation.

Introduction

Why is the information domain so important to the Ukraine conflict?

Because wars are not won by military hardware alone: as Russia is finding in Ukraine, superiority in military capability is not enough to assure success. Wars are won by human actions and human decisions. Some decisions are taken on the battlefield – to stand and fight, or to flee. But many crucial decisions are taken elsewhere: by foreign political leaders, who must decide how far to go in defence of their values and interests in supporting either side. In democracies, politicians cannot go further than the public will support, so individual citizens' beliefs count, too. In every case – on and off the battlefield – beliefs drive behaviour.

The information domain has been a key battleground in Russia's war in Ukraine. Every significant actor with a stake in the war – whether involved directly in the fighting or not – is engaged in a constant battle to shape beliefs about the conflict, in the hope of influencing the actions of others to their advantage. All information campaigns, whether based on truth or on deliberate lies, share a common goal: to embed beliefs that will ultimately shape behaviour. All successful campaigns leverage some combination of cognitive, social and affective emotional factors. And once a belief is formed it can be stubbornly difficult to change.[i]

In this conflict, we have seen Russia use a familiar playbook of methods, both overt and covert, in the information domain. Disinformation – the deliberate generation or amplification of false information – has played a significant role. New and emerging technologies are becoming increasingly prominent, both in enabling the rapid proliferation of disinformation in the digital sphere, but also in providing new opportunities to counter its spread. To enable effective responses, it is important to dissect why Russia has deployed disinformation tools in the way that it has, and how successful it has been.

Targets and Tools

Who is Russia seeking to influence?

The Russian government's greatest fear is popular unrest at home: thus the Russian public is the first and most important audience for Russian disinformation, with the main objective being to shape attitudes in line with the Kremlin's desired course of action. A second major focus of Russia's information operations is Ukraine – both the Ukrainian military (to erode morale and disrupt operations) and the broader population, with the desire of instilling fear and stoking disaffection with the Kyiv government. Challenging political consensus in the West is also an important goal, as is the promotion of pro-Russian and anti-Western narratives with other international audiences, including China, India, Africa and the Middle East.

A blunt instrument frequently used in Russian disinformation is denials and lies by Russian government officials – no special technical capabilities are required here. The most barefaced of Russia's lies has been to deny that there was a war going on at all. It framed action in Ukraine as a 'special military operation' with limited aims of protecting the Donetsk and Luhansk republics from the actions of a 'Nazi' Ukrainian regime egged on by NATO. Why? Russia hoped to maintain the false narrative of a limited military intervention, especially as it had not prepared its population for an extended, large-scale war. Russia also sought to blunt international condemnation of its aggression. This has been broadly unsuccessful, with 141 countries voting in the UN General Assembly to demand an end to Russia's offensive, and the UN's International Court of Justice issuing a legally binding declaration that Russia should immediately end its operations in Ukraine.

Russia has also mounted covert deception activity in the form of 'false-flag' operations intended to frame Ukraine for attacks in the Donbas, creating a pretext for Russia's intervention.[ii] These operations can achieve their influence aims, even if they fail to stand up to scrutiny under serious investigation. Emotive images of alleged attacks on civilian infrastructure in the Donbas received huge coverage in Russian state media in the run-up to the invasion, helping to justify military action among a domestic audience, rather than convince international observers. A series of Western intelligence releases 'pre-bunking' Russia's plans probably reduced their traction with international audiences.

Russia invaded. Information operations accompanied Russia's initial action, intended to sap the morale of Ukrainian Armed Forces. According to an adversarial threat report published by Meta (the parent company of Facebook), disinformation surged in the initial phase of the conflict. Russia-aligned hackers compromised dozens of social media accounts belonging to Ukrainian military personnel and attempted to upload video footage of Ukrainian forces surrendering.[iii] However, Russia underestimated the strength of anti-Russian feeling among Ukrainians and was unable to dent a steadfast and dogged Ukrainian resolve. In contrast, Ukraine's leadership and citizenry launched powerful social media campaigns. Engaging, emotive videos, showing against-the-odds resistance at Snake Island or the bravery of Ukrainian civilians in facing down Russian tanks, reached huge audiences, as did Zelensky's authentic, direct messages to the Ukrainian nation. Ukraine's 'will to fight' was quickly cemented into its national self-image and that of Western audiences.

