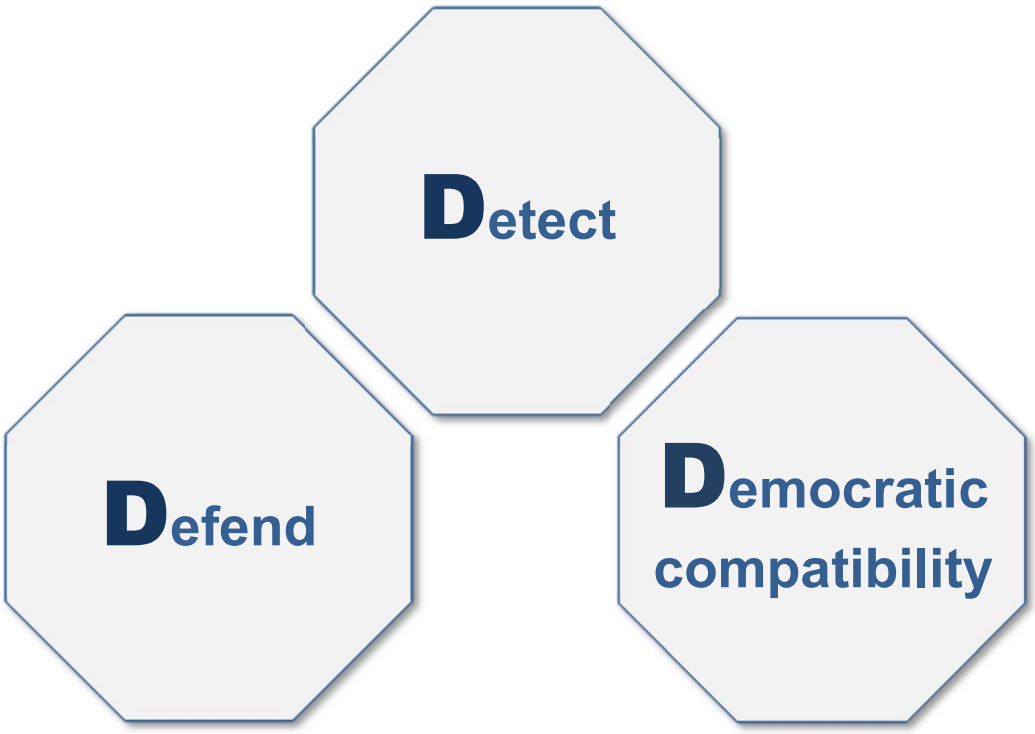
Intelligent Biology.

Cognitive defense of the Joint Force in a digitizing world

Report for the Pentagon Joint Staff Strategic Multilayer Assessment Group

Nicholas D. Wright

– v1 July 2021 –



The United States Department of Defense Joint Staff Strategic Multilayer Assessment Group sponsored this research as part of a project for Headquarters Air Force, to examine Integrating Information in Joint Operations (IIJO). For further information contact Intelligent Biology (www.intelligentbiology.co.uk).

This report is one of a coherent family of *Intelligent Biology* products that together provide a framework for successful influence across the spectrum of competition, including the Gray Zone. All are available www.intelligentbiology.co.uk, including:

- Wright, ND (2019) ***From Control to Influence: Cognition in the Grey Zone*****,** Intelligent Biology (v3).
- Ed. Wright ND, (2018) ***AI, China, Russia and the Global Order****: Technological, Political, Global, and Creative Perspectives,* U.S. Dept. of Defense Joint Staff.

About the author

Dr Nicholas Wright is affiliated with Georgetown University, University College London (UCL), Intelligent Biology and New America. He combines neuroscientific, behavioral and technological insights to understand decision-making in politics and international confrontations, in ways practically useful for policy. He regularly works with Governments. He has numerous academic and general publications. He received a medical degree from UCL, a BSc in Health Policy from Imperial College London, has Membership of the Royal College of Physicians (UK), has an MSc in Neuroscience and a PhD in Neuroscience both from UCL.

Acknowledgements

The author thanks Jason Spitaletta, Carl Hunt, Larry Kuznar, Natalie Thompson, Todd Veazie and others for helpful comments. All errors remain the author's.

Contents

Executive summary

How can we defend the humans in the Joint Force—and its key support networks—from adversarial information operations in our digitizing world?

Service personnel, their families and friends are human. Adversaries and other destabilizing forces threaten to sow discord and disruption amongst these humans, in order to degrade collective capabilities. Such threats can harness the powerful new digital technologies immersing our lives. Effectively defending the Joint Force's humans from information threats is crucial to protect its competitive capabilities: in our current era of Gray Zone competition, during escalation scenarios, and in war.

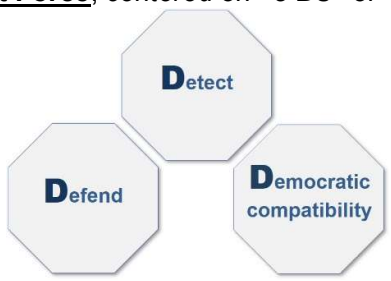
Part I defines the task. ***Who are we defending from what – and how is this new?***

"Who" includes humans like the 1.3 million active-duty personnel, and their families who provide key cognitive resilience and influences. Some four million hold clearances at secret or above. At home or work, on U.S. soil or abroad, on myriad digital devices, it's a huge attack surface with many tempting target audiences.

They face information threats from two sources: external adversaries' information operations (e.g. as NATO troops face from Russia; see Box 1); and extremists from their own society (e.g. as German Special Forces face now; see Box 2). Digitization changes the character of such threats, not least by merging "domestic" and "foreign": a key distinction in U.S. law, history and new *Interim National Security Strategic Guidance* (2021).

"Deepfakes" illustrate how such threats' character may evolve. These Artificial Intelligence (AI)-generated media can be tools of mass-produced disinformation, or of exquisite "active measures" (intelligence operations to shape political decisions, see p. 8). But deepfakes used alone exert limited influence. Instead, they are one tool in "combined arms" information operations alongside dual-use tech like micro-targeting (analyses of personal data to identify a specific audience's interests in order to influence their actions). This illustrates the real threat to the Joint Force: a vast market failure where huge commercial spending builds profitable dual-use *offensive* tech (like micro-targeting) but little for *defense* that's mostly just a business cost.

Part II is an evidence-based response for the Joint Force, *centered on* ***"3 Ds"*** *of* ***Detect, Defend,*** *and* ***Democratic compatibility****.*



1. DETECT: The U.S. must build capabilities to detect and characterize adversary influence operations against the Joint Force – who is targeted, by what means and for what purposes? They must work at multiple scales.

1.A. Detecting coordinated campaigns

Recommendation 1.A.i. *Build the integrated human, AI and organizational capabilities for* ***counter-intelligence at scale****, which can detect adversary information operations at the scales of relevance (e.g. a few to millions of people).*

Recommendation 1.A.ii. *The* ***"seam" between "domestic" and "foreign"*** *should remain a very clear distinction but it also presents a challenge for detecting threats that straddle that seam. Clear responsibilities should exist for how organizations coordinate to detect such threats, alongside dedicated resources.*

1.B. Detecting specific instances of mis- and dis-information

Recommendation 1.B.i. *Build the integrated human, AI and organizational capabilities to detect specific instances of mis-/dis-information (e.g. deepfakes).*

1.C. Detecting how adversaries shape the terrain over years

Recommendation 1.C.i. *Identify long-term risks to the Joint Force's humans as other nations, beyond trusted allies, shape the information tech immersing their lives (e.g. China's* ***TikTok*** *controls algorithms and big data on half of U.S. youths).*

2. DEFEND: Humans' cognition will always contain vulnerabilities as targets for disruption (e.g. work and life stresses or grievances), which the Joint Force can help minimize and so deny to others. Social, family and mental support matters, as does "media education" although it is no panacea. Individuals must also be given the technological means to defend themselves online, for which low cost, practical options exist. But the individual scale is not enough.

Mass personalization of influence operations is coming (as seen for retail and healthcare) in which personal data will be a key weapon and must be defended (e.g. sensitive medical data; or data collected by China's TikTok). Damaging personal data on millions in the Joint Force or their families can be injected into their social networks. New human-AI teams and organization are needed at multiple scales:

2.A. Defending at the individual human scale

Recommendation 2.A.i. *Enhance* ***social, family and mental support*** *with a particular focus on predictable periods of vulnerability (e.g. moving postings).*

Recommendation 2.A.ii. *Give individuals the technological* ***means to defend themselves*** *online. Low cost, practical options exist.*

Recommendation 2.A.iii. *Training and practical help concerning social media security (e.g. use high privacy settings and do not use apps like TikTok), the clear expectations set by the Joint Force (e.g. on privacy settings, or legal requirements for political posts) and practical ways to evaluate social media content.*

2.B. Defending individuals at the organizational scale

Recommendation 2.B.i. *Defend the data of individuals in the Joint Force and its support networks – and, against conventional wisdom,* ***build silos*** *for their data.*

Recommendation 2.B.ii. *Prepare defenses for adversary campaigns so they can deploy deterrence, offense, emergency preparedness and emergency response.*

Recommendation 2.B.iii. *Use* ***evidence-based influence*** *methods for defense.*

Recommendation 2.B.iv. *Anticipate vulnerability at the* ***seam between domestic and foreign****. Defending it needs a coordinated, funded organizational response.*

2.C. Defense against shaping of the information terrain over years

Recommendation 2.C.i. *Build defensive advantages into the information terrain, and prevent competitors developing strategic advantage via platforms like TikTok.*

Recommendation 2.C.ii. *Encourage a thriving news ecosystem, with trusted messengers for the multiple audiences in the Joint Force and its support networks.*

3. DEMOCRATIC COMPATIBILITY: Place new capabilities within ethical, legal and political frameworks that render them compatible with a free society. Maintain Posse Comitatus and intelligence oversight, whilst also mitigating the gaps and lack of agility they entail. Restraint is not just a bug of the U.S. system, it is a strength.

Introduction

"To seduce the enemy's soldiers from their allegiance and encourage them to surrender is of special service, for an adversary is more hurt by desertion than by slaughter."

– Flavius Vegetius Renatus, c. 378 AD

"Only an alert and knowledgeable citizenry can compel the proper meshing of the huge industrial and military machinery of defense with our peaceful methods and goals, so that security and liberty may prosper together."

– President Dwight Eisenhower's Farewell Address, 1961

Put yourself in the shoes of an adversary. The Joint Force possesses great strengths, but also inevitable vulnerabilities that can be exploited by a thousand cuts over time. Sowing discord and leveraging grievances within a military or between a military and its society: these are age-old strategies in Western or in East Asian societies.

From that adversary's point of view, the millions of humans in the U.S. Joint Force and its support networks offer a vast patchwork of target audiences—a tempting smorgasbord—most of whom can now be neatly identified, characterized and reached via social media or other digital means. In 2021 Facebook ad services still kindly let (almost) anyone target U.S. military personnel (Newman, 2021). In 2021, China's AI-powered Tiktok is used by 21% of U.S. adults (similar to Twitter's 23% or Reddit's 18%) and by a staggering 48% of those aged 18-29 (Pew, 2021). A challenge to defend.

But, *over*-reactions to information threats do the adversary's job for them. In a global gray zone competition between democratic and authoritarian states, keeping democracy healthy at home is key. Extended witch-hunts or requirements for total political homogeneity are unlikely to help.

Success for the Joint Force is, then, to react effectively, but within the democratic constraints of a free society. That is:[1]

> *to deny the adversary their objectives by preserving the value of the Joint Force's human and organizational resources (or, in the event of a successful attack, recovering lost value), and to do this without damaging the health of U.S. democracy.*

Part I of this report delineates the challenge for the Joint Force in our digitizing world. Part II describes a practical, effective response to this challenge through a strategy centered on **"3 Ds": Detect, Defend,** and **Democratic compatibility.**

[1] This definition of success is adapted from (Denning, 1999), with the addition of human and democratic elements.

