Australian Government
Department of Defence

EDTAS
EMERGING DISRUPTIVE TECHNOLOGY
ASSESSMENT SYMPOSIUM

INFORMATION AND INFLUENCE

INSIGHTS PAPER

EDTAS 2023-2024

EDTAS 2023-2024 Insights Paper

Produced by:

DSTG Technology Futures & Foresight and Atturra in collaboration with
DSTG Information Warfare STaR Shot

LDI: Feb 2024

Contents

Executive Summary 4

Introduction 6

Background 6

Definitions 7

Historical Context 7

Aim and Scope of Insights Paper 8

Key Insights 10

Digitised Existence and Open-Source Data 11

Evolution of Generative AI and Synthetic Media 16

Vectors of Influence and Disinformation 23

Automation and AI in the Cyber Domain 28

Tools for Managing the Information Environment 32

Additional themes 37

Australian Perspective 38

Outlook 44

2023-2024 EDTAS Campaign Going Forward 46

References 47

Glossary 49

Executive Summary

Malicious activity in the information environment is one of the most significant challenges facing Australia today, affecting both national security and the landscape of modern warfare. For this reason, information warfare is a priority area in the Defence Strategic Review.

As articulated in the Defence Science and Technology Group (DSTG) 'More, together' strategy, meeting these complex challenges requires focused, mission-directed research effort across the national science and technology enterprise[1]. The Emerging Disruptive Technology Assessment Symposium (EDTAS) campaigns bring together experts across academia, industry and Defence to explore new disruptive trends and technological developments and to contribute to shaping future strategy, policy and programs.

The 2023-2024 EDTAS campaign looks at the information environment, specifically focusing on information and influence. The first stage of the EDTAS campaign has sought insight across each of these areas via a series of interviews with subject matter experts (SMEs), identified by Defence Science and Technology Group (DSTG). The collective outcomes of these interviews are synthesised and summarised in this Insights Paper.

Five key themes we derived from interviews and are discussed in this document:

1. **Digitised existence and open-source data** arising from proliferation of networked devices and greater online engagement are gradually erasing the separation between physical and online worlds. It increases vulnerability to cyber-attacks and influence, and enables mass surveillance by different actors, creating new vectors of influence. Opportunities lie in utilising the available data in research on combating mis/disinformation, regulation of data practices, and building population resilience to cyber-attacks and influence.

[1] Defence Science and Technology Group (n.d.)b

2. **Evolution of generative artificial intelligence (AI)** and synthetic media brings the capacity for rapid, cost-effective creation of multi-modal content. AI systems can mimic human interactions and integrate a range of different functions. The risks lie in amplification of mis/disinformation and individualised targeting at scale. Opportunities can be found in development of trustworthy AI systems and AI tools for countering mis/disinformation. In the military domain, effective human-AI partnering can strengthen the decision-making process.
3. **Vectors of influence and disinformation** are best understood at the intersection of computer science and social and behavioural sciences. These disciplines are significant due to the key role of individual and collective behaviours in cyber activity and influence. Maintaining a narrow focus on technology risks missing the benefits of interdisciplinary research: understanding human activity online, improving cybersecurity, and countering mis/disinformation.
4. **Automation and AI in the cyber domain** widen the access, reach and scale of cyber-attacks and speed up the decision-making loop in both civilian and military contexts. Trustworthy automation and AI can help counteract these effects, while addressing the skills shortage in cybersecurity.
5. **Tools for managing the information environment** comprise systems for detecting, attributing, and measuring the impact of influence campaigns, as well as simulation, experimentation and wargaming methods that incorporate information environment effects and AI. These are subject to competitive development across the globe and will be crucial to the management of the information environment going forward.

Recommendations from the Australian perspective include establishment of coordinated sovereign capability in cybersecurity and AI, establishing mechanisms for interdisciplinary projects and Defence partnerships, support for research and development ecosystems, and improving the capacity for commercialisation of innovative ideas.

Introduction

Background

The Emerging Disruptive Technology Assessment Symposium (EDTAS) campaigns bring together a diverse range of experts across academia, industry and Defence for in-depth exploration of emerging technologies and trends. The aim of EDTAS campaigns is to identify key challenges and opportunities for Australia and to help shape future strategy, policy and programs for Defence and national security[2]. The collaborative nature of the EDTAS campaigns reflects the Defence Science and Technology Group (DSTG) 'More, together' strategy which seeks to deliver strategic advantage through mission-focused research[3].

One of the most significant and complex challenges facing Australia today arises from malicious online activity and disinformation campaigns, which endanger individual rights, threaten social cohesion and trust in democratic institutions and impact national security. Information warfare is a significant part of the evolving nature of military conflicts.

The 2023-2024 EDTAS campaign considers the information environment, specifically focusing on information and influence.

This topic is explored in three phases:

- Subject-matter expert (SME) consultation
- Information and Influence Symposium, and
- Military Implications Symposium

This Insights Paper summarises the key themes from the SME consultation phase of the EDTAS campaign.

[2] Defence Science and Technology Group (n.d.)a
[3] Defence Science and Technology Group (n.d.)b

Definitions

The information environment includes people, organisations and systems that collect, process, share and act on information as well as the information itself. Australian doctrine describes the information environment in terms of three dimensions: physical, cyber and human.

Within Defence, information warfare is defined as the contest for the provision and assurance of information to support friendly decision-making, whilst denying and degrading that of adversaries. It includes a range of information-related activities including cyberwarfare, electromagnetic spectrum operations and information operations/influence activities. Information warfare spans the spectrum of conflict from the shaping, influencing and deterrence activities characteristic of sub-threshold activities through to the non-kinetic effects within the cyber domain during high intensity conflicts.

Historical Context

Although recent technological advances have sharpened the focus on the cyber dimension, the strategy of influence in the information environment is pervasive in human history. From printing press, radio and television through to the multi-layered communication channels of social media and artificial intelligence (AI) chatbots, the mass media tools of influence have continued to evolve, often with surprising and disruptive effects[4].

Information warfare, including grey-zone activities, has been a recurring feature during military operations. For example, the deception campaign leading up to the invasion of Normandy in WWII featured a combination of radio broadcasts and creation of fake armies and infrastructure in order to mislead the German High Command about the location of the impending Allied invasion of Europe[5]. Whereas during military operations in Afghanistan, US officials had to counter disinformation campaigns directed against US troops[6].

[4] Rosenberg (2023)
[5] Latimer (2001)
[6] Zucchino (2009)

Today, the strategic implications of deliberate disinformation campaigns are widely recognised, with organisations such as North Atlantic Treaty Organization (NATO) forewarning that such campaigns offer adversaries a way of bypassing the traditional battlefield and a means of radically altering Western societies by pushing international norms and behaviours[7].

In Australia, the recent release of the Defence Strategic Review (DSR) has put a spotlight on information warfare as a priority area for Defence, DSTG, and the newly formed Advanced Strategic Capabilities Accelerator (ASCA). The DSR directs resources to crucial future-focused joint capabilities such as information warfare. EDTAS seeks to assist in this process by identifying disruptive technologies that could impact the future of information warfare and by connecting industry, academia and government researchers with end users to better understand these opportunities and their military implications.