Fighting back through open data

Disinformation campaigns leverage digital technology for unparalleled reach.

But technology can also be an effective tool in countering online disinformation. In the current phase of the conflict, the balance of advantage is with those who seek the truth about progress in Russia's campaign.

Data about this conflict have been more accessible to Western audiences than ever before. Commercial satellite imagery showed Russia's military build-up around Ukraine's borders in the weeks preceding the invasion. Video footage uploaded from smartphones and dashcams showed military equipment on the move. NASA's Fire Information for Resource Management System (FIRMS), designed to use satellite observations to detect active fires, showed in near-real time the location of thermal anomalies indicative of conflict. With Russian encrypted comms failing, Russian units resorted to high frequency radio communications: their communications could be intercepted and analysed using cheap, commercially available equipment and the results posted online.

Online forums have been a key enabler in creating and connecting communities of open source intelligence (OSINT) analysts. Crowd-sourced analysis can connect large populations of motivated, skilled observers with vast quantities of checkable data. The collective efforts of OSINT analysis, citizen-journalists and military enthusiasts to analyse the facts on the ground make it much more difficult for false information to remain uncontested.

The conflict has also seen unprecedented efforts by Western intelligence agencies to declassify and release sensitive material into the public domain.[iv] By outlining Russia's plans for invasion and revealing Russia's attempts at falsifying a pretext for action, these declassifications have helped to counter Russia's disinformation among Western audiences.[v] Journalists and open-source analysts have been able to match intelligence revealing Russia's intent with observable activity on the ground – making these declassifications more powerful.

Emerging technology is likely to further accelerate the success of OSINT analysis, by lowering the barriers to entry for analysts in interpreting publicly-available data. Machine learning capabilities can assist in automated recognition of military vehicles from satellite imagery or social media, reducing reliance on the rare talents of expert military analysts.[vi] Facial recognition software can enable the identification of soldiers and their association with specific units. Cloud based machine translation tools enable analysts to interpret text and audio even if they lack the relevant language skills. Open-source data analysis techniques can reveal patterns that are hallmarks of disinformation networks, such as similarities in the behaviour or language used by accounts across different platforms.

An interesting shift in Russian tactics is the recent emergence on pro-Russian social media channels of fake fact-checking outfits, claiming to debunk Ukrainian-generated 'fakes' showing destroyed Russian military units or devastation of civilian infrastructure by Russian strikes. The goal, according to researchers at Media Forensics Hub,[vii] is to inject doubt among Russian-language audiences as they encounter real images of wrecked Russian military vehicles. This is an evolution of an established Russian tactic, polluting the information environment with a range of sometimes-contradictory narratives. By increasing the sense of uncertainty about which sources to trust, audiences are left not knowing what to believe.

Outrage at Russia's actions has catalysed a more muscular response by Western social media companies, galvanising action on the policies, capabilities, and operations to take down or neuter sources of Russian disinformation. Social media companies have made it more difficult for disinformation to spread unchecked, through efforts to label state-sponsored content and prevent its monetisation. But while policies have evolved rapidly, enforcement has been piecemeal – one study released by the Centre for Countering Digital Hate reckoned that, of a sample of c.3600 recent articles posted by Russian state news sources, Facebook was failing to label 91% of the posts as state-sponsored.[viii] DemTech researchers also found that only about half of Russian embassy Twitter accounts were labelled in accordance with Twitter's policies.[ix] Nonetheless, even with patchy implementation, the dominance of US tech giants means their actions have a disproportionate effect among Western audiences.