Part I. Who are we defending from what – and how is this new?

Who are we defending?

The U.S. currently has 1.3 million active-duty service personnel. They comprise less than one-half of one percent of the U.S. population, and their demographics are not identical to the broader U.S. population.[2] The military is, for instance, younger than the civilian population, with about 70 percent of enlisted Marines aged twenty-four years old or younger. They comprise diverse audiences in other ways: consider life as an Army Private (Helmus et al., 2018), or life as a senior cyber specialist. They are widely deployed outside the U.S., particularly in Germany, Italy, the UK, Japan and South Korea. The Army alone notes it has some "187,000 soldiers deployed worldwide in 140 countries on six continents" (Cancian, 2020).

But the Joint Force and its key support networks are far broader than this. There are some 1.02 million reserves (Congressional Research Service, 2021). In 2019 about 4.2 million personnel from the federal government and government contractors were cleared for access to secret and above information, of whom some 3.6 million were Department of Defense employees or contractors (Lopez, 2019).

Moreover as the poet John Donne wrote, "No man is an island, entire of itself", with families and friends being hugely important sources of both resilience and influences. Defending families is not a new concern for the U.S. military, for instance as attested by debates about the presence of families in South Korea (Stewart, 2020). Nearly two-thirds of privates in a recent report said a family member influenced their decision to join, and that families provide key support when problems arise (Helmus et al., 2018). In the old maxim, the "military recruits a soldier/sailor but retains a family."

Any enterprise with millions of humans (however well those humans are selected or supported), will also include a reasonable number (even if a low proportion) who will suffer from financial stresses, mental health, marital or other life problems. Life events can be part of the cause, although not the sole cause, of vulnerabilities even leading to extremist radicalization – as we know from extremism of many types amongst the U.S. population (Brown et al., 2021). They can increase risk factors and decrease protective factors to provide fertile ground for grievances that adversaries can channel.

Furthermore, the millions of humans in the Joint Force are all human and all exhibit cognitive biases and use heuristics. Heuristics like familiarity can make social media users believe false news stories, while sharing of false news stories may be largely driven by inattention (Pennycook & Rand, 2021) and novelty (Vosoughi et al., 2018). The members of the Joint Force also all vary along cognitive or personality dimensions that mean some people, for example, are more likely to perceive conspiracies (Bowes et al., 2021) or develop highly polarized political views (Rollwage et al., 2018, 2019).

[2] Adversaries might begin identifying target audiences within this 1.3 million using characteristics such as demography, race or where recruits are enlisted from. For a recent discussion of these characteristics across services see e.g. (Council on Foreign Relations, 2021).

Now consider these myriad audiences from outside, from the perspective of an adversary. They present a tempting array of potential target audiences for information operations, not least as U.S. citizens are already analyzed by companies like Facebook or by "digital brokers" who aggregate remarkably detailed data about U.S. citizens and sell it on. In addition to purchasing data or insights from U.S. sources, other sources like China's data-hoovering TikTok—used by some 48% of U.S. citizens aged 18-29 (Pew, 2021)—can help. Even if an individual in the Joint Force is not on TikTok or other social media, their siblings, parents, children or friends may be – providing wonderfully rich data for understanding that audience, finding vulnerabilities and ways to influence.

When are we defending them?

The humans in the Joint Force are not static, and as they move through different contexts—e.g. across times of the day, their careers and their lifespans—this also exposes potential vulnerabilities. Digital media magnify this challenge because now adversaries can easily target specific time periods, as we see in standard commercial targeting by U.S. big tech companies for entirely reasonable commercial purposes.[3] And, of course, these are dual-use offensive technologies.

Consider the following contexts that might expose targets for information threats.

- *On and off duty; and on personal digital devices or on work devices.* Individuals at work on a work digital device may expect monitoring and digital protection. But at home on personal devices monitoring is often (rightly) impossible and protections may be much lower – and this makes it a potentially appealing vulnerability. Similarly, many are at work on personal devices, which is still common at the time of writing.
- *During their careers in the Joint Force*. Transitions such as new postings or promotions can be points of increased vulnerability for service personnel and families.
- *Life-course: Before, during and after military careers.* Many of 2030's service personnel will be today's teenage TikTok users, today's mental or sexual health clinic users and today's risk-taking teenagers. Before their military careers, vast amounts of data will be harvested. After a military career, transitions to civilian life can be difficult, which matters as there are nineteen million U.S. veterans (Schaeffer, 2021).

Now, once more, step back and consider these vulnerabilities from the perspective of an adversary. Many vulnerabilities can never be eliminated and only mitigated – and that mitigation is crucial to help deny objectives to that adversary. To defend the humans of the Joint Force.

But vulnerabilities only matter in so far as they might be exploited.

[3] This is prominent on the Google and Facebook advertising websites, or online marketing advisers such as https://neilpatel.com/.

What are we defending them from?

In information defense[4], the goal is to preserve the value of the resources or, in the event of a successful attack, recover lost value. Value in this case is the ability of the humans in the Joint Force to coordinate and collaborate to successfully carry out their missions, which requires a level of commitment and trust in those around them and in the broader organization.

Adversaries and other destabilizing forces threaten to sow discord and disruption amongst the humans in the Joint Force, in order to degrade collective capabilities. These humans face information threats from two sources: external and internal, which could also be termed foreign and domestic. We discuss in turn and then discuss the distinctions and links between them.

External adversary information operations: State and non-state actors threaten to disrupt, degrade or distract the Joint Force. As old as warfare itself, contemporary examples include Russian operations against Ukrainian and NATO troops, described in Box 1. Consider the following four overlapping ways the external challenge can manifest.

- *"Active measures":* Semi-covert or covert intelligence operations to shape an adversary's political decisions, these were used and developed extensively by Warsaw Pact countries in the Cold War including against U.S. and allied militaries. Scholar Thomas Rid describes three key features (Rid, 2020). First, active measures are not spontaneous lies by politicians, but are the methodical output of large bureaucracies (typically intelligence agencies). Second, they all contain an element of disinformation (e.g. forged content). Third, they are always directed against an end, usually to weaken a targeted adversary (e.g. creating wedges between groups or trust in societies), although they may have a single narrow objective (e.g. against a specific weapons system, like in the 1970s/80s rousing of European opposition to the U.S. "neutron bomb").
- *"Foreign Influence Efforts":* A recent Princeton study identified and described seventy six "Foreign Influence Efforts" in which foreign governments have used social media to influence politics in a range of countries by promoting propaganda, advocating controversial viewpoints, and spreading disinformation (Martin et al., 2020). They define these as: (i) coordinated campaigns by one state to impact one or more specific aspects of politics in another state, (ii) through media channels, including social media, by (iii) producing content designed to appear indigenous to the target state.
- *"Sharp power":* Sharp power is an approach to international affairs that typically involves efforts at censorship or the use of manipulation to sap the integrity of independent institutions (Walker, 2018). Broader than just "information operations", it seeks to use the openness of Western societies

[4] Offense aims to disarm an adversary, whilst defense aims to deny them their objective. Success with information requires effective offense and defense, although this report focusses on defense.

against them and applies multiple instruments of national power, as seen with Chinese sharp power against Australia and New Zealand.

- *Long-term shaping and recessed capabilities:* Billions of smart devices are penetrating our homes globally. If the toasters, refrigerators, telephones and lights in the homes of most citizens in the U.S., its allies and partners have authoritarian surveillance capabilities baked into their design – clearly this is a challenge for all U.S. institutions including the Joint Force (Wright, 2020b). Now consider the long-term recessed capabilities that an adversary can derive from TikTok's data and analyses about millions of humans in and around the Joint Force (Boxes 3 and 4). Or consider "smart cities" being built across the world, which currently center on surveillance and will be future battlespaces. All this will shape the information terrain to the benefit of U.S. adversaries.

Extremists from within their own society: All societies and their institutions will face challenges from extremism, because there will always be extreme humans. German special forces provide a contemporary example of how this can get out of hand within military institutions, discussed in Box 2.

A recent analysis of the U.S. showed that, in 2019, some 1.5 percent of all domestic terrorist incidents were linked to active-duty and reserve personnel, and 6.4 percent were linked in 2020 (Jones et al., 2021). In human societies, and therefore their national security organs, there will always exist extreme views, conspiracy theories, misinformation and disinformation, incitements to violence and ideas from grievance entrepreneurs on the left and right of politics. The challenge for the Joint Force is how to minimize their numbers and, given the special capabilities they possess, their impacts on society and on the Joint Force's collective capabilities.

Distinctions—and links—between external and internal threats: The U.S. has long distinguished between internal and external threats. Americans have long historical traditions that abhor military involvement in civilian affairs, at least under ordinary circumstances. These find tangible expression, for example, in the 19th century Posse Comitatus Act, which forbids the Army (and other military organs) to execute civil law except where expressly authorized (Elsea, 2018). The reason for such distinctions is that a powerful military, while often needed to protect from external predation, could be used internally to control the civilian population. In 2021 the U.S. military machine and other external security organs are incredibly powerful, and if that power were directed internally it could spell disaster for U.S. democracy. Thus, the long-term health of U.S. democracy requires the clear-cut distinction between "external" and "internal."

But although the internal-external distinction is enormously beneficial, it also means that *a fundamental challenge for the U.S. will always be a "seam" between external and internal that adversaries can exploit.* A great analogy was related to me by a retired U.S. Army Colonel who originally started as a beat cop in Houston[5]:

[5] Colonel (Rtd.) Carl W. Hunt, Ph.D., private communication.

"working "across the seams," [is] something that all government and commercial enterprises struggle with. As a beat cop in Houston (Radio Patrolman), and subsequent crime analyst searching for patterns of criminal activity throughout the city and region, analyzing and working across the seams of the geographic separations of the city were most troublesome. The "seams" between the six police substations and areas of responsibility in a city the size of Houston were indeed problematic. There were criminal entities that knew both the geographic boundaries and times of shift change (another seam if you will) that existed within the Houston Police Department and they routinely sought to exploit those self-imposed separations of authority and responsibility to their benefit."

A free society can mitigate but never eliminate the vulnerability from this seam, certainly outside a total war.[6] Once a Russian narrative, for example, gets picked up by U.S. news outlets, the U.S. Government cannot employ all available instruments of power against this threat because it can now propagate as protected speech. Authoritarian adversaries—like the USSR during the Cold War—may suffer from other weaknesses relative to democracies, but they have powerful domestic organs to shut down undesired information. This particular seam poses a greater challenge for the U.S..

The Joint Force must then live with this seam, which requires recognizing and managing significant linkages across this seam:

- External and internal information threats often combine, not least as existing domestic social discord is a prime target for foreign adversaries. Soviet Cold War active measures, for instance, leveraged existing fissures (Rid, 2020). The same types of underlying grievances (e.g. social inequalities) can also fuel both internal and external threats. Moreover, both internal and external threats can harness the same tactics with information, e.g. spreading mis-information and dis-information.[7]
- Many of the key actions needed to defend the Joint Force against both external and internal information threats are the same. Part II describes these, such as mental health and social support.
- The new *Interim National Security Strategic Guidance* (INSSG, Biden, March, 2021) describes how economic globalization, pandemics and other transnational forces drive information challenges that require *responses coordinated across the internal-external seam*. How do these transnational challenges affect Middle America's middle class?
- Digitization changes the character of information threats because hugely denser inter-connections between societies blend "domestic" with "foreign" and so vastly increase the attack surface for external adversaries. Sowing

[6] Even in war, this must be time limited.