Aim and Scope of Insights Paper

This Insights Paper summarises five key themes arising from the EDTAS SME consultations conducted in March-April 2023:

1. Digitised existence and open-source data
2. Evolution of generative AI and synthetic media
3. Vectors of influence and disinformation
4. Automation and AI in the cyber domain
5. Tools for managing the information environment

Each theme includes a description of key terms and concepts, and a summary of emerging trends, disruptive effects and opportunities.

[7] Masakowski & Blatny (2023)

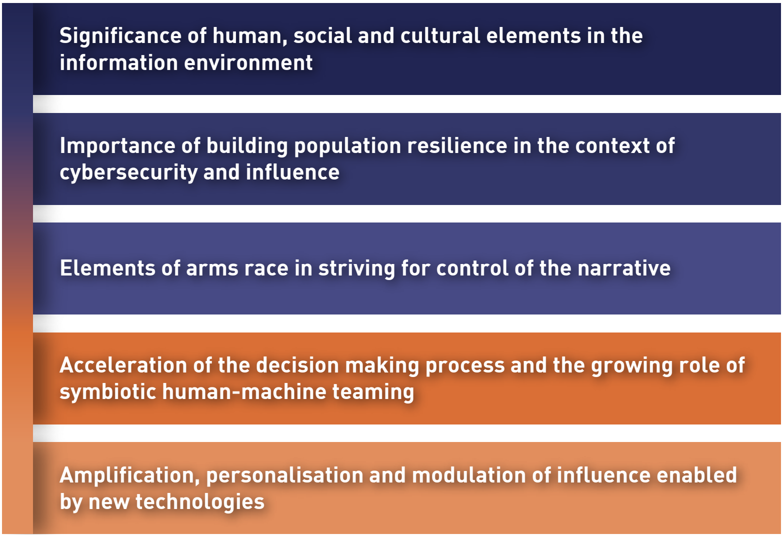
Whilst offensive cyber operations are a key component of information warfare, the classification of work in this space places it outside the scope of this document. The topics of communications and electronic warfare warrant in-depth exploration in their own right and are not addressed in this EDTAS campaign.

The Insights Paper sets the scene for the next phase of the EDTAS campaign, the Information and Influence Symposium. The Symposium will explore these themes within a structured process that immerses the participants in the art of the possible within the realm of information and influence.

Key Insights

> *'The disruption won't come from any one technology, but convergence of different technologies that will change the way we do things and think about things.'* – SME interviewee

This section further explores the five key themes identified during the SME consultations, while providing explanation of key terms and concepts and giving historical examples. Figure 1 below highlights common discussion threads across all five topics.



Significance of human, social and cultural elements in the information environment

Importance of building population resilience in the context of cybersecurity and influence

Elements of arms race in striving for control of the narrative

Acceleration of the decision making process and the growing role of symbiotic human-machine teaming

Amplification, personalisation and modulation of influence enabled by new technologies

Figure 1. Key aspects of the information environment and influence

DIGITISED EXISTENCE AND OPEN-SOURCE DATA

'Information doesn't respect boundaries or borders.' – SME interviewee

Key terms and concepts[8]

Advances in communication networks, computational power and proliferation of networked devices have enabled digitisation of most aspects of life. The key aspects of increasingly digitised existence include the following:

Distributed approach to large-scale sensing, storage, computation, decision making, and research and development, enabled by edge computing, low-cost sensor networks, and decentralised production (through AI-assisted design, democratised access and additive manufacturing).

Generation of massive amounts of data, which is exploited through advanced analytical methods, including AI, and advanced mathematics to provide hitherto impossible insights at speed. This can include collection of personal data such as biometrics, location, movement and micro-expressions, including for building individual user profiles over time.

Blending of the physical, cyber and human dimensions, which creates new cognitive or physical realities: Internet of Things (IoT), smart city ecosystems, 'digital twins' of real world systems, and virtual worlds and communities such as Meta's 'metaverse'.

Similarly, in the increasingly digitised military domain, functions such as command and control (C2), cyber operations, intelligence, surveillance and reconnaissance (ISR), and control of un-crewed systems all rely on and generate vast quantities of data, which contributes to the common operating picture.

[8] Masakowsky & Blatny (2023); Rosenberg (2023)

Emerging Trends

Digital connectivity, greater engagement on the Internet, and proliferation of networked devices (critical infrastructure, cars, household items, implants) is erasing the separation between the information environment and physical world.

Increasingly digitised existence, in turn, generates enormous volumes of different types of data, which has become a valuable resource for private firms and nation states. In fact, a number of technology companies are now taking on some of the power and functions of nation states.

In the military domain, companies such as Microsoft and StarLink increasingly supplement or provide supporting infrastructure for communications and data management – both within and outside of formally contracted arrangements. The Ukraine-Russia conflict has shown that actions of single companies can contribute significantly to a country's defence.

Disruptive Effects

Proliferation of networked devices and services creates multiple points of vulnerability for cyber-attacks and data theft, which has been used to perpetuate identity theft, financial fraud and blackmail.

At the same time, the greater number of devices that track motion, location and online exchanges pave the way for mass surveillance. Surveillance technology and tactics often go under the radar with little oversight or monitoring and with little understanding as to who holds and controls the data. In many cases, this data is taken off-shore and could be used by state and commercial entities to promote their agenda.

Democratic societies are particularly vulnerable to shaping activities driven by big data analytics. Individual data may be captured in order to micro-target people for political and other purposes. In some cases, this creates social polarisation along party or cultural lines that supersedes national interests.

Furthermore, large technology companies are increasingly challenging the sovereignty of nation states, often operating by offsetting the cost of damage inflicted by their platforms to the taxpayer. When governments attempt to negotiate or constrain these companies, they have variable success, as shown in the case of the Australian Government making changes to the media code after suffering the impacts of Facebook's news ban in Australia, which affected essential services.

Additionally, online communities, such as Meta's 'metaverse', create avenues for parties with commercial and other interests to manipulate narratives and messaging to the community members.

Historical Example

In 2018, it was exposed that Cambridge Analytica, a private British company, harvested data from millions of individuals and groups on Facebook in order to target them with tailored messages. Cambridge Analytica used both qualitative and quantitative data to build psychological profiles that informed the design of targeted content, with the goal of influencing public opinion at scale. Although the effectiveness of Cambridge Analytica methods has been called into question, this incident demonstrated the potential risks associated with exploitation of data from online interactions.[9]

[9] Ebbott et al. (2021)

Opportunities

Just as the use of data associated with digitised and networked existence creates a range of risks, a number of strategies can be applied to benefit from and safeguard the generated data:

- Investing in data science and data analytics will help build new theories and models for human interactions online in order to combat mis/disinformation.
- Regulation of data ownership and legislation that keeps certain types of data within the citizen's country will help safeguard against targeting of users by adversaries.
- Building resilience of the population through informed public debate, education, training and awareness campaigns will increase the security of individual citizens.
- Regulation of content will help combat persistent bias, mis/disinformation and deliberate manipulation online.
- Putting the onus onto the technology platform companies to mitigate and account for the damage their platforms inflict will strengthen citizens' rights and the quality of online engagement.

EVOLUTION OF GEN-AI AND SYNTHETIC MEDIA

'There's something attractive, powerful, useful, and mysterious there and it's going to change our lives fundamentally and we don't know how. It's appropriate to be a little bit worried.' – SME interviewee

Key terms and concepts[10]

The last decade has witnessed an explosion in AI research and applications. Task-specific models, which have dominated the AI landscape to date, are being overtaken by multimodal (sound text, video, voice, etc.) foundation models (MFMs). This type of generative AI performs well over a wide variety of tasks and with different types of input and output. MFMs are trained on wide-ranging data at scale so that they can perform a variety of downstream tasks.