Defence at home, new audiences abroad

Russia's fear of Western strengths on the information battlefield is evident in its extreme, defensive approach to its domestic information space.

Russia has taken unprecedented steps to insulate its population from anything that counters the state-sanctioned narrative of events. A combination of stiff legal penalties, bans on independent and foreign media, and the forced pull-out of Western social media companies has left most Russians dependent for news on channels that the Kremlin controls. Whether their support is synthetic or genuine, polling shows a large majority of Russians support the official narrative that that Russia's 'special military operation' is a necessary response to the threat from a 'Nazi' regime.[x] Russia's domestic audience will remain the first and most important target for Kremlin disinformation – and the primary purpose of these information operations continues to be to suppress the risk of popular discontent. Information control helps the Kremlin maintain support for its actions in the face of rising human and economic consequences of the war. The increasingly punitive costs of dissent play a role, too. The Kremlin's levers of domestic control allow it to re-frame events in the conflict to suit its goals - dialling escalatory tensions up, or down, as required.

Technology also presents some potential ways around Russia's lockdown of its information space: the use of privacy protecting, anonymous communication tools such as Tor enables access to alternative, independent news sources for the Russian people, bypassing the authoritarian controls of the Russian state. Even older technology such as shortwave radio has been revived, harking back to Radio Liberty & Radio Free Europe from the Cold War.[xi]

Outside the Western 'information theatre', Russia is having more success: audiences in China, India, Africa and the Middle East have a more sympathetic view of Russia's actions. New research suggests that non-Western countries may be a key target of pro-Russian online propaganda: research by Carl Miller at Demos suggests that a set of Twitter accounts in BRICS countries, which had previously tweeted about regional political issues, switched to propagating pro-Russian narratives through coordinated inauthentic activity as the war began.[xii]

Looking ahead

Russia's disinformation efforts against the West will continue. Looking ahead, we might make some educated guesses about how they could evolve.

We might see more sophisticated attempts at disinformation for battlefield effect. In early March, a poor-quality 'deep fake' video of Ukraine's President Zelensky briefly circulated, before being debunked by Meta.[xiii] It was unconvincing – showing an unusually wooden, stilted speech by Zelensky standing behind a podium, calling on Ukrainians to lay down their arms. But it nonetheless provides a window into how information operations could evolve. There is no obvious technical barrier to producing a more convincing fake video, given the large quantities of video and audio footage of Zelensky now available. There is growing evidence of deep-learning systems being used to generate credible audio, stills and increasingly video footage. The missing ingredients for a credible deepfake are probably the non-technical aspects. A more convincing operation might use an actor who could portray Zelensky's gestures more convincingly, and a format that mimicked his 'selfie' walkabout videos. Crucially, such an operation could choose a moment where a call for political compromise or surrender was more plausible, and have a series of channels ready to amplify the message.

Russian disinformation to complicate the attribution of responsibility for war crimes is already in full flight.[xiv] Images of atrocities have galvanised further sanctions, and with Russia intent on the use of brutal military tactics to break Ukrainian resistance, more evidence of atrocities is likely to emerge. A major evidence-gathering effort is already underway, with broad support from Western governments, to support a future judicial process. But any international war crimes trial will take time– and successful disinformation campaigns could muddy the waters in the court of public opinion, much sooner.

We should expect to see more use of disinformation to deter the West, inflating the threat of consequences for actions against Russia. Ukrainian resistance relies on Western help to resupply military forces and stave off economic and humanitarian collapse: if Russia can fracture this support it will be at a significant advantage. Russia will likely attempt to play on fears about the possibility of escalation to try to limit Western support to Ukraine. We might also expect to see Russia play up other threats, in particular, the threat of unconventional or asymmetric responses. These messages are not just aimed at political leaderships - mobilising popular concern is an aim in itself.