[7] Misinformation can be defined as the spreading of unintentionally false information. Examples include Internet trolls who spread unfounded conspiracy theories or web hoaxes through social media, believing them to be true. Unlike misinformation, disinformation is intentionally false. Examples include planting false news stories in the media and tampering with private and/or classified communications before their widespread release. (Congressional Research Service, 2020)

discord in Delaware used to be tricky for foreign adversaries, now they can do it from bed.

- Finally, many powerful digital technologies are built for U.S. companies to influence domestic audiences for commercial ends (to buy more beverages or a new phone) but can be harnessed by external adversaries. As the INSSG (Biden, 2021) highlights: inside and outside come together in ways that the U.S. did not well anticipate.

The next sub-section explores a tangible example of the new digital technologies—"deepfakes" that form part of a broader suite of dual-use offensive digital technologies—in order to help better anticipate the future character of information challenges.

Box 1 Russian information operations against Ukrainian and NATO troops

Shortly after fighting started in eastern Ukraine in 2014 soldiers deployed to the combat region received "fake texts" to threaten and demoralize them: "Ukrainian soldiers, they'll find your bodies when the snow melts" or "Nobody needs your kids to become orphans". Other text messages aimed to undermine unit cohesion and morale. Texts, which often appeared to come from fellow soldiers, claimed the commander had deserted and that "We should run away." Text messages sent to one's phone are much harder to ignore than leaflets or radio messages.

Russia also combines information and kinetic operations, as the following example described. A text message to a soldier first tells him he is "surrounded and abandoned." Ten minutes later, his family receives (via his recent contacts) a text message stating, "Your son is killed in action." Family and friends then likely call him to see if the news is true. Seventeen minutes after the initial text message, he receives another message telling him to "retreat and live", and then shortly after that an artillery strike follows to the location where the large group of targeted cell phones were detected. This blurs the geographical boundaries between the front line and the home front.

NATO troops deployed in the Baltics and Poland to deter Russia have also been targeted. This includes having their Facebook accounts hacked, having data erased, or receiving a message stating "Someone is trying to access your iPhone" that includes a map with Moscow at its center. This may intimidate soldiers, let them know that Russian intelligence forces are tracking them and that their data is at risk.

Texts that falsely announce infidelity and injuries are sent to NATO soldiers' loved ones back home, as recently described by Commander Michael Widmann of NATO's Co-operative Cyber Defence Centre of Excellence based in Tallinn, Estonia (The Economist, 2021). "It throws you off," he said. When in April 2021 he led the world's biggest military cyber-exercise, NATO's Locked Shields 2021, it included the hacking of participants' mobiles.

Russia has also targeted local support networks for the U.S. military in Europe, attempting to decrease its military readiness and that of its NATO allies. Russian media outlets have, for instance, reached out to mayors of towns outside of the Hohenfels training area in Germany, inquiring about military training noise disrupting the local population.

In 2020, a Canadian-led NATO battle group in Latvia was targeted by a pandemic-related disinformation campaign ahead of a major exercise, which commanders said they believe originated in Russia (Brewster, 2020). Reports circulated in Baltic and Eastern European media outlets falsely suggesting that the contingent at Camp Adazi in Kadaga, outside the capital of Riga, had "a high number" of Covid-19 cases.

This Box draws on (Beehner et al., 2018) except where otherwise referenced.

Box 2 Germany: Neo-Nazi infiltration of Special Forces (past decade-now)

In June 2020 the German Defense Minister, Annegret Kramp-Karrenbauer ordered the partial dissolution of the elite “Kommando Spezialkräfte” (KSK) (BBC, 2020). Twenty members of the elite force were suspected of right-wing extremism. Ms Kramp-Karrenbauer told the Süddeutsche Zeitung newspaper that the KSK had "become partially independent" from the chain of command, with a "toxic leadership culture", and that some 48,000 rounds of ammunition and 62kg (137lb) of explosives had disappeared. Before this, in May 2020, police had already seized explosives and weapons at the home of a KSK soldier.

The military's broader problem with far-right supporters emerged in 2017 with a Bundeswehr officer named as “Franco A.”. The soldier was first detained by Austrian police in February after he tried to retrieve a handgun he had hidden in a toilet at Vienna airport (BBC, 2017a). His “right-wing extremist mindset” was first flagged by a supervisor after he had submitted his master’s thesis at the French military academy Saint-Cyr in December 2013 – but he continued to serve, during which he reportedly lived a double life as a Syrian refugee and planned political assassinations before his 2017 arrest (Denney, 2021).

Following his arrest, inspections were ordered on all military barracks when Nazi-era memorabilia was found at two of them (BBC, 2017b). In January 2020, military intelligence said there had been almost 600 suspected far-right supporters in the army over the past year (BBC, 2020). Many of those suspected of far-right links are thought to be sympathetic to Germany's main opposition Alternative für Deutschland (AfD) party (Denney, 2021).

In part this concentration of extremism within the AfD reflects the relatively limited formal political opposition to the Government compared to a country like the U.S. or UK – for most of Chancellor Angela Merkel’s fifteen years in office her center-right party governed *with* the main center-left party. That left little effective choice at the ballot because essentially whoever one voted for, one got the government.

In 2019, the AfD estimated six percent of its party members were career soldiers. AfD members of the Bundestag (Parliament) include more former career and regular soldiers than any other faction, prompting the German tabloid Bild to ask in a headline: “Will Alternative for Germany become the new soldier party?” The AfD was itself recently placed under formal surveillance by Germany’s domestic intelligence agency for suspected right-wing extremism, and it argues against the defense ministry’s methods to address far-right ideology (Angelos, 2021).

How is this new? Deepfakes as an example of evolving challenges

Deepfakes help illustrate how the character of information threats may evolve with technology, particularly AI and the digital technologies. Deepfakes can be defined as AI-generated synthetic media (e.g. images, video or audio) that most commonly involve a person saying or doing something that they did not say or do.[8]

Deception and forgery are old. Entire books have been fabricated. The infamous *"Protocols of the Elders of Zion"* first appeared in 1903 and was largely copied from an obscure, French-language political satire (Zipperstein, 2020).

What is new here is largely that new technology makes powerful tools of fakery available much more cheaply, rapidly, easily and widely. Deepfakes arose chiefly from *dual-use* technology, as a by-product of AI advances and civilian uses of synthetic images for entertainment. Indeed, deepfake-like technologies will likely become in widespread use globally for synthetic personal assistants, retail assistants and in healthcare applications like aiding those whose disability affects their speech.

The specific AI advances that made deepfakes possible occurred around 2012, in an AI technique called "deep learning." Deep learning[9] radically improved AI's ability to perceive things such as images, audio or video (for accessible discussions see (Wright, 2019b)). Deep learning AI is now very good at *perceiving* images or sounds – and essentially turning those computer programs back-to-front instead *generates* images or sounds. This makes convincing "deepfakes", which emerged around 2018 to make fake pornography.

Deepfakes provide adversaries with new openings for mis- and dis-information in three main ways:

- *Unexpectedness*: Currently, as individuals and collectives we are poorly prepared for realistic fake videos, pictures or audio – and so when deepfakes are used creatively the surprise they cause can help catch our attention (Wright, 2019a) or even, sometimes, fool us into believing they may be real. Importantly, this will likely fade as we become used to deepfakes. Emails or "CGI" in movies seemed new once.
- *Mass-produced disinformation:* Deepfakes are now easily mass produced and "broadcast." This might be done in a cheap and dirty way, for example by enterprises like the "troll farms" of Russia's "Internet Research Agency" or the Macedonian entrepreneurs who spewed out fake posts for U.S. social media audiences (Singer & Brooking, 2018). More sophisticated and

[8] No universally accepted definition exists. A good recent report by scholar Tim Hwang provides the following (Hwang, 2020): "In this paper, the term "deepfakes" refers to the broad scope of synthetic images, video, and audio generated through recent breakthroughs in the field of ML [machine learning, a form of AI], specifically in deep learning. This term is inclusive of ML techniques that seek to modify some aspect of an existing piece of media, or to generate entirely new content. While this paper emphasizes advances in neural networks, its analysis is relevant for other methods in the broader field of ML. The term "deepfakes" excludes the wide range of techniques for manipulating media without the use of ML, including many existing tools for "cutting and pasting" objects from one image to another."

[9] Many computational methods for AI are from a field called **machine learning** which is the set of techniques and tools that allow computers to 'think' by creating mathematical algorithms based on accumulated data. **Deep learning** is one method for machine learning that uses neural networks with at least one "hidden layer." For an accessible discussion see Chapter 1 (Wright, 2019b), www.intelligentbiology.co.uk.

potentially much higher impact ways to use broadcast information are described in Box 3.

- *More exquisite "active measures":* High quality, carefully crafted deepfakes also have uses. Reportedly, for example, a deepfake of the voice of a company boss successfully fooled a subordinate into transferring a large sum of money, a financial crime that required detailed knowledge of the company (Bateman, 2020). In another potential use, high quality deepfakes might be hidden amongst troves of genuine stolen media and be leaked to the media – a trick successfully used with other media during and after the Cold War (Rid, 2020).

But deepfakes are also quite limited, particularly if one remembers the key fact that *the main aim of mis-/dis-information is to create effects in audiences*. Limitations include:

- Quality *or* quantity? Highly convincing deepfakes still require a lot of computational power and, more importantly, data about the people to be faked. Videos of Tom Cruise can be faked so well in large part because we have very many videos of Tom Cruise. Short of a new technological leap (e.g. in the ability to generalize learning from small amounts of data) we are unlikely to see mass-produced and highly convincing deepfakes soon.
- Detection mechanisms to catch deepfakes are good and can catch up quickly with advances in deepfake manufacturing (Hwang, 2020). For this reason adversaries may keep back their novel deepfake tech to act as a "Zero-day" exploit, which is a vulnerability previously unidentified by the defender so that they have zero days of notice to fix it before damage is done. To help mitigate such novelty, defenders should set up a "zoo" to share deepfakes (Hwang, 2020).
- Who is the *creative talent*? Most importantly, a convincing picture or video requires ideas and points of leverage about target audiences – and creative, talented people to create effects. Will it be funny, or shocking, or believable enough?

Thus, deepfakes used alone will likely only exert limited influence.

Instead, deepfakes will most likely be useful as one tool in **"combined arms" information operations** to create effects in audiences, much like the German *Panzer* forces combined infantry, tanks and artillery to devastating effect. Consider some other tools amongst which deepfakes can sit.

- ***Cropping real media***, or mislabeling media with a ***fake context***, can be as effective as fancy AI to create effects in audiences (see Fig. 1 below). Combining deepfakes with other types of fakery can help keep things fresh for audiences and create problems for defensive content moderators (human and/or AI) – particularly if "ironic" or "funny" versions are used to push the boundaries of what is allowable.
- ***Conversational systems*** can drive realistic fake bot identities on social media, which can be given plausible "faces" by deepfakes. ***Social bots*** are algorithmic software programs designed to interact with, or send information to, humans. Bots powerfully amplify commercial messages (Confessore et al., 2018). Again they have political uses. Bots published

perhaps a fifth of all 2016 U.S. presidential election and a third of all Brexit referendum tweets. They may have spread propaganda in 50 countries (S. Bradshaw & Howard, 2019). Most bots are not yet powered by sophisticated AI, although they are becoming available (Woolley, 2020) to further semi-automate campaigns. Conversational "chatbots" or AI personas form another huge commercial and research area (Vogel & Wright, 2019).

- ***Microtargeting*** is a form of online targeted advertising that analyses personal data—a role AI can play—to identify a specific audience or individual's interests in order to influence their actions (ICO, 2019; Tiku, 2017). Facebook's original social network produced the data that afforded commercial microtargeting (Naughton, 2019) – and then that microtargeting apparatus afforded political use. Cambridge Analytica's abuses weren't certain to happen, but the platforms currently afford offensive political use.