Early examples of MFMs are pre-trained large language models (LLMs) including Google's Bard and OpenAI's GPT-4. These and other models have been developed by private enterprises, sometimes in an open manner, and released for public consumption as either free or paid service. Subsequently, an ecosystem has emerged around MFMs with a host of websites, services and plug-ins that utilise them to do everything from writing essays to designing, planning and executing scientific experiments.

One of the uses of AI techniques is in augmenting chatbots – programs that simulate a human conversation. While not used in all chatbots, LLMs and conversational AI techniques such as natural language processing are a key feature of many modern chatbots, allowing the programs to understand the user's questions, produce relevant content and generate conversational responses.

Generative adversarial network (GAN) is a class of machine learning framework used in generative AI that has gained prominence in generation of synthetic media and fake content. GAN is based on a contest between two neural networks: a generator that creates content and a discriminator that assesses the authenticity of content, with both neural networks learning dynamically.

[10] Goldstein et al. (2023); consultation with DSTG Strategic Futures Project SMEs (26 Oct 2023)

This approach enables generation of realistic images and full or partial manipulation of videos in real time (e.g. live faces swap), creating the capability for production of synthetic content, deep fakes and dissemination of targeted disinformation at scale.

Known issues within AI models include propagation of bias inherent in the training data, and errors in output presented as statement of fact. The potential for harm increases with complexity of task, which can cause misalignment of AI objectives with the original intent.

Emerging Trends

Large investments by nation states as well as private firms, such as Microsoft, Alphabet, Amazon and Meta, have underpinned the incredible evolution of generative AI and its proliferation across many sectors.

Generative AI now supports a range of cheap, easy, readily accessible tools for creation and automation of content in different modes: speech, video, audio, and image. AI systems can have conversations and arguments like humans; post tweets and retweets; create photo-quality images, artworks, videos, memes, and audio; and clone voices of real human beings. The quality and speed of generated content, including deep fakes, continues to improve at an astonishing rate.

While currently relatively primitive, AI-enabled bots will grow increasingly human-like and capable of complex personalised interactions that integrate speech, audio and video. While some synthetic content will not be discernible to the human eye in the next five years, in the next decade it will become impossible to differentiate between humans and AI in online interactions. The emergence of hybrid human-machine relationships will impact individuals and societies in ways that are difficult to predict.

At the same time, AI assistants and plug-ins increasingly integrate different aspects of life: shopping, cooking, writing, religious practices, political participation, fashion, and sports. In the future, an AI assistant may know all the details of one's life and serve to provide tailored advice and information, replacing search engines and effectively becoming one's window to the world.

It is also recognised that restricted access to algorithms and training data, combined with a number of poorly understood emergent effects, has led many to describe sophisticated AI as 'Black Box' systems lacking of clarity around the logic underlying AI outputs and the motivations behind AI development.

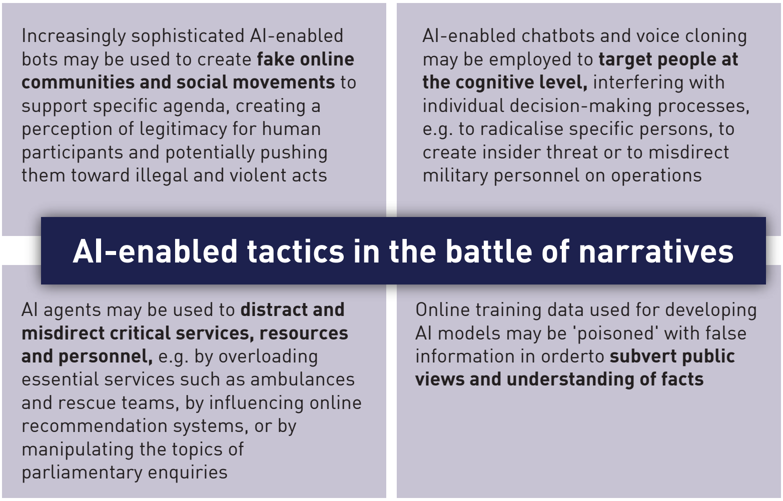
Disruptive Effects

'Every second person on the street is a proxy soldier in a war they don't even know is being waged.' – SME interviewee

The ability to generate convincing synthetic content in different modes can be used to amplify, target and automate disinformation campaigns, increasing their scale and reach and reducing costs. AI systems will serve to identify areas of vulnerability and potential avenues of attack as well as generate the relevant code.

Synthetic content and deep fakes will become highly personalised and targeted, creating a 'firehose of disinformation'. This capability will be used by actors with political and social agendas to disrupt social and political discourse and propagate ideological messages, leveraging individual targeting at scale. Messaging will be fine-tuned via behavioural models of individual users based on longitudinal data from online interactions, including on social media. Monitoring of user's reactions will enable modulation of interactions in real time.

The advent of AI assistants and human-like AI conversational agents opens up another avenue for targeted influence and manipulation: commercial, political and ideological. Figure 2 summarises the examples of disruptive influence effects that AI agents may enable.



AI-enabled tactics in the battle of narratives

Increasingly sophisticated AI-enabled bots may be used to create **fake online communities and social movements** to support specific agenda, creating a perception of legitimacy for human participants and potentially pushing them toward illegal and violent acts

AI-enabled chatbots and voice cloning may be employed to **target people at the cognitive level,** interfering with individual decision-making processes, e.g. to radicalise specific persons, to create insider threat or to misdirect military personnel on operations

AI agents may be used to **distract and misdirect critical services, resources and personnel,** e.g. by overloading essential services such as ambulances and rescue teams, by influencing online recommendation systems, or by manipulating the topics of parliamentary enquiries

Online training data used for developing AI models may be 'poisoned' with false information in orderto **subvert public views and understanding of facts**

Figure 2. Examples of potential AI-enabled tactics in the battle for control of narrative

The potential consequences are significant: altered outcomes of elections, social fragmentation, and misallocation of resources and erosion of trust.

Historical Example

At the start of the Ukraine-Russia conflict, a deep fake was shared on Twitter, comprising a synthetically generated video of President Zelensky talking of surrendering to Russia and urging Ukranians to put down their weapons. Although almost immediately discredited in this instance, this kind of disinformation can affect both morale and actions of the local population.[11]

[11] Wakefield (2022)

Opportunities

'We need mechanisms for trust.' – SME interviewee

The debate around regulation and trustworthiness of AI models has focused on the notions of validity, security, explainability, and responsibility. These mean that an AI model should do what it is intended to do, be resilient to adversarial conditions, provide logical and relevant justification for its outputs, and comply with ethical and legal frameworks.[12] The Australian Government AI ethics principles include human, societal and environmental wellbeing; human-centred values; fairness, privacy protection and security; reliability and safety; transparency and explainability; contestability; and accountability.[13]

The SME discussions have similarly highlighted the need for explainable and predictable AI, with regulation of data and algorithms, in order to ensure that the logic of AI outputs and the evidence behind them can be understood and to check for bias. This may be achieved through development of sovereign AI capability, by building on open-source models with access to training data, and/or by employing smaller models tailored to specific tasks.