As some platforms strengthen their defences to disinformation, Russia will shift to other platforms and adapt its tactics. In particular, we can anticipate that Russia will focus on amplifying 'organic' narratives advanced by fringe actors or mainstream ones that suit its aims. Emerging technology tools to counter disinformation are of less help if the sources of false narratives are genuine, Western voices. Critical thinking – applying rigour to who and what you choose to believe – will always be the best defence.

About the Authors

This article has been authored by CETaS Experts, who have chosen to publish anonymously to articulate unhindered views.

References

i Ecker, U.K.H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., and Amazeen, M. A. 'The psychological drivers of misinformation belief and its resistance to correction', Nature Reviews Psychology 1, p. 13–29. Available at: https://www.nature.com/articles/s44159-021-00006-y

ii Madhani, A., Cook, L., and Fraser, S. 2022. 'US says new intel shows Russia Plotting false flag attack.' Available at: https://apnews.com/article/russia-ukraine-business-europe-belarus-jens-stoltenberg-43c9151532de706a2edec5684dfcf07d

iii Meta. 'Meta's Adversarial Threat Report, First Quarter 2022.' Available at: https://about.fb.com/news/2022/04/metas-adversarial-threat-report-q1-2022/

iv The Independent. 2022. 'Secret intelligence has unusually public role in Ukraine war.' Available at: https://www.independent.co.uk/news/world/americas/us-politics/ukraine-ap-vladimir-putin-russia-jeremy-fleming-b2049839.html

v Adam, K. 2022. 'How UK intelligence came to tweet the lowdown on the war in Ukraine.' Available at: https://www.washingtonpost.com/world/2022/04/22/how-uk-intelligence-came-tweet-lowdown-war-ukraine/

vi Gupta, R., Reed, C., Rohrbach, A., and Darrell, T. 2022. 'Accelerating Ukraine intelligence analysis with computer vision on synthetic aperture radar imagery.' Available at: https://bair.berkeley.edu/blog/2022/03/21/ukraine-sar-maers/

vii Silverman, C. and Kao, J. 2022. 'In the Ukraine conflict, fake fact-checks are being used to spread disinformation.' Available at: https://www.propublica.org/article/in-the-ukraine-conflict-fake-fact-checks-are-being-used-to-spread-disinformation

viii Center for Countering Digital Hate. 2022. 'Facebook failing to label 91% of posts containing Russian propaganda about Ukraine.' Available at: https://counterhate.com/blog/facebook-failing-to-label-91-of-posts-containing-russian-propaganda-about-ukraine/

ix Oxford Internet Institute. 2022. 'Russian propaganda targets non-Western audiences on Twitter; Chinese sources amplify Russian disinformation.' Available at: https://demtech.oii.ox.ac.uk/war-in-ukraine-and-disinformation-newsletter-12-april-2022/#continue

References

x Levada-Centre. 2022. ‘Putin’s approval rating.’ Available at: https://www.levada.ru/en/

xi Saunders, T. 2022. ‘BBC turns to dark web and shortwave radio to bring outside news to Russian people amid wartime crackdown.’ Available at: https://inews.co.uk/news/technology/bbc-dark-web-russia-censorship-putin-1534974

xii CASM Technology. 2022. ‘Message-based community detection on Twitter.’
Available at: https://files.casmtechnology.com/message-based-community-detection-on-twitter.pdf

xiii Hatmaker, T. 2022. ‘Meta takes down deepfake of Ukraine’s President Zelensky surrendering.’ Available at: https://techcrunch.com/2022/03/16/facebook-zelensky-deepfake/?guccounter=2

xiv Disinfo. 2022. ‘Disinformation to conceal war crimes: Russia is lying about atrocities in Bucha.’ Available at: https://euvsdisinfo.eu/disinformation-to-conceal-war-crimes-russia-is-lying-about-atrocities-in-bucha/#

Centre for Emerging Technology and Security

CETAS.TURING.AC.UK