Much of this offensive tech is dual-use, which raises the question: why is so much powerful offensive tech available? Answering this question brings to light the real tech threat to the Joint Force in information operations: a vast market failure where huge commercial spending builds profitable dual-use *offensive* tech (like micro-targeting) but little for *defense* that's mostly just a business cost.

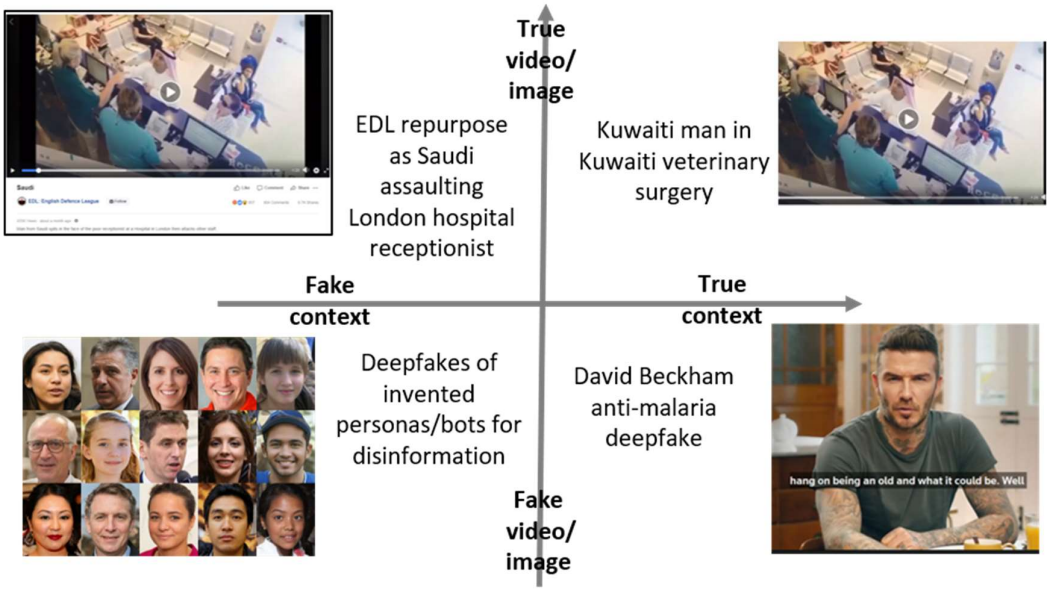


Figure 1 Deepfakes and context in disinformation campaigns. A true video of an angry Kuwaiti man (top right) was repurposed on social media by the far-right "English Defence League" in a fake context (top left). Deepfakes can enable footballer David Beckham to promote an anti-malaria campaign across many languages (bottom right), or malign actors can create fake faces for fake personas (bottom left). [10]

[10] The top two images are taken from (Banet & Lebel, 2018); the bottom left images are fictitious faces used for illustration here but not created for fake personas/bots, taken from (Leskin, 2019); and the bottom right image is taken from (Reuters, 2019).

How will new defenses be built? Market failures and defense against dual-use offensive tech

The biggest cause of the challenge from AI and the new digital technologies is a profound **market failure**. Colossal efforts are made to develop advertising and marketing tools—understandably given global ad spending on social media alone of some $84 billion in 2019 (McCarthy, 2019)—but who is incentivized to build tools that *defend* populations, including in the Joint Force, from influence or manipulation?

Defending people is often just a business cost. Facebook, for instance, took a big hit as these costs emerged. The Financial Times described how “Investors were particularly spooked in July 2018 by warnings from the company itself about the huge financial costs of tackling problems such as disinformation, data protection and other online abuses (T. Bradshaw, 2020). For investors potential costs are investment risks. This largely explains why Facebook concentrated on cheaper AI-heavy responses (Murgia & Murphy, 2019) even though, as Box 4 describes, many expensive humans are also needed.

Moreover, human-AI teams are not just needed to protect the Joint Force itself, but also local support networks in key regions of global contest. Facebook anticipates most user growth in the Asia-Pacific, but fails globally to transfer moderation manuals—let alone providing required local adaptation—and struggles with translation across its dozens of supported languages (Fick & Dave, 2019).

It is unclear whether *technologically* AI will benefit information offense or information defense more – but clearly the *market* incentives strongly favor offense over defense.

Box 3 Mass-personalization of influence – the Joint Force in an escalation scenario

Mass personalization of retail has already been rolled out at scale by companies like Amazon, where recommendations based on big data contribute to vast profits. Mass personalization of healthcare that tailors treatments to individual patients, rather than big groups, is clearly on the way (Kent et al., 2018). Facebook can be considered a company that sells the ability to influence humans in a highly precise and targeted ways but at mass scale (Véliz, 2020). Meanwhile China's tech titans like Alibaba, with a market capitalization of over 550 billion USD as of May 2021, are no slouches at using big data for mass personalization (Laubscher, 2019). Why would mass personalization not also be applied to information operations in security?

Personal data on the members of the Joint Force will be a key fuel, such as from medical data, TikTok use, financial data or romantic dating sites. Not just about the members of the Joint Force themselves, data about family can be very helpful: that fun genomic data bought as a birthday present may reveal that somebody's father is not who they think he is; or a partner had an affair.

So many options for use. Firstly the data can train AI, with some human help, to find tempting target audiences in the Joint Force. Secondly, weaponized personal data on millions in the Joint Force or their families can be injected into their social networks.

Leaking damaging data on members of the Joint Force at key moments, such as during a China-U.S. escalation scenario or limited war, might have some utility. And whilst embarrassing details are coming out, what about slipping in some invented damaging data where none exists? It could be targeted at individuals' social media. Or perhaps troves of analyzed data could be released by "free speech" third parties like a new Wikileaks. Or simply an Ashley Madison style data dump. Financial problems, gambling habits, computer pornography, sexual health treatments, the results of drug or alcohol tests, sexual experimentation…

It could never happen, one might say. Soviet Cold War information operations included publishing "Who's who in CIA" listing agents and others incorrectly labelled as agents, to which the U.S. responded by publishing their own list of KGB agents (Rid, 2020). In a China-U.S. escalation scenario or limited war, would the U.S. hold off attacking parts of the digital authoritarian apparatus, such as the "Social Credit System", by which the Chinese Communist Party (CCP) increasingly maintains its authority? That might be perceived as a threat to regime security (the CCP's top priority) and thus perhaps more escalatory than leaking personal data about humans in the Joint Force.

New human-AI teams and organization are needed at the scale of this defensive challenge.

Box 4. AI – strengths, weaknesses and TikTok

The AI-related technologies comprise the cutting edge of the broader digital technologies. By the term "AI" here I refer to a constellation of AI-related technologies that together provide powerful, wide-ranging and new capabilities: AI more tightly defined, machine learning, big data, and digital things (e.g. the "internet of things"). Together they enable a new industrial revolution, taking the vast reams of data produced by the computers and internet – and turning it into useful information. None is entirely new, but recent big improvements (particularly from "deep learning" around 2012) mean together they have revolutionary applications.

However, these advances have not been uniform, and we must understand two key strengths and two key limitations. **AI is currently good at two things:**

(1) **Perception**, e.g. perceiving images or speech, or some patterns in big data.

(2) AI also improved when choosing actions in **tasks that are bounded enough** to be very well described by vast amounts of (often labelled) data, e.g. logistics in a warehouse.

Thus, real-world impacts now relate largely to perception (e.g. perceiving faces or speech) or some bounded decision tasks (e.g. logistics). Continued rollout in these areas will likely dominate for the next few years at least.

But **AI's two key current limitations**, have meant that rolling AI out in the real world, let alone at scale, has proven very tough in many fields (e.g. medicine despite all the hype). These are:

(1) AI deals badly with **context**, so humans are often needed to make even common-sense judgements.

(2) AI requires **huge amounts of often labelled data**, so that setting up datasets is often a crucial precondition.

Thus, because of AI's current limitations it requires extensive human involvement to help deal with context (i.e. **human-machine teams** rather than AI alone); and current efforts will likely try to **acquire large amounts of data** that includes both **broad data** (e.g. via **TikTok**) as well as **"ground truth" data** to act as labels (e.g. medical or financial data via U.S. tech companies, data brokers or espionage).

TikTok data alone is valuable, with 689 million active users outside China (Sehl, 2021) and a further 550 million users at its twin platform inside China called Douyin (Liao, 2021). TikTok's high quality AI algorithms are now being sold to help other companies personalise their websites and apps (T. Bradshaw, 2021). Together this is powering the parent company, Bytedance, to a valuation of some $250 dollars (Chen et al., 2021). And TikTok is shaping the terrain in ways opaque to the U.S.: Western young people write newspaper articles thanking TikTok for diagnosing their mental health disorders (Bosely, 2021).

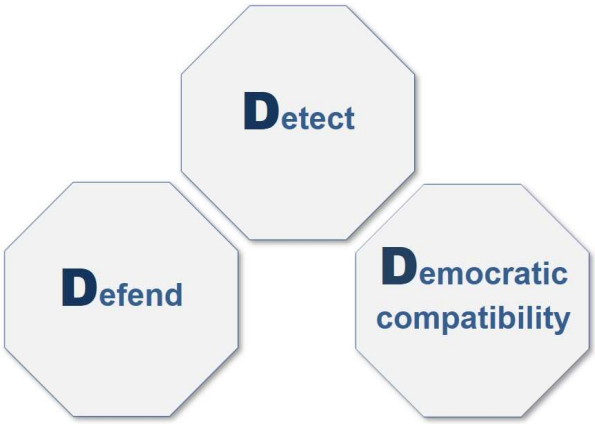
But this does not even count the additional value for information operations by combining the broad TikTok data with "ground truth" data on U.S. citizens obtained legally and via espionage. Given that 48% of those aged 18-29 use TikTok (Pew, 2021), this will open valuable new lines to exploit within the Joint Force.

Part II. Interventions – what can the Joint Force do?

What does success look like against these evolving information threats? Across the spectrum of competition—from peace through the gray zone to war—the basic aim is the same:

> *to deny the adversary their objectives by preserving the value of the Joint Force's human and organizational resources (or, in the event of a successful attack, recovering lost value), and to do this without damaging the health of U.S. democracy.*

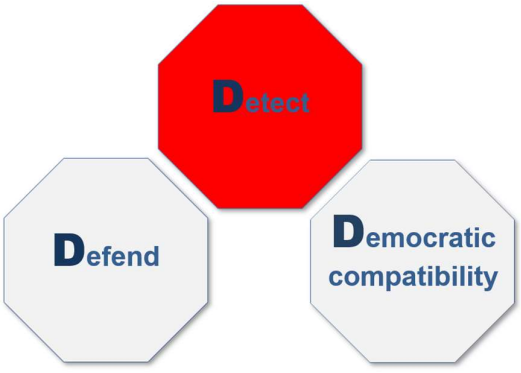
This can be achieved through a strategy centered on **"3 Ds": Detect, Defend,** and **Democratic compatibility**. All three are necessary, and none alone is sufficient. Part II discusses each in turn.



But before that discussion it is important to note that although this report focuses mostly on gray zone competition—the current situation with China and Russia, and likely to continue over the next few decades—there is a small but real chance of escalation to war. Thus, strategy for countering information threats must also anticipate how to apply the "3 Ds" during escalation and war.

Moreover, one of the most often ignored contingencies is that great power war, even in our nuclear age, is perfectly likely to continue for many years – as they have so often before despite initially optimistic forecasts. Again, although beyond this report's scope, some thought must anticipate how the "3 Ds" apply in a prolonged great power war.

1. DETECT



Without effective capabilities to detect adversary influence operations, the Joint Force will be blind. It must have capabilities to detect and characterize adversary influence operations against the Joint Force, in order to grasp who is targeted, by what means and for what purposes.