The use of AI tools for propagating disinformation may be countered by building population resilience through education and awareness campaigns. This is particularly important for educating the public about the strengths, weaknesses and appropriate use of technologies like ChatGPT, and about the approaches that leverage generative AI to influence individuals and to create disinformation at scale. At the same time, trusted information sources will help align online narratives with public interest and democratic ideals.

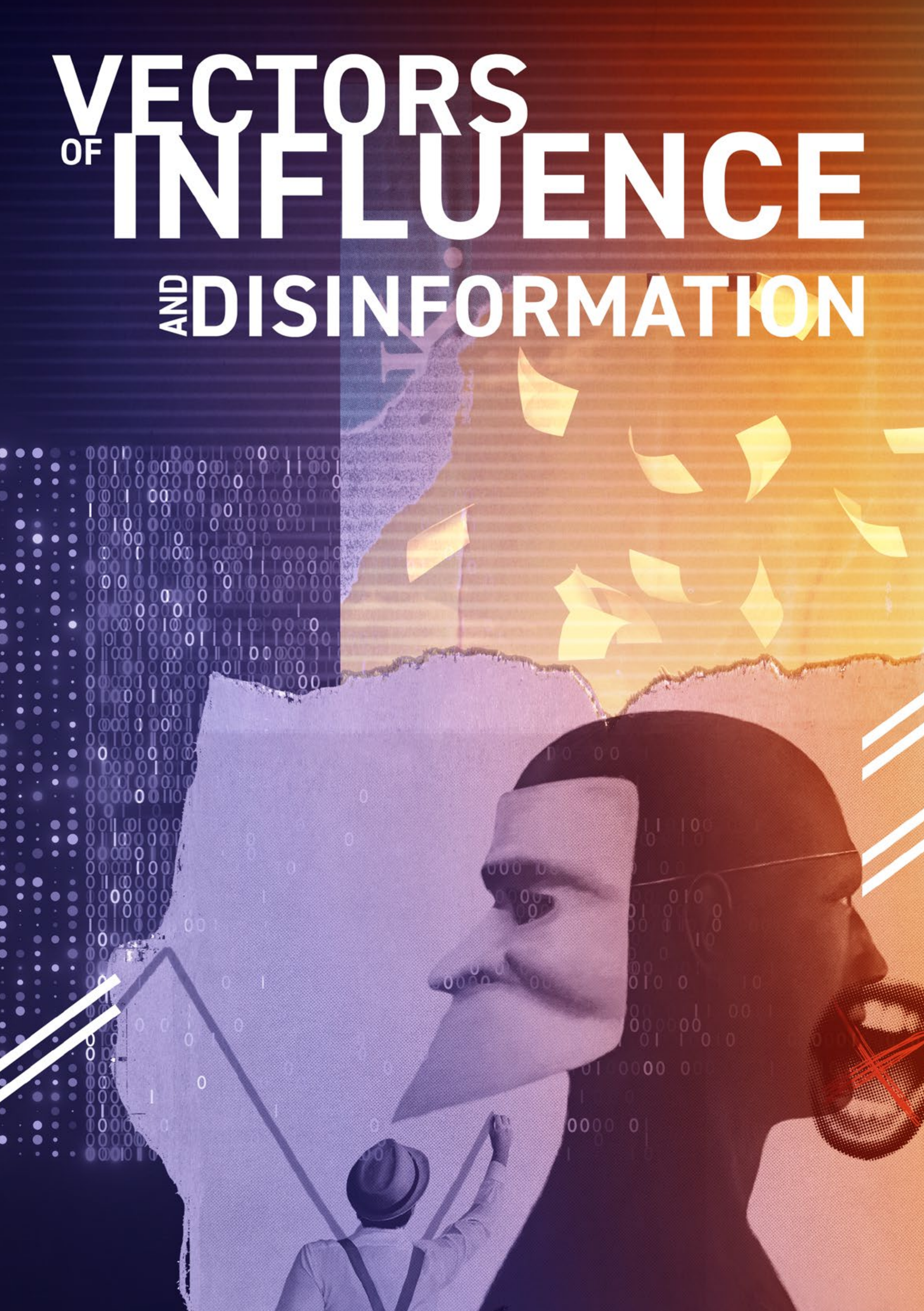
[12] Masakowski & Blatny (2023)

[13] Australian Government Department of Industry, Science and Resources (n.d.)

Another opportunity lies in leveraging both machine and human intelligence to build new models of decision-making processes, particularly under conditions of uncertainty. Developing the fundamental science that supports effective teaming of human and machine cognition will make AI into a partner for human operators, including in the military context.

While the nature of the evolving AI systems has increasingly raised questions around emergence, consciousness, and the risks of superhuman intelligence, there may also be opportunities associated with building 'thinking machines' – AI models that would work in a similar way to a human mind, developing a sophisticated model of the world over time. AI systems capable of abstract thinking do not yet exist, but may offer more valuable support in complex strategic planning in the future.

VECTORS OF INFLUENCE AND DISINFORMATION



'This cognitive aspect is an important thing ... That's where the big leaps will happen. All we've got now is clever maths.' – SME interviewee

Key terms and concepts[14]

Vectors of influence and disinformation are best understood at the intersection of computer science and social and behavioural sciences. Social and behavioural sciences seek to understand individual and group behaviours, patterns, processes and structures. These disciplines (e.g. cognitive, organisational and social psychology, sociology, and ethnographic research) provide insight into human factors within complex socio-technical systems. The associated research practices use a wide range of theories, analytical frameworks, methods, and techniques.

In the context of information environment and influence, social and behavioural sciences enable understanding of the affordances, modalities and vectors of online influence at the micro (individual), meso (group) and macro (society) levels. This includes the study of mechanisms that adversaries may use to leverage online behaviours and profiles to conduct remote influence and disinformation campaigns at scale.

Social and behavioural sciences help elucidate the mechanisms of influence such as amplifying and exploiting social and political divides, disseminating rumours to generate collective anxiety, and exploiting cognitive errors in the decision making process. These disciplines can also help develop strategies for building societal resilience to disinformation and influence.

Understanding the mechanisms of influence is further enabled by contributions from neuroscience, political science, linguistics, and cultural and legal studies.

[14] Ebbott et al. (2021); Masakowsky & Blatny (2023); consultation with DSTG IWSS SMEs (Dec 2023)

Emerging Trends

Exploitation of individual and collective human behaviour is a key part of influence campaigns and cyber-attacks. The scientific understanding of the underlying mechanisms continues to expand. The key challenges lie in the translation of refined models and theories from the lab environment into the complex real world, and in linking models of behaviour at micro, meso and macro levels.

Additionally, there is a growing need to understand the differences between human and machine cognition and how they interact, as human and machine cognition are fundamentally different. Understanding these processes is important in interpreting AI outputs and for leveraging machine intelligence in human decision making, even as human-oriented interfaces are becoming increasingly intelligent.

Disruptive Effects

Many conversations about information environment, information warfare and influence focus on technology; this focus is too narrow and doesn't account for human factors. Most AI developers are not researching cognition, and the value of interdisciplinary approaches is often underestimated. In fact, the greatest impact in the information environment is likely to come from understanding of the mechanisms and vectors of influence at different levels of human interactions, and this is a body of knowledge that could be used by adversaries to disrupt social structures.

Additionally, there is a growing need understand how AI intersects with and affects human behaviour because it is becoming increasingly important in the decision-making process. Lack of understanding of the fundamental differences between human and machine cognition means that the two are not combined effectively, thus affecting the ability to utilise AI in complex settings for planning and wargaming.