New human-AI teams will be needed—AI alone is currently very far from able to cope (Box 4)—as will new will new organizational capabilities. Detection

must also operate at multiple scales that we discuss in turn, which includes: (1.A.) coordinated campaigns; (1.B.) specific instances of mis-/dis-information; and (1.C.) how adversaries may shape the information terrain over years.

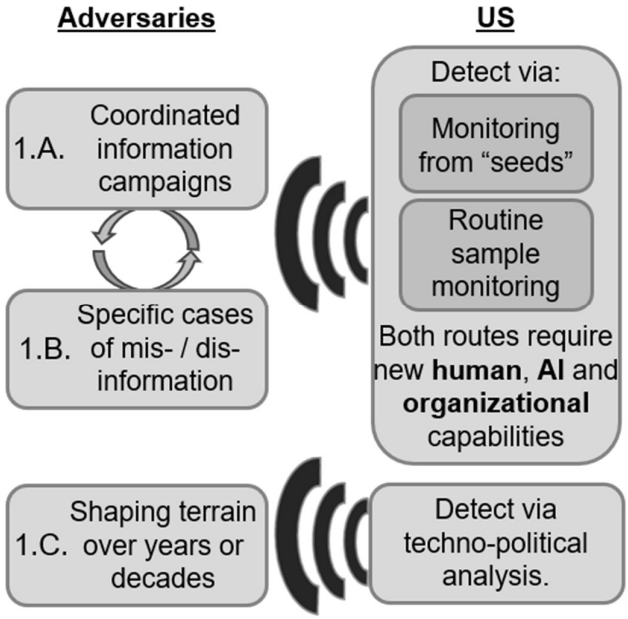


Figure 2. Detection at multiple scales.

1.A. Detecting coordinated campaigns

External adversaries like China or Russia can mount large, coordinated campaigns employing "combined arms" information operations. Capabilities adequate to detect such external threats will also be sufficient to detect domestic threats. Characterizing coordinated adversarial campaigns will involve detecting such aspects as their:

- Multiple "Lines of Effort", each of which may use particular techniques to achieve particular ends (e.g. sowing discord between racial or political groups in the Joint Force to provoke demonstrations or reduce morale);
- Activities across multiple platforms (e.g. Twitter, Facebook, WhatsApp);
- Exquisite active measures "narrowcast" against specific targets (e.g. the painstaking planting of fakes amongst stolen documents);
- Mass-produced information operations "broadcast" to wider audiences in and around the Joint Force; and
- Activities that exploit the internal/external "seam" faced by the Joint Force, e.g. as external adversaries use servers based in the U.S. to conduct their information operations.

Where these campaigns are primarily conducted via digital media—as they almost certainly will be now—then detecting them is essentially a challenge of ***counter-intelligence at scale***. This requires human, AI and organizational capabilities.

In terms of organizational capabilities, it requires systems that can coordinate and integrate across multiple sources of intelligence (not only digital, and derived from sources including the Joint Force, intelligence community and cyber sources) in order

to characterize digital social networks at the required scale. The *scale of relevance* depends on the audiences targeted within the Joint force, which may be dozens, thousands or even millions of humans.

Detection through counter-intelligence at scale may occur via two complementary pathways, both of which should be built by the Joint Force:

Firstly, routine monitoring of social media/online interactions of a sample of those within the Joint Force and its support networks who have already submitted themselves for enhanced monitoring, in order to detect adversary activities that can then be investigated more closely.

> Is a Russian or Chinese campaign, for example, targeting the Joint Force or its support networks? To help answer this question, human-AI teams could monitor a small sample of individuals with security clearances for limited periods of time. For security clearances, publicly available social media data can already form part of the vetting process, and it is anticipated that moves towards "continuous evaluation" may also integrate social media data.[11] The intrusiveness of such routine monitoring can be reduced by focusing where possible on actors who communicate *at* these individuals, particularly actors from outside the U.S.[12], rather than on what these individuals in the Joint Force themselves do. More broadly, such routine monitoring should be conducted with appropriate oversight and to the minimum degree needed to detect and characterize adversarial activities while maintaining the trust of members of the Joint Force.

Secondly, build wider (but still carefully limited) digital investigations of social media/online activity by groups in the Joint Force based upon investigative "seeds" derived from other legitimate sources.

> For example, individuals identified as compromised by other sources (as seen in the German military in Box 2) could be used as start points from which to investigate the social media/online activities of others with whom they interacted. "Honeypots" could be used to entice malign actors. Other potential starting seeds may involve interactions between members of the Joint Force and adversary influence networks that have been identified by other means (e.g. via the mapping of adversary networks of social bots, or via intelligence from other sources).

Importantly, both pathways should *detect at the speed of relevance*, so that they can take actions rapidly enough through private sector social media/online networks to dampen the spread of mis-/dis-information (e.g. via blocking or removal).[13] Such public-private links will require appropriate oversight.

---

[11] For existing use see e.g. www.dni.gov/files/NCSC/documents/products/CE_FAQ_7_July_2020.pdf. Various public reports over a number of years have discussed social media in continuous monitoring see e.g. (Ogrysko, 2019; Stevenson, 2016). A recent Rand report also noted that digital personal conduct is seen as an emerging risk for younger generations of potential clearance holders (Posard et al., 2021).

[12] This helps protect the privacy of those who are *not* in the Joint Force. Data from verifiable U.S. citizens who are *not* in the Joint Force and who communicate *at* the Joint Force must also be treated carefully to protect their privacy, e.g. by ensuring deletion within a specific period of time.

[13] Timing matters in many ways when combating misinformation. When a debunking message is shown relative to a misinforming headline, for example, affects its impact (Brashier et al., 2021).

Finally, providing sufficient experience to adequately *train* the human, AI and organizational components of this counter-intelligence at scale may be helped by working outside the Joint Force – e.g. with allies and partners who are subject to more active adversarial information operations (e.g. the Ukraine, Baltic states, Taiwan or Middle Eastern states).

Recommendation 1.A.i. for the Joint Force: *Build the integrated human, AI and organizational capabilities for* ***counter-intelligence at scale****, which can detect adversary information operations at the scales of relevance (e.g. dozens of individuals or hundreds of thousands of individuals).*

DARPA has funded programs to build AI that can address parts of this problem (e.g. "INCAS" or "COMPASS"), even though DARPA's entire budget is very small compared to the big tech companies who make dual-use offensive tech. Moreover, tech must not overshadow the organizational systems needed to integrate "seeds" and routine monitoring that are just as crucial for success.

Recommendation 1.A.ii. for the Joint Force: *The* ***"seam" between "domestic" and "foreign"*** *should remain a very clear distinction but it also presents a challenge for detecting threats that straddle that seam. Clear responsibilities should exist for how organizations coordinate to detect such threats, alongside dedicated resources to facilitate coordination.*

Threats that straddle the internal-external seam should not become "some other organization's problem" – which would suit adversaries very well.

Rising to the INSSG's challenge of bringing together domestic and foreign policy (Biden, 2021) raises tough legal, ethical and bureaucratic problems in our rapidly digitizing world. Meeting them requires U.S. domestic and foreign facing organizations to collaborate in ways that are democratically compatible.

1.B. Detecting specific instances of mis-/dis-information

Below the scale of detecting coordinated campaigns, it is also crucial to build capabilities to detect specific instances of misinformation, such as a particular fake story or specific deepfake (Fig. 1). These are important both to reduce the noise of untrustworthy information in the information environment and also to feed into the mechanisms for detecting coordinated campaigns (Fig. 2).

Again, AI alone is insufficient and the required capabilities include both human-AI teams and organizational innovations. Consider the example of deepfakes "broadcast" as part of a mass information operation, for which counter-intelligence at scale must employ:

- technology for deepfake detection;
- trained humans who can add contextual understanding that helps defeat adversary "combined arms" techniques to avoid detection (e.g. use of "irony");
- organizations like a "deepfake zoo" to share deepfakes so that many deepfake detectors can learn, as well as organizational links that can

distribute knowledge about specific deepfakes at the speed of relevance so social media platforms can stop them being uploaded or shared.

Human, AI and organisational innovation will also be needed to combat deepfakes used in the more exquisite "active measures" described in Part I.[14]

Recommendation 1.B.i. for the Joint Force: *Build the integrated human, AI and organizational capabilities to detect specific instances of mis-/dis-information (e.g. deepfakes).*

Data can be acquired through the tightly limited means described above for detecting coordinated campaigns (1.A.). Technology and processes can be acquired from private sector platforms with experience moderating content, but should be implemented with greater care and oversight – essentially by an accountable cyber military police or counter-intelligence at scale.

1.C. Detecting how adversaries shape the terrain over years

Finally, on a timescale of years or decades, the Joint Force must also look ahead to detect longer-term strategic threats: adversaries are reshaping the globe's information infrastructure in ways that shift strategic advantage away from the Joint Force and its humans.

The U.S. benefits enormously from having shaped the global information terrain in which the Joint Force's humans live and work – through U.S. global tech giants, its position at the center of global communication networks (particularly as part of the "Five Eyes" intelligence-sharing apparatus) and global financial networks (e.g. the "SWIFT" banking system).

But now China's heft enables it to shape the global information terrain more in its favor: from global AI standard setting, to the global social media platform TikTok, to building global undersea and outer space communication networks. The humans in the Joint Force will be ever more immersed in digital technologies, and the U.S. must be able to detect the risks posed as China increasingly shapes this terrain.

Consider the Chinese social media company TikTok (Box 4). It is hugely popular with young Americans. Tiktok owns and shapes the terrain on its platform: not only the *algorithms* that determine what people see and what becomes popular; but also how millions of users are split into *target audiences*, how insights are derived about those target audiences, and how effective those insights are for driving *influence*. What can the U.S. reliably detect on such terrain? And, of course, TikTok is harvesting vast amounts of *data* to build up profiles of many of the humans who will populate the present and future Joint Force.

It is unclear how the Joint Force can be adequately defended whilst crucial social media are owned or operated by nations beyond trusted allies and partners.

[14] For two good reports specifically on deepfakes, rather than the broader range of challenges addressed here, see (Bateman, 2020; Hwang, 2020).

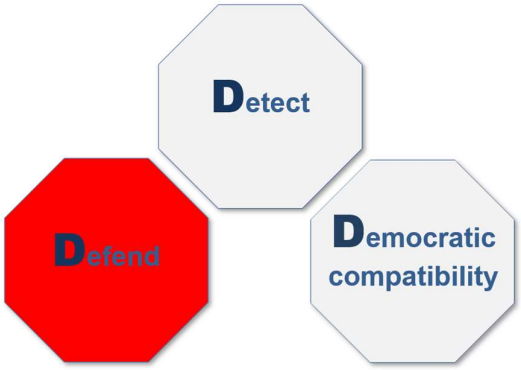
Recommendation 1.C.i. for the Joint Force: *Identify long-term risks to the Joint Force's humans as other nations, beyond trusted allies, shape the information tech immersing their lives (e.g. China's* ***TikTok*** *controls algorithms and big data on half of U.S. youths).*

Who will build detection capabilities across these scales?

Finally, the Joint Force should recognize that "the market" on its own will neither build and deploy all the capabilities needed to detect adversary information operations, nor surmount the challenge posed by a hugely popular Chinese owned or embedded entity like TikTok. However, requirements for the security of the Joint Force are seen as important in U.S. politics, and other interest groups, such as those for copyright theft or children's rights[15], have successfully used clear and forceful arguments to force changes from big tech.

Thus, the Joint Force must help correct the market failures by specifying the capabilities needed, arguing for them in appropriate fora and where necessary funding their development.

2. DEFEND



Individuals' cognition will always contain vulnerabilities as targets for disruption, which the Joint Force can help minimize and so deny to others. No panacea can exist. Instead, minimizing these vulnerabilities requires ongoing improvements to defense at multiple scales: the individual human; coordinated campaigns; and shaping the information environment over years or decades. We discuss each in turn.

2.A. Defending at the individual human scale

The Joint Force can help defend an individual's vulnerabilities, enhance their resilience and give them the technological tools to defend themselves online.

Recommendation 2.A.i. for the Joint Force: *Enhance* ***social, family and mental support*** *with a particular focus on predictable periods of vulnerability (e.g. moving postings).*

Helping individuals effectively when they have problems renders them harder targets for grievance merchants both foreign and domestic. The Joint Force should strive for continuous, evidence-based improvement to provision of pastoral care, mental health services, family support and training for future

[15] To protect copyrighted content on Youtube, Google has since 2007 applied a system (now called Content ID) that compares uploaded videos to audio and video files registered by content owners. By 2018 they claimed to have invested $100 million in Content ID and paid over $3 billion to rightsholders (Sawers, 2018). Other companies have similar systems (Constine, 2020). While far from perfect in enforcement, children have various protections online in the US (Keller & Dance, 2019; Kelly & Alexander, 2019).

employment. Predictable periods of enhanced vulnerability should be anticipated, for example providing additional social support for military families as they go through transitions like new postings. None of this is "sexy" policy, and it is no panacea, but it can help reduce vulnerabilities in potentially attractive target audiences for adversaries.

The Joint Force can help harness new digital tools. These are effective and can, for example, help increase access and reduce costs (Fu et al., 2020; Karyotaki et al., 2021; Rosen et al., 2021). However, enhanced cyber security for this is key.

Recommendation 2.A.ii. for the Joint Force: *Give individuals the technological* ***means to defend themselves*** *online. Low cost, practical options exist.*

"Choice Shield" is one example of an early-stage project that emerged from interactions among academics from the cognitive and communication sciences, civil society, and technologists.[16] It aims to afford people the ability to choose what they see on social media—without censorship—using an app or browser extension. Users can decide how many manipulated images they want to see, and select which organizations will rate those images as manipulated (e.g. they could choose a *CNN*-approved or a *Fox News*-approved filter). The code and tools will be open source. Conceptual prototypes piloted in two 2019 studies showed considerable appetite among users for control over what they saw on Facebook. Tools that give users more tailored control already exist in a more limited version[17], and revealingly there was some legal pushback from Facebook before the 2016 U.S. election changed the political climate (Blue, 2013).

Another more limited but later stage example is "NewsGuard", which uses a team of journalists to rate news sources for reliability (e.g. the *New York Times* or *Fox News*) and should be free to all personnel in the Joint Force.[18] Unlike Choice Shield individuals cannot choose which filter provider is used to screen information (e.g. some may want a *CNN* filter, others a *Fox News* filter) and there would not be the ability to plug in a filter to remove deepfakes etc.

Influential scholars such as Francis Fukuyama have more recently advocated such approaches, which he describes as "middleware." Middleware is generally defined as software that rides on top of an existing platform and can modify the presentation of underlying data (Fukuyama et al., 2021).

Clearly such tools will mostly appeal to the already "news savvy" unless it can be bundled with browsers, apps or similar (Molla, 2019). But they can provide one more useful improvement in a challenge for which no silver bullet can exist.

Moreover, increasing demand for better information, not just limiting supply, is also important. Considerable demand exists for disinformation, with fake

[16] My collaborators were: Karen Dill-Shackleford (Fielding U.), Aurie Babarinsa (Carnegie Mellon U., formerly at Twitch) Jevin West (U. of Washington,) Don Grant (Resolutions Teen Center) and Richard Petty (Ohio State U.). Contact nick@intelligentbiology.co.uk for data or further details.
[17] See for example https://socialfixer.com/index.html and https://www.media.mit.edu/projects/gobo/overview/.
[18] Personal communication. To view NewsGuard see www.newsguardtech.com/.

news traveling further and faster than true news (Vosoughi et al., 2018). Thus, if such tools were adopted the Joint Force should also seek to increase the propensity for people to use tools like these as part of broader training outlined next.

Recommendation 2.A.iii. for the Joint Force: *Training and practical help concerning social media security (e.g. use high privacy settings and do not use apps like TikTok), the clear expectations set by the Joint Force (e.g. on privacy settings, or legal requirements for political posts) and practical ways to evaluate social media content.*

Telling people that they are "illiterate" about media is unlikely to provide a good route to help them better defend themselves. Nor will telling people they need "education" in how to think properly. But in a changing world everyone needs to learn new things, and the Joint Force can help provide training and an environment that helps individuals better defend themselves – and do so in ways that aren't time consuming or (too) boring.

Designers of training programs should put themselves in the shoes of the audience—in this case the humans in the Joint Force—and find ways to offer things that the *audience* values. This will also be a slow process, so should be repeated over the person's career.

Training can address simple things like how to ensure that privacy settings are set as high as possible, what apps might be risky or what the rules (e.g. the Uniform Code of Military Justice) actually mean for social media. Training in more formal settings can be aided by friendly help via outreach at places like local shops or other communal areas about how to actually change the settings on phones or other devices (which big tech makes deliberately very hard to understand).

Clear expectations from the Joint Force must be established in the minds of the individuals in the Joint Force.[19] For example it should be clear that high privacy settings are expected, so it will be frowned upon to do otherwise. Similarly, clear codes for political content exist on posting political content (and that also specifically encourage political engagement)[20], and it should be clearly understood that breaking these existing rules can and will result in disciplinary action up to and including legal sanctions.[21] The point is to change social norms in the Joint Force.

Engaging documentaries or other media explaining adversaries' use of social media—e.g. by China in Taiwan or Russia in its near abroad—and the broader social media business models may also be helpful for some.

Training in how to evaluate social media content can also be given – although crucially neither touted as "media literacy education" nor seeming to imply "education" to change their political views. Instead, the—actually true—

[19] This draws on evidence for the power of "norms" to drive behavioural and cultural change. For a review of this evidence see Wright ND *"From Control to Influence: Cognition in the Grey Zone"* Version 3 with updates, 2019 is available at www.intelligentbiology.co.uk/s/From-Control-to-Influence

[20] E.g. www.army.mil/socialmedia/

[21] A recent example was a TikTok video by two deployed Michigan Army National Guard soldiers (Winkie, 2020)

stated purpose should be that adversaries aim to sow discord and degrade the Joint Force's capabilities via information operations played out over years, and some simple techniques[22] can help people better defend themselves against that manipulation, and help them be less likely to inadvertently help adversaries manipulate others in their communities.

Finally, given the recent high profile of "media literacy" we must also note that whether "media literacy" interventions actually make an impact is poorly understood, and even in the best case scenario are unlikely to be a magic bullet. As a recent RAND review notes "there is little causal, evaluative research in the ML [media literacy] field that isolates the effects of ML interventions" (Huguet et al., 2019). Fashionable as "education" currently is, other aspects of training and practical help may be more helpful and should also be employed and evaluated.

2.B. Defending individuals at the organizational scale

Individuals in the Joint Force cannot defend themselves alone. They must entrust their information to others, such as medical facilities, banks or personnel departments, who must guard it. They are embedded in social networks, organizations and families and cultures that influence their behavior. And the individuals in the Joint Force face threats from capable state and non-state actors.

Recommendation 2.B.i. for the Joint Force: *Defend the data of the individuals in the Joint Force and its support networks – and, contrary to the conventional wisdom,* ***build silos*** *for their data.*

AI is as good as the data it trains on – and data about the humans in the Joint Force is enormously valuable because adversary campaigns require data to understand and thus influence target audiences in the Joint Force. As Box 4 describes, the integration of data is critical and often only governments have the incredibly valuable "ground truth" data (e.g. tax returns) that acts like labels for the broader data (e.g. smartphone or TikTok usage), or government heavily regulates who can access data (e.g. medical records or genetic data). How can the Joint Force protect that data?

Siloing different sources of data about an individual is one key principle (Wright, 2020b). Received wisdom amongst many in the public and private sectors is to break down "silos", in which data in one department is isolated from the rest of the organization, much like grain in a farm silos.[23] The received wisdom is wrong. Dangerously so. To be sure, creating or preserving silos requires a trade-off because some data-sharing can bring efficiencies, but crucially it is a trade-off. The disastrous Chinese hack removing intimate data about 22 million security-cleared employees from the U.S. Office of Personnel Management illustrates an inherent problem of building a giant

[22] For an accessible recent review of available techniques see e.g. (Huguet et al., 2019). Simple options could be chosen, appropriate for the various audiences within the Joint Force and its support networks.
[23] Public sector examples include flagship World Bank reports (World Bank, 2016, 2021), and for the private sector in the Harvard Business Review (Wilder-James, 2016).

honeypot (Perera, 2015; Sanger, 2018).

Newly enhanced cyber security is also key for all these silos – whatever the Joint Force is already doing, it is almost certainly not enough. This must include supporting private sector partners.

Finally, strongly consider preventing the use of services like TikTok that cannot protect significant aspects of their users' data.

Recommendation 2.B.ii. for the Joint Force: *Prepare defenses for adversary campaigns so they can deploy deterrence, offense, emergency preparedness and emergency response – even though deterrent and offensive capabilities should be sparingly employed outside escalation or war.*

It is beyond the scope of this report to cover these topics in detail, but some points are important to raise.

Capabilities for offense and deterrence should remain relatively limited in scale compared to other aspects of defense, and should largely be kept in reserve for escalation scenarios/war. This is because all bureaucracies have the tendency to try and enhance their role and these capabilities pose particular challenges for democratic compatibility (see the later section on this topic).

Emergency preparedness and response are important to enable measured responses to events and not to over-react. This was arguably seen with aspects of the U.S. domestic surveillance after 9/11 or in some East Asian countries for surveillance of Covid-19 (Wright, 2020a).

Recommendation 2.B.iii. for the Joint Force: *Defensive information operations should employ* ***evidence-based methods for influence****.*

Defenders should use tools shown by evidence to be effective. One evidence-based framework[24] for successful influence can be broken down into three areas:

- Audience: Put the target audience's decision-making at the heart of the influence process.
- Messages: Tailor messages to maximise impact.
- Messengers: Messengers to deliver those messages must be perceived by audiences to be an appropriate voice or means of delivery.

Wright (2019)[25] provides detailed evidence-based influence, with Chapters 2-4 addressing the audience, message and messenger respectively.

Recommendation 2.B.iv. for the Joint Force: *Anticipate vulnerability at the* ***seam between domestic and foreign****. Defending it needs a coordinated, funded organizational response.*

The recent INSSG (Biden, 2021) articulates the merger of domestic and foreign policy. Operationalizing this aim requires both maintaining this

[24] This is one framework based in evidence. Others exist, based in more or less evidence.

[25] Produced for SMA, Wright ND *"From Control to Influence: Cognition in the Grey Zone"* Version 3 with updates, 2019 is available at www.intelligentbiology.co.uk/s/From-Control-to-Influence.

important distinction and to enable U.S. information defenses to coordinate from both sides.

2.C. Defense against shaping of the information terrain over years

Building strategic advantage into the information terrain over years or decades will be crucial for successful defense of individuals and organizations.