In the longer term, risks may arise due to development of artificial generalised intelligence or a system with emergent properties that may be learning at a superhuman rate. These risks need to be assessed in the context of human timescales, human cognitions and human needs.

Opportunities

Understanding and managing the information environment requires an interdisciplinary approach, bringing disciplines such as psychology, political science, cognitive science and sociology into the discussion of online influence. Interdisciplinary studies are essential for advancing the research into human interactions in online environments and for interpreting the vast volumes of generated data, building models of human behaviour at micro, meso and macro levels.

Potential interdisciplinary research directions in this space include the following:

- Analysis of available data and development of tools for interpreting the data with involvement of cultural experts
- Understanding the psychological impacts of cyber-attacks and disinformation campaigns and how an adversary could use prediction of human behaviour
- Involvement of linguists to understand how language and narratives are used to build communities and how language and narratives reflects assumptions and worldviews
- Advancing understanding of machine cognition, its evolution and the way machines process information
- Development of effective human-oriented interfaces for processing large amounts of information and supporting semantic communication to human users.

An interdisciplinary approach with involvement of social and behavioural sciences will strengthen strategies for combating mis/disinformation, polarisation, cyber-attacks and hate crimes, improving cognitive security and resilience of the population, and countering false narratives. It will support the development of tools for detecting intent and reasoning effectively about the presented information.

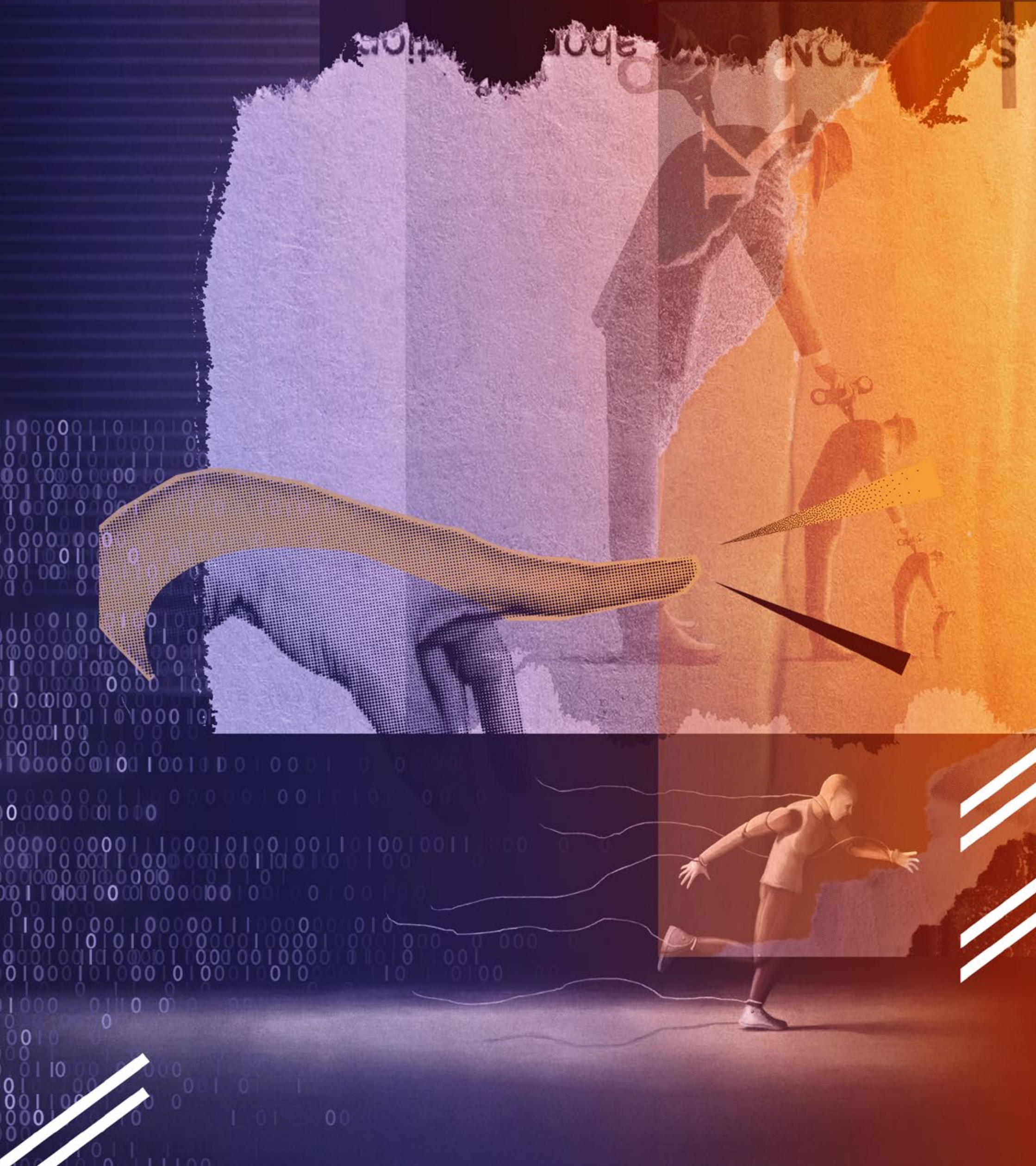
From the decision-making perspective, and particularly in the military domain, interdisciplinary approach to wargaming will help with design and test of theories and models for symbiotic human-machine decision-making processes. For military analysts, development of effective AI partner tools can augment situational awareness while providing actionable insights. Human-oriented interfaces and semantic communication will feature in intelligent systems that collect, analyse and disseminate information, while generating multi-modal content for communication.

Historical Example

One of the effective influence strategies in the Ukraine-Russia conflict has been amplification of particular events with a strong emotional component, such as the defiant stand of Ukranian troops at Snake Island in Feb 2022. The audio recording of the exchange between Ukraine's small military continent and Russia's Black Sea Fleet went viral on social media, garnering millions of views. This approach on behalf of Ukraine has played a big part in securing international support and military assistance of other nations.[15]

[15] Grobarcik (2023)

AUTOMATION AND AI IN THE CYBER DOMAIN



> *'Given that machine learning technology is becoming a library that people can download and use, or a service that people can use, what might future attacks look like when attackers are able to automate the generation of their attacks using, particularly, generative AI?'*
> – SME interviewee

Key terms and concepts[16]

The cyber domain has seen progression from simple automation (e.g. with use of bots for low-tech tasks) to development of complex autonomous systems. The key distinction lies in the latter's capacity for learning and a degree of independent decision-making.

AI and automation is also the topic of interest in cyber operations, particularly in assisting with detection and response to cyber-attacks, as well as rapid data analysis.

An important concept for discussing the effects of automation and AI in the cyber domain is their impact on the OODA loop (Observe-Orient-Decide-Act): a four-step approach to decision-making that encompasses filtering of data input, putting it in context and reaching a decision for action.

Emerging Trends

The next ten years will see a significant shift away from manual human processes in the cyber domain toward automated, AI-enabled systems or other rule-based approaches. Both cyber-attacks and cyber-defence increasingly leverage automation and AI, including creating and testing relevant code.