Recommendation 2.C.i. for the Joint Force: *Build defensive advantages into the information terrain, and prevent competitors developing strategic advantage via platforms like TikTok.*

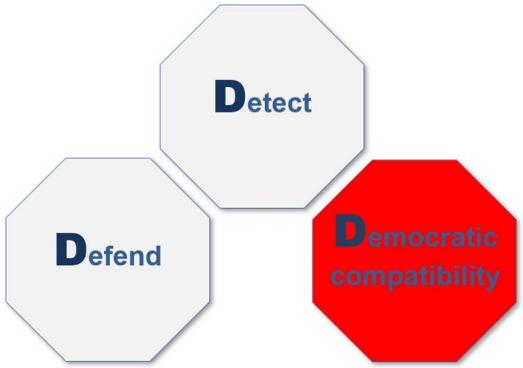
If the Joint Force conducts defensive information operations on TikTok, it will literally be competing on terrain designed by a highly capable competitor. If TikTok is not banned in the U.S., the Joint Force should ban its use where it can and discourage its use more broadly. The same is true of a host of smart cities and other technologies where the U.S. and its allies should aim to set standards, and where the Joint Force can help identify and mitigate potential information threats. Beyond the scope of this report, building alternatives to 5G technologies (in which China is strong), gaining leadership in 6G and reinvigorating U.S. innovation alongside allies will be key.

For discussion see the companion report for this SMA effort: Wright (2021) *The future character of information in strategy,* www.intelligentbiology.co.uk.

Recommendation 2.C.i. for the Joint Force: Encourage and where necessary build a thriving news and information ecosystem, with trusted messengers tailored to the distinctive needs of the multiple audiences in the Joint Force and its support networks.

Like local newspapers that help report on stories of local interest and importance, the human communities in the Joint Force require trusted and engaging news sources. This is unlikely to be met by market forces alone.

3. DEMOCRATIC COMPATIBILITY



U.S. success in Cold War information operations rested in large part on what the U.S. chose not to do. Scholar Thomas Rid's recent book, *Active Measures* (2020), describes how U.S. information operations showed considerable restraint after the early stages of the Cold War against the far more aggressive and well-resourced Soviet Union apparatus. U.S. Cold War capabilities were employed—not, of course, without imperfections—within ethical, legal and political frameworks that rendered them compatible with a free society. So too must the new U.S. capabilities as the Joint Force has embarked on a new Gray Zone conflict also likely to last decades. And this time it's digital.

First, maintain the seam between domestic and foreign – and manage the vulnerabilities that brings. This is a central theme of the new INSSG (Biden, 2021). The Joint Force should build detection and defensive capabilities without eroding safeguards such as Posse Comitatus and intelligence oversight – which as discussed above requires new methods of coordination to mitigate the gaps and lack of agility this seam entails. In addition, many of the key actions needed to defend the Joint Force against both external and internal information threats are the same (e.g. mental health and social support), and thus focusing more on these will raise fewer domestic problems.

Second, some detection and defensive capabilities carry fewer risks to democracy (e.g. enhancing social and mental health support) than others (e.g. building offensive information capabilities that can turn inwards) – and greater emphasis can be placed on safer options, which I have tried to emphasize in this report. Put simply, focus on minimizing vulnerabilities in ways that pose the least dangers to democracy, for example by helping individuals become more resilient, protecting their data and reducing TikTok's large U.S. presence.

Third, build robust ethics into the cultures and processes of the organizations and individuals charged with detecting and defending against adversary information operations. One must remain grounded in the realities these communities face because while ethics is crucial, not least for effectiveness and success, it is a topic that often makes practitioners' eyes glaze over a little when bombarded with well-meaning high-level admonitions. It is also often perceived as yet another hurdle for getting things done effectively. Thus, recognise that ethics are a key component of longer term success and provide concrete, practical guidelines (e.g. Box 5).

Fourth, ensure existing frameworks for democratic oversight are fit for current and near-future technologies – such as digital counter-intelligence at scale.

BOX 5: Omand's six ethical guidelines

Sir David Omand served as Director of the UK's Government Communications Headquarters and as the first UK Security and Intelligence Coordinator.

He proposed six ethical guidelines for security and intelligence agencies (Omand, 2006). The guidelines are designed for all intelligence activity but have here been adapted to the specific case of influence.

1. There must be sufficient sustainable cause

The 'target' of the proposed intelligence activity must be "capable of doing real damage either to the interests of the nation or to the lives and livelihoods of its citizens"

2. There must be integrity of motive

The motives of those designing, authorising, and implementing influence activity (both government and non-government partners) should be clear. Where non-government individuals and groups are involved, will their identity and their interests be appropriate protected?

3. The methods to be used must be proportionate

The techniques and methods deployed in an influence activity should be considered against the seriousness of the threat that the activity is designed to counter. Are the methods – which may include deception and/or intrusion into others' privacy – justified and proportionate?

4. There must be right authority

Government-initiated influence activities must be authorised at a "sufficiently senior level, and with accountability within a chain of command". Such activities may also be subject to independent oversight. Individuals involved in influence activities should have a way of raising concerns and issues of conscience without fear while preserving secrecy.

5. There must be reasonable prospect of success

Government-initiated influence activities should have the prospect of success in relation to their intended outcome. However, they must also take proper account of the "risks of unintended consequences, or of political or diplomatic damage, if the operation were exposed, and judge them acceptable—including applying the golden rule 'do unto others as you would be done by'."

6. Recourse to secret intelligence must be a last resort

Covert influence activities should only be considered when there are no reasonable alternatives that might include less sensitive or non-secret sources thus "avoiding all the possible moral hazards and trade-offs that a covert campaign may involve".

Conclusion

This report began by asking you to put yourself in the shoes of an adversary. From that point of view, the digital technologies now give you wonderful opportunities to reach deep into U.S. society, so that you can directly target myriad audiences in the millions of humans in the Joint Force and its support networks. The future looks bright if you have technical sophistication *and* creativity, which can be combined for information operations.

Powerful AI might help you build new influence tools. As of June 2021 China arguably has the world's most powerful called "Wu Dao 2.0" (Heikkilä, 2021). Creative ways to understand people in the US also helps. In May 2021 the Chinese fast fashion brand Shein overtook Amazon to become the most downloaded shopping app in the U.S. (Gapper, 2021). You might have many reasons for confidence.

But, what would an adversary *not* want the US to do in response?

You wouldn't want the U.S. to take long-term, effective measures to reduce vulnerabilities in the Joint Force, which reduce the impact of active measures or information operations. You might want the debates dominated by well-meaning voices telling many military personnel that they are "illiterate" at understanding media, set against other voices claiming there are no problems, and so together fueling the very discord you want to see. Instead, you would not want the U.S. to act *and* to do so with restraint, to avoid overreacting and to avoid further damaging U.S. democratic civil-military safeguards. You would want the Joint Force to focus on technological quick fixes like AI, and not on combining humans, AI and tough organizational reforms like defending the vulnerable seam between "domestic" and "foreign." You would hope that for U.S. organizations, the threats straddling the domestic-foreign seam will remain "some other organization's problem." You would not want the U.S. to rise to the INSSG's challenge of bringing together domestic and foreign policy (Biden, 2021) any more than you wanted the previous U.S. administration's focus on great power competition.

No simple answers exist. But many things—encapsulated by a strategy centered on Detect, Defend, and Democratic compatibility—would make the humans in the Joint Force and its support networks a more frustrating bunch for an adversary to influence.

References

Angelos, J. (2021, March 5). *Who wants to be a soldier? Germany grapples with far-right extremism in its ranks*. POLITICO. https://www.politico.eu/article/germany-armed-forces-far-right-extremism/

Banet, R., & Lebel, S. (2018, July 26). *No, this it not a video of a Saudi assaulting a London hospital receptionist*. Fact Check. https://factcheck.afp.com/no-it-not-video-saudi-assaulting-london-hospital-receptionist

Bateman, J. (2020). *Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2020/07/08/deepfakes-and-synthetic-media-in-financial-system-assessing-threat-scenarios-pub-82237

BBC. (2017a, April 27). German soldier posed as Syrian refugee and "planned attack." *BBC News*. https://www.bbc.com/news/world-europe-39733952

BBC. (2017b, May 7). Germany searches all army barracks for Nazi material. *BBC News*. https://www.bbc.com/news/world-europe-39835609

BBC. (2020, June 30). Germany far right: Elite KSK commando force "to be partially disbanded." *BBC News*. https://www.bbc.com/news/world-europe-53237685

Beehner, L. M., Collins, L. S., & Person, R. T. (2018). The Fog of Russian Information Warfare. In B. S. Loudon & M. D. Vertuli (Eds.), *Perceptions Are Reality: Historical Case Studies of Information Operations in Large-Scale Combat Operations*. Army University Press.

Biden, J. R. J. (2021). *Interim National Security Strategic Guidance*. EXECUTIVE OFFICE OF THE PRESIDENT WASHINGTON DC. https://apps.dtic.mil/sti/citations/AD1124337

Blue, V. (2013). *Popular plugin Social Fixer surrenders to Facebook legal menacing*. ZDNet. https://www.zdnet.com/article/popular-plugin-social-fixer-surrenders-to-facebook-legal-menacing/

Bosely, M. (2021, June 3). *TikTok accidentally detected my ADHD. For 23 years everyone missed the warning signs | Matilda Boseley*. The Guardian. http://www.theguardian.com/commentisfree/2021/jun/04/tiktok-accidentally-detected-my-adhd-for-23-years-everyone-missed-the-warning-signs

Bowes, S. M., Costello, T. H., Ma, W., & Lilienfeld, S. O. (2021). Looking under the tinfoil hat: Clarifying the personological and psychopathological correlates of conspiracy beliefs. *Journal of Personality*, *89*(3), 422–436. https://doi.org/10.1111/jopy.12588

Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation. Working Paper 2019.2*. Project on Computational Propaganda.

Bradshaw, T. (2020, January 10). *Facebook shares hit record high, surpassing 2018 peak*. https://www.ft.com/content/5fda80c8-33a3-11ea-9703-eea0cae3f0de

Bradshaw, T. (2021, July 4). ByteDance starts selling AI that powers TikTok to other companies. *Financial Times*. https://www.ft.com/content/bed7cba1-db7a-49c7-9d57-06fd19e14e10

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5). https://doi.org/10.1073/pnas.2020043118

Brewster, M. (2020, May 24). *Canadian-led NATO battlegroup in Latvia targeted by pandemic disinformation campaign | CBC News*. CBC. https://www.cbc.ca/news/politics/nato-latvia-battle-group-pandemic-covid-coronavirus-disinformation-russia-1.5581248

Brown, R. A., Helmus, T. C., Ramchand, R., Palimaru, A. I., Weilant, S., Rhoades, A. L., & Hiatt, L. (2021). *Violent Extremism in America: Interviews with Former Extremists and Their Families on Radicalization and Deradicalization*. https://www.rand.org/pubs/research_reports/RRA1071-1.html

Cancian, M. F. (2020). *U.S. Military Forces in FY 2021: Army*. Center for Strategic and International Studies. https://www.csis.org/analysis/us-military-forces-fy-2021-army

Chen, L. Y., Liu, C., & Huang, Z. (2021, March 30). ByteDance Valued at $250 Billion in Private Trades. *Bloomberg.Com*. https://www.bloomberg.com/news/articles/2021-03-30/bytedance-is-said-valued-at-250-billion-in-private-trades

Confessore, N., Dance, G. J. X., Harris, R., & Hansen, M. (2018, January 27). The Follower Factory. *The New York Times*. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Congressional Research Service. (2020). *Defense Primer: Information Operations*.

Congressional Research Service. (2021). *Defense Primer: Reserve Forces*. https://crsreports.congress.gov

Constine, J. (2020). Pex buys Dubset to build YouTube ContentID for TikTok & more. *TechCrunch*. https://social.techcrunch.com/2020/03/05/legalizing-remix-culture/

Council on Foreign Relations. (2021). *Demographics of the U.S. Military*. Council on Foreign Relations. https://www.cfr.org/backgrounder/demographics-us-military

Denney, S. (2021, March 17). *The German Far Right Doesn't Need to Win Elections to Be Dangerous*. Lawfare. https://www.lawfareblog.com/german-far-right-doesnt-need-win-elections-be-dangerous
Denning, D. E. R. (1999). *Information Warfare and Security*. ACM Press.
Elsea, J. K. (2018). *The Posse Comitatus Act and Related Matters: The Use of the Military to Execute Civilian Law*. Congressional Research Service.
Fick, M., & Dave, P. (2019, April 23). Facebook's flood of languages leave it struggling to monitor content. *Reuters*. https://www.reuters.com/article/us-facebook-languages-insight-idUSKCN1RZ0DW
Fu, Z., Burger, H., Arjadi, R., & Bockting, C. L. H. (2020). Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: A systematic review and meta-analysis. *The Lancet Psychiatry*, *7*(10), 851–864. https://doi.org/10.1016/S2215-0366(20)30256-X
Fukuyama, F., Richman, B., & Goel, A. (2021, January 26). *How to Save Democracy From Technology*. https://www.foreignaffairs.com/articles/united-states/2020-11-24/fukuyama-how-save-democracy-technology
Gapper, J. (2021, June 18). Shein leaves Boohoo and Zara on fast fashion's shelf. *Financial Times*. https://www.ft.com/content/4d079978-65f4-47b8-a572-ed1f7b6a8e61
Heikkilä, M. (2021, June 9). *Meet Wu Dao 2.0, the Chinese AI model making the West sweat*. POLITICO. https://www.politico.eu/article/meet-wu-dao-2-0-the-chinese-ai-model-making-the-west-sweat/
Helmus, T. C., Zimmerman, S. R., Posard, M. N., Wheeler, J. L., Ogletree, C., Stroud, Q., & Harrell, M. C. (2018). *Life as a Private: A Study of the Motivations and Experiences of Junior Enlisted Personnel in the U.S. Army*. https://www.rand.org/pubs/research_reports/RR2252.html
Huguet, A., Kavanagh, J., Baker, G., & Blumenthal, M. S. (2019). *Exploring Media Literacy Education as a Tool for Mitigating Truth Decay*. https://www.rand.org/pubs/research_reports/RR3050.html
Hwang, T. (2020). *Deepfakes: A Grounded Threat Assessment*. Center for Security and Emerging Technology. https://cset.georgetown.edu/research/deepfakes-a-grounded-threat-assessment/
ICO. (2019, May 8). *Microtargeting*. https://ico.org.uk/your-data-matters/be-data-aware/social-media-privacy-settings/microtargeting/
Jones, S. G., Doxsee, C., & Hwang, G. (2021, April 12). *The Military, Police, and the Rise of Terrorism in the United States*. https://www.csis.org/analysis/military-police-and-rise-terrorism-united-states
Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. genannt, Furukawa, T. A., Cuijpers, P., & Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration. (2021). Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. *JAMA Psychiatry*, *78*(4), 361–371. https://doi.org/10.1001/jamapsychiatry.2020.4364
Keller, M. H., & Dance, G. J. X. (2019, September 28). The Internet Is Overrun With Images of Child Sexual Abuse. What Went Wrong? *The New York Times*. https://www.nytimes.com/interactive/2019/09/28/us/child-sex-abuse.html, https://www.nytimes.com/interactive/2019/09/28/us/child-sex-abuse.html
Kelly, M., & Alexander, J. (2019, November 13). *YouTube's new kids' content system has creators scrambling*. The Verge. https://www.theverge.com/2019/11/13/20963459/youtube-google-coppa-ftc-fine-settlement-youtubers-new-rules
Kent, D. M., Steyerberg, E., & Klaveren, D. van. (2018). Personalized evidence based medicine: Predictive approaches to heterogeneous treatment effects. *BMJ*, *363*, k4245. https://doi.org/10.1136/bmj.k4245
Laubscher, H. (2019). *Tmall 2.0 Goes Big On Customization*. Forbes. https://www.forbes.com/sites/hendriklaubscher/2019/07/11/tmall-20-unlocks-game-changing-brand-building-opportunities/
Leskin, P. (2019). *The AI tech behind scary-real celebrity "deepfakes" is being used to create completely fictitious faces, cats, and Airbnb listings*. Business Insider. https://www.businessinsider.com/deepfake-tech-create-fictitious-faces-cats-airbnb-listings-2019-2
Liao, R. (2021, February 18). TikTok's China twin Douyin has 550 million search users, takes on Baidu. *TechCrunch*. https://social.techcrunch.com/2021/02/17/short-video-search-douyin-tiktok/
Lopez, C. T. (2019). *DOD to Take Over Background Checks by Fiscal 2020*. U.S. DEPARTMENT OF DEFENSE. https://www.defense.gov/Explore/News/Article/Article/1886923/dod-to-take-over-background-checks-by-fiscal-2020/
Martin, D. A., Shapiro, J. N., & Ilhardt, J. G. (2020). *Trends in Online Influence Efforts (V.2.0)*. https://scholar.princeton.edu/sites/default/files/jns/files/trends_in_online_influence_efforts_v2.0_aug_5_2020.pdf
McCarthy, J. (2019). *Social media ad budgets continue to grow at "expense of print", up 20% in 2019*. The Drum. https://www.thedrum.com/news/2019/10/07/social-media-ad-budgets-continue-grow-expense-print-up-20-2019

Molla, R. (2019, February 13). *It will take more than NewsGuard's team of journalists to stop the spread of fake news*. Vox. https://www.vox.com/2019/2/13/18220746/real-journalists-fake-news-newsguard
Murgia, M., & Murphy, H. (2019, November 8). *Can Facebook really rely on artificial intelligence to spot abuse?* https://www.ft.com/content/69869f3a-018a-11ea-b7bc-f3fa4e77dd47
Naughton, J. (2019, January 20). "The goal is to automate us": Welcome to the age of surveillance capitalism. *The Observer*. https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook
Newman, L. H. (2021, January 28). Facebook Ad Services Let Anyone Target US Military Personnel. *Wired*. https://www.wired.com/story/facebook-ad-targeting-us-military/
Ogrysko, N. (2019, September 6). *The future of continuous evaluation is just about here, and it has a different name*. Federal News Network. https://federalnewsnetwork.com/workforce/2019/09/the-future-of-continuous-evaluation-is-just-about-here-and-it-has-a-different-name/
Omand, D. (2006). Ethical Guidelines in Using Secret Intelligence for Public Security. *Cambridge Review of International Affairs*, *19*(4), 613–628. https://doi.org/10.1080/09557570601003338
Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*.
Perera, D. (2015). *Lawsuit seeks relief from cyberspying - CIA and OPM: Rethinking the silo*. POLITICO. https://www.politico.com/tipsheets/morning-cybersecurity/2015/07/lawsuit-seeks-relief-from-cyberspying-cia-and-opm-rethinking-the-silo-212543
Pew. (2021, April 7). *Social Media Use in 2021*. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/
Posard, M. N., Ellinger, E., Ryan, J., & Girven, R. S. (2021). *Updating Personnel Vetting and Security Clearance Guidelines for Future Generations*. https://www.rand.org/pubs/research_reports/RRA757-1.html
Reuters. (2019). *David Beckham's "deep fake" malaria awareness video | Reuters Video*. https://www.reuters.com/video/watch/david-beckhams-deep-fake-malaria-awarene-idOVA9QVUY3
Rid, T. (2020). *Active Measures: The Secret History of Disinformation and Political Warfare* (Illustrated edition). Macmillan USA.
Rollwage, M., Dolan, R., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology*.
Rollwage, M., Zmigrod, L., de-Wit, L., Dolan, R. J., & Fleming, S. M. (2019). What Underlies Political Polarization? A Manifesto for Computational Political Psychology. *Trends in Cognitive Sciences*, *23*(10), 820–822. https://doi.org/10.1016/j.tics.2019.07.006
Rosen, C. S., Morland, L. A., Glassman, L. H., Marx, B. P., Weaver, K., Smith, C. A., Pollack, S., & Schnurr, P. P. (2021). Virtual mental health care in the Veterans Health Administration's immediate response to coronavirus disease-19. *The American Psychologist*, *76*(1), 26–38. https://doi.org/10.1037/amp0000751
Sanger, D. E. (2018). *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age*. Penguin Random House USA.
Sawers, P. (2018, November 7). YouTube: We've invested $100 million in Content ID and paid over $3 billion to rightsholders. *VentureBeat*. https://venturebeat.com/2018/11/07/youtube-weve-invested-100-million-in-content-id-and-paid-over-3-billion-to-rightsholders/
Schaeffer, K. (2021, April). The changing face of America's veteran population. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2021/04/05/the-changing-face-of-americas-veteran-population/
Sehl, K. (2021, May 5). 23 Important TikTok Stats Marketers Need to Know in 2021. *Social Media Marketing & Management Dashboard*. https://blog.hootsuite.com/tiktok-stats/
Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The Weaponization of Social Media*. Houghton Mifflin Harcourt.
Stevenson, P. (2016). *Security clearance investigations to include social media activity*. National Guard. https://www.nationalguard.mil/News/Article/799056/security-clearance-investigations-to-include-social-media-activity/
Stewart, P. (2020, December 3). U.S. military families in South Korea? Top U.S. general wants a rethink. *Reuters*. https://www.reuters.com/article/usa-military-korea-idUSKBN28D3DK
The Economist. (2021, May 22). NATO increasingly sees its soldiers' phones as a liability. *The Economist*. https://www.economist.com/europe/2021/05/22/nato-increasingly-sees-its-soldiers-phones-as-a-liability?frsc=dg%7Ce
Tiku, N. (2017, May 21). Welcome to the Next Phase of the Facebook Backlash. *Wired*. https://www.wired.com/2017/05/welcome-next-phase-facebook-backlash/
Véliz, C. (2020). *Privacy is Power: Why and How You Should Take Back Control of Your Data*. Bantam Press.
Vogel, A., & Wright, N. D. (2019, May 10). *Alexa Is Both Friend and Sales Robot. That's a Problem.* Slate Magazine. https://slate.com/technology/2019/05/alexa-amazon-voice-assistant-conflict-interest-regulation.html

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
Walker, C. (2018). What Is “Sharp Power”? *Journal of Democracy*, *29*(3), 9–23. https://doi.org/10.1353/jod.2018.0041
Wilder-James, E. (2016, December 5). Breaking Down Data Silos. *Harvard Business Review*. https://hbr.org/2016/12/breaking-down-data-silos
Winkie, D. (2020, October 19). *Deployed soldiers face punishment for their ‘message to liberals’ video*. Army Times. https://www.armytimes.com/news/your-army/2020/10/16/deployed-soldiers-face-punishment-for-their-message-to-liberals-video/
Woolley, S. (2020). *We’re fighting fake news AI bots by using more AI. That’s a mistake.* MIT Technology Review. https://www.technologyreview.com/s/614810/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/
World Bank. (2016). *World Development Report 2016: Digital Dividends*. World Bank Publications.
World Bank. (2021). *World Development Report 2021: Data for Better Lives*. The World Bank.
Wright, N. D. (2019a). *From control to influence: Cognition in the Grey Zone (v3)* (p. 158). Intelligent Biology. www.intelligentbiology.co.uk
Wright, N. D. (Ed.). (2019b). *Artificial Intelligence, China, Russia, and the Global Order*. Air University Press.
Wright, N. D. (2020a, April 13). *Coronavirus and the Future of Surveillance*. https://www.foreignaffairs.com/articles/2020-04-06/coronavirus-and-future-surveillance
Wright, N. D. (2020b). *Artificial Intelligence and Democratic Norms: Meeting the Authoritarian Challenge*. National Endowment for Democracy. https://www.ned.org/sharp-power-and-democratic-resilience-series-artificial-intelligence-and-democratic-norms/
Zipperstein, S. J. (2020). The Conspiracy Theory to Rule Them All. *The Atlantic*. https://www.theatlantic.com/politics/archive/2020/08/conspiracy-theory-rule-them-all/615550/