[16] Ebbott et al., (2021), DARPA (n.d.), Masakowsky & Blatny (2023)

Disruptive Effects

'Things are getting faster, bigger, scarier.' – SME interviewee

Automation and AI increases access, reach and scale of cyber-attacks. The intensity and frequency of attacks will grow. Critical infrastructure, transportation systems, communication networks, utilities, health services, government services and military systems can all come under attack from state and non-state actors, individuals and terrorist groups. Economic markets can and have been manipulated. These activities can cause widespread disruption, physical damage, loss of sensitive information and negative psychological impacts on the population.

At the same time, the use of AI systems can compromise the security of information submitted by the users, as was the case when Samsung developers released classified code by entering it into ChatGPT in 2023[17]. Corporations putting their trust in AI for monitoring and service security may not be accounting for vulnerabilities in the AI models. People creating software by compiling packages available online may be introducing vulnerabilities.

The effect of automation and AI on the speed of decision-making is important in the context of Defence and national security. There is a risk of falling behind the adversary in the speed of the OODA Loop and the robustness of decision-making if effective countermeasures are not implemented and if the technological edge is not maintained. Getting inside the adversary's decision loop requires faster decision process, and automation and autonomy are the key.

Historical Example

One of the earlier examples of the use of automated systems for influence can be gleaned in tactics of the Russian Internet Research Agency (IRA). It is estimated that during the 2016 US elections, IRA used between 36,000-50,000 bots to amplify specific messages online, creating the impression of a groundswell support for particular views.[18]

[17] Ray (2023)
[18] Ebbott et al. (2021)

Opportunities

‘Human effort doesn’t scale. We need to be investing in automation.’
– SME interviewee

Law enforcement, government and Defence will need to prepare for countering hostile actors using automation and AI in the cyber domain. Building secure systems and cyber-defences can reduce the number of attacks, improve detection and strengthen response. This requires incentives and investments to create technologies that are secure by design, rather than transferring this responsibility onto the end user. It also requires ability to measure the security of systems through formal metrics.

Countering automated systems requires an understanding of how these systems work and where they fail. Leveraging secure, trustworthy automation and AI in the cyber domain can help detect suspicious activity, understand vulnerabilities and prioritise response at scale, alleviating the skills shortages in the cybersecurity field. Areas of opportunity for automation and AI include:

- Data governance systems for determining what data organisations hold, its sensitivity, where it is being stored and how it is being used, as well as protection and deletion of unused data
- Developing systems that shape adversary decision-making processes and actions, directing them to parts of the network where their identity and activity can be tracked
- Setting up systems for quick access to and dissemination of trustworthy information
- Establishing sovereign servers and computation infrastructure for development of in-house AI systems, supporting various research and development (R&D) projects, and ensuring the security of information relevant to Defence and national security.

TOOLS FOR MANAGING THE INFORMATION ENVIRONMENT

'We need a generative AI test range in the same way we have cyber test ranges and missile test ranges.' – SME interviewee

Key terms and concepts[19]

Detecting, measuring and countering influence requires an understanding of the effects of multiple variables. Today, data analytics and social network analytics provide new tools for analysing the impact and effectiveness of influence campaigns.

New theories, models, and combination of both qualitative and quantitative metrics measured over time will improve situational awareness and ability to monitor the impact and effects of influence campaigns and countermeasures.

One of the emerging tools is simulation technology that can create complex, data-intensive environments that emulate real-world operations or events with use of rich, dynamic models. For example, computational cognitive science, which leverages mathematical modelling, can be used to develop a simulated environment for understanding and predicting cognitive behaviour.

Different types of simulation and AI systems can also be integrated within military wargames to help develop warfighting concepts, train commanders and analysts, explore scenarios, and assess force options.

Emerging Trends

Current research programs explore the potential of models, tools and data-sets for management of information environment: detecting, mapping and predicting narratives, detecting and attributing fake content (e.g. cloned voice), getting early warning of disinformation campaigns, linking online activity to real life effects, discrediting the sources, understanding and taking control of the narrative, improving transparency of information and reducing corruption.

[19] Ebbott et al. (2021), Masakowski & Blatny (2023), EDTAS Insights Paper SME interviews (2023)

Disruptive Effects

Even before the advent of automation and AI in the cyber domain, the ability to generate mis/disinformation has far exceeded the ability to detect, attribute and prevent it. This has been partly due to the lack of tools that could assist with these functions and regulate cyber activity in real time.

Development of formal tools for management of information environment is now the focus of competitive R&D across the world – it is an arms race in managing and exploiting the information environment. Failure to invest in such tools creates a risk of falling behind in the battle for control of narrative. An added challenge is that as mechanisms for detection of nefarious online activity evolve, so do the techniques for counteracting and evading them.

The clandestine and protean nature of cyber activity in different countries adds to the challenge of building accurate simulations that would enable strategic planning in the information environment.

Historical Example

The Russian IRA used off-the-shelf software and tools as well as digital marketing metrics to track public interests and opinions, combining qualitative and quantitative analysis. These early formal tools for managing the information environment helped the organisation to identify and target specific audiences, and amplify cultural and political divisions.[20]

[20] Ebbott et al. (2021)

Opportunities

Managing the information environment requires a set of models, metrics and tools that would improve the sensemaking and situational awareness and safeguard the decision making process. These kinds of tools are important to Defence and national security, law enforcement and commercial sectors. Such broad applicability creates a massive global market for these technologies, which is a strong financial incentive for development.

Formal mathematical techniques and metrics offer a structured approach for managing the information environment, ensuring security and reliability of own systems, assisting with detection of influence campaigns, mitigating deception and mis/disinformation, and countering dissemination of propaganda. Formal methods are more reliable than trial and error testing as they provide proof rather than guesses. Specific research directions that present areas of opportunity include the following:

- Developing processes and tools for authentication of content
- Establishing cybersecurity safeguards that check online traffic and generate alerts
- Building systems for measuring bias in online information and automating the flagging of bias in content
- Creating approaches for detecting background coordination of influence campaigns, such as can be seen when unrelated accounts start to coordinate their actions
- Using specific tools for checking reliability of systems
- Operationalising psycholinguistics through data analysis, e.g. in measuring similarity of users through their language
- Developing predictive techniques that analyse online communications, look for patterns in information spread over time and examine how social groups react to different types of content.

Simulation, experimentation and wargaming present another promising approach in management of the information environment, with capacity to incorporate technological, human and social elements and environments for testing different approaches and strategies. Existing research projects aim to build simulation environments for examining social media and influence campaigns. This type of simulation would provide a safe and cost-effective environment for training, planning and testing and refining strategies without revealing capabilities or intentions to the adversary. Leveraging simulation, experimentation and wargaming towards understanding of the information environment is an interdisciplinary endeavour; it requires behavioural scientists to assist with design on experiments and assess human interactions, sociologists to consider changes in norms, and legal experts to contextualise strategies within regulatory frameworks.

Military wargaming now needs to step beyond the constraints and objectives of traditional wargaming to incorporate AI, cyber operations, and political, social and individual effects. Fidelity of military experimentation can be improved by examination of real-life case studies, such as those presented by the Ukraine-Russia conflict. Apart from leveraging technologies such as LLMs to generate scenarios at scale, wargaming will benefit from incorporating human-machine teaming models in the decision-making loop.

Additional themes

The SME discussions touched on several themes additional to the five topics described in this Insights Paper, which present opportunities for dedicated research effort in their own right:

- **Autonomous systems in complex environments.** Convergence of robotics, automation and AI has meant that sophisticated automated and autonomous systems are now entering the real world: self-driving cars, robots, drones and various co-pilot systems. There is a growing acceptance of these systems, particularly in replacing humans for dangerous and repetitive tasks. These systems are becoming better at handling complex and varying environments with fewer data inputs. However, in both military and civilian domains, proliferation of autonomous systems requires better understanding of machine decision-making under conditions of uncertainty.
- **Advances in computation and network technologies.** A number of emerging technologies will enable new capabilities in the information environment. 6G networks will dramatically increase the scale and speed of data transfer. Novel efficient computing approaches such as neuromorphic computing will enhance edge processing, AI capacity for learning and adaptation and rapid response.
- **Quantum computing.** Quantum computing, in particular, will dramatically impact the decision-making loop and may shape competition in the cyber domain in the future. Quantum computing algorithms will affect the security of some classical encryption systems that are widely used today. For both Defence and national security, the initial challenge lies in developing verifiable post-quantum encryption protocols.

Australian Perspective

'If we're going to really develop sovereign capability and make those innovative leaps, we're absolutely going to have to be taking more risks and expecting some failures along the way and accepting that as part of the strategy.' – SME interviewee

Table 1 outlines the key challenges as well as strategies for positive growth in the Australian context that have been put forward in the SME consultations. The table is structured in terms of social, technological, environmental, economic, political, legal, ethical and military aspects. Several consistent recommendations comprise:

- Establishment of sovereign capability in cybersecurity and AI
- Development of pathways for interdisciplinary research
- Support for R&D ecosystems, fundamental science and long-term programs
- Improving capacity for commercialising and operationalising innovations.

Table 1. Key challenges and strategies for the future in the Australian context

Key Challenges	Strategies for the Future
Social	
Mis/disinformation will impact social cohesion, integrity of public discourse and trust in institutions. The challenge for Australian government will be in balancing regulatory intervention in the information environment with the principles of individual privacy and freedom of speech. The lack of broad technological and cyber literacy often reduces people's confidence with use of online systems and may leave some groups vulnerable to cyber-attacks and influence. At the same time, there is a lack of cybersecurity awareness and culture in many commercial organisations, with few economic incentives that would encourage transparency around cyber-attacks.	Supporting the R&D efforts to model human interactions in the information environment will help enact strategies that protect the national infrastructure and counter mis/disinformation online. Education, training and awareness campaigns will improve public resilience to cyber-attacks and online influence and enable appropriate response in the event of large-scale attacks. Continuous upskilling of technical experts will help meet the emerging challenges such as automation of cyber-attacks. Additionally, the onus should be placed on developers to build secure technological platforms and on data-collecting organisations to protect and manage this data in an ethical manner.

Key Challenges	Strategies for the Future
Technological	
Despite several areas of excellence and high calibre of researchers, Australia, as a nation, does not currently have a strong sovereign capability in cybersecurity and AI. This is due to lack of central coordination, established R&D ecosystems, targeted education pathways and long-term funding systems, which often results in loss of experts to other countries. There are very low numbers of Australian PhD students in cybersecurity compared with other nations. Technology focus has led to few education pathways for interdisciplinary research and projects. The value of social and behavioural sciences in management of information environment often goes unrecognised. Sophisticated infrastructure for computation and AI research is largely sourced from overseas. Australia lacks the domestic capability for development and manufacture of key infrastructure in communications and computation that would allow creation of systems like ChatGPT.	The Australian Government can lead the establishment of sovereign capability in cybersecurity and AI through several strategies: + Defining the high-level plan and the desired end-state for sovereign capability + Creating formal structures for integration of academia, industry and Defence + Establishing interdisciplinary initiatives that bring together technology experts, social and behavioural scientists, and legal and political SMEs + Protecting specific sectors and critical infrastructure that underpin the pathway to a digitally secure and resilient nation, and + Assessing data requirements and opportunities for international partnerships Long-term funding structures will foster R&D ecosystems and development of fundamental science, whereas agile approach and acceptance of risk will expand the national capacity for innovation, experimentation and adoption of new technologies. Education pathways can incorporate tailored and flexible cybersecurity programs with larger numbers of students and faster intakes, as well as strategies for recruiting talent from overseas. Australia needs to review and secure the supply chains for critical components and infrastructure and apply a security-focused approach to integration of hardware and software, while balancing sovereign production with opportunities presented by trusted partnerships.

Key Challenges	Strategies for the Future
Environmental	
Development of large AI models is energy-intensive due to use of servers, cooling systems, and other auxiliary equipment. This can impact the environment through increase in carbon emissions. Another challenge common across the ICT ecosystem is generation of electronic waste (e-waste) as the swift pace of technological advancement necessitates regular updates and replacement of hardware.	E-waste management strategies will become increasingly important in Australia and in other countries. Development of more efficient computing approaches will reduce energy requirements and associated emissions.
Economic	
In Australia, novel research and innovative ideas, including in cybersecurity and AI, can be difficult to commercialise and operationalise. There is often a large gap between academic research/prototype development and conversion into commercial product or deployed capability.	Transition of novel research into commercial and deployed systems requires stable, long-term funding for applied science projects that would cover the entire pathway from scientific discovery through to technology maturation, implementation, and education of users.
Low tolerance for failure in project acquisition leads to risk-averse culture and permeates through to early R&D projects, stifling innovation. The preference is to buy or import technology, which reduces domestic expertise. Industry investment in R&D is relatively low in Australia compared with other countries.	Risk management approaches need to allow for failure as part of the innovation process. Australia will benefit from economic and regulatory structures that foster a culture of information sharing and transparency in the cybersecurity space, enabling rapid response and learning.

Key Challenges	Strategies for the Future
Political	
Many nations, including Australia, are examining the growing power and influence of large technology companies and effects of new technologies on social cohesion. The debate about management of the information environment is sometimes impacted by the lack of technical insight at the political level, in part due to the failure of the S&T communities to communicate effectively with the decision-making authorities.	Political support needs to focus on development of sovereign capability in cybersecurity and AI, establishment of enabling national infrastructure, and regulation of data management practices by private firms. Enforcement of policy will require additional human resources. The S&T community can benefit from effective communication strategies and formal engagement mechanisms for conveying key information to the decision-making authorities.
Legal	
Access to new online systems such as generative AI transcends national borders, making it difficult to regulate locally. Enforcing laws is constrained by jurisdictional issues and the largely anonymous nature of the online world. Data collection and mass surveillance encroaches on individual privacy and freedom of speech; it may be enacted by third party commercial organisations and foreign nation states. IP laws can create obstacles to information sharing and partnerships in a commercially competitive environment.	Australia's resilience to influence campaigns will benefit from participation in the establishment of international laws and norms, supported by robust domestic legislation. Legal frameworks will be required for assigning responsibility for the output of AI models, tagging of synthetic content, and conducting cybersecurity research. New structures (e.g. DARPA-like) will be required for projects where IP is retained by the Commonwealth and is shared for R&D purposes.

Key Challenges	Strategies for the Future
Ethical	
The use of online data for research and monitoring presents ethical challenges for governments and researchers, broaching questions around the rights to privacy and norms of online behaviour. One of the ethical challenges in the battle of narratives is balancing communication and persuasion vs manipulation and coercion. In the Australian context, this goes to the need to safeguard the integrity of the democratic processes and the autonomy of our citizens. AI systems are known to have inbuilt bias arising from training data, which introduces risks when using these systems to support government functions and research.	Australia needs to support trusted sources of information and rigorous fact checking mechanisms that with the broader context of information and not only with single statements. At a broader level, Australia needs a national concept of truth and trusted facts, as well as collective agreement on the ways of ascertaining the truth. Measuring accuracy in reporting and public dialogue will become increasingly important. Balanced strategies are required for mitigating bias in AI systems and to ensure that Australian government acts lawfully, with transparency, and in the way that aligns with the liberal democratic values. AI systems should be built with ethical constraints.
Military	
A key constraint for developing sovereign cybersecurity capability is the classification of work and the requirement for security clearances. It is largely impractical for all researchers in the field to hold security clearances, particularly with the large cohorts of international students. Classification of work also reduces Australian researchers' contribution to the international body of work. At the same time, Defence competes with the private sector for talent.	Proactive development of cybersecurity capability in Defence requires investment in research, training and wargaming approaches that incorporate cyber operations and AI. The issue of security clearances can benefit from examining which parts of research need to be classified, where it may be possible to use proxy problems, and what information may be declassified to enable further research. Economic incentives and clear career paths will support Australian students entering the cybersecurity research field. Funding is also needed to develop fundamental science and sovereign critical infrastructure.

Outlook

Areas of Opportunity

The main areas of opportunity for Australia going forward lie in establishment of interdisciplinary projects, Defence partnerships, and novel education pathways formed with industry participation. New capabilities will also draw on international partnerships.

Interdisciplinary projects, including development of tools and simulated environments, require participation from computer sciences and AI researchers as well as social and behavioural scientists and related disciplines. These will support the study of cybersecurity, influence and human interactions online, with the view of developing sophisticated models connecting individual and group behaviours.

Defence has an opportunity to influence development of sovereign capability in cybersecurity and AI, and to shape teams that bring together military practitioners, academia and industry. Developing a centre of excellence in information warfare and influence will support the capacity for detecting, defending against, and responding to cyber-attacks and influence campaigns; it will improve situational awareness and strengthen the decision-making process. Cybersecurity and information warfare research requires allocation of classified and unclassified work, translation of novel research into classified areas by organisations such as DSTG, and maintaining operational relevance. This work can be supported through formal mechanisms for managing relationships and coordinating overlapping objectives of different organisations.

Australia has academic areas of expertise in AI research, computer vision, simulation, and ethics of AI use. However, the increasingly interdisciplinary and applied nature of research in cybersecurity, information warfare and influence requires central coordination and flexibility in design of university degrees and career paths.

The Australian industrial complex often holds an advantage in quicker development times, lower bureaucratic burden and solution-oriented processes. Australia hosts many innovative and successful technology start-ups. Industry can assist universities in building new degrees, developing pipelines of expertise that are relevant to current problems and that address emerging skills shortages. This process will be enabled by clear definitions, concepts and priorities in information warfare and influence, and by support for commercialisation pathways.

DSTG's 'More, together' strategy highlights the importance of collaboration with like-minded international partners who face similar challenges. There are strong centers of expertise in allied countries, including the UK National Cybersecurity Centre[21], the US Defence Advanced Research Projects Agency (DARPA)[22] and others. NATO has developed a Centre of Excellence in Strategic Communications to support the collective understanding, harmonisation and enhanced training and education in the various disciplines[23]. International partnerships will build a shared capability in cybersecurity and AI, and become invaluable resources for information-sharing.

[21] National Cyber Security Centre (n.d.)
[22] Defense Advanced Research Projects Agency (n.d.)
[23] NATO (n.d.)

2023-2024 EDTAS Campaign Going Forward

The DSTG Information Warfare STaR Shot aims for the future research into information and influence to incorporate the following aspects:

- Development of effective mechanisms for interdisciplinary research
- Understanding of the fundamental science of influence, and not only the technological effects
- Examining the impacts of convergence and synchronisation of effects and modalities in the information environment and beyond
- Developing expertise in different types of information environments, which may be bounded by national, cultural, religious or linguistic parameters, as well as varying access to technologies

These principles, alongside the five key themes described in this Insights Paper, inform the objectives and the design of the EDTAS events going forward. The Information and Influence Symposium will bring together experts from academia, industry and Defence to explore the key areas of risk and opportunity for Australia and to chart the way forward.

References

Australian Government Department of Industry, Science and Resources (n.d.). Australia's Artificial Intelligence Ethics Framework. Retrieved from https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles.

DARPA (n.d.). Cyber Grand Challenge (CGC). Retrieved from https://www.darpa.mil/program/cyber-grand-challenge.

Defence Science and Technology Group (n.d.)a. Emerging and Disruptive Technologies Assessment Symposium (EDTAS). Retrieved from https://www.dst.defence.gov.au/NextGenTechFund/emerging-disruptive-technology-assessment-symposium-edtas.

Defence Science and Technology Group (n.d.)b. More, together: Defence Science and Technology Strategy 2030. Retrieved from https://www.dst.defence.gov.au/strategy.

Defense Advanced Research Projects Agency (n.d.). Creating breakthrough technologies and capabilities for national security. Retrieved from https://www.darpa.mil/.

Ebbott, E., Saletta, M., Stearne, R., Webb, M., Dowling, M.-E., Farina, M., Young, G. & Job, P. (August 2021). Understanding Mass Influence. Three case studies of contemporary mass influence platforms and campaigns. Produced for the Department of Defence by: The University of Adelaide, The University of Melbourne, University of New South Wales, Edith Cowan University and Macquarie University.

Goldstein, J.A., Sastry, G., Musser, M., DiResta, R., Gentzel, M. & Sedova, K. (2023). Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. Stanford University.

Grobarcik, D. (2023). Snakes, stamps and javelins: how Ukranian information and influence operations brought the fight to Russia. Irregular Warfare Inititative. Retrieved fromhttps://irregularwarfare.org/articles/snakes-stamps-and-javelins-how-ukrainian-information-and-influence-operations-brought-the-fight-to-russia/.

Latimer, J. (2001). Deception in War. New York: Overlook Press.

Masakowski, Y.R. & Blatny, J.M. (Eds) (2023). Mitigating and Responding to Cognitive Warfare. NATO STO-TR-HFM-ET-356.

National Cyber Security Centre (n.d.). The National Cyber Security Centre. Helping to make the UK the safest place to live and work online. Retrieved from https://www.ncsc.gov.uk/.

NATO Strategic Communications Centre of Excellence (n.d.). Retrieved from https://stratcomcoe.org/.

Ray, S. (02 May 2023). Samsung bans ChatGPT among employees after sensitive code leak. Forbes. Retrieved from https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/?sh=6ab278b96078.

Rosenberg, L. (2023) The Metaverse and conversational AI as a threat vector for targeted influence. In: 2023 IEEE 13th Computing and Communication Workshop and Conference (ICCWC), Las Vegas, NV, pp. 0504-0510.

Wakefield, J. (2022). Deepfake presidents used in Russia-Ukraine war. BBC News. Retrieved from https://www.bbc.com/news/technology-60780142.

Zucchino, D. (2009). Fighting Afghan information war. Los Angeles Times. Retrieved from https://www.latimes.com/archives/la-xpm-2009-jun-11-fg-afghan-information11-story.html.

Glossary

AI	Artificial intelligence
ASCA	Advanced Strategic Capabilities Accelerator
C2	Command and control
ChatGPT	Chat Generative Pre-trained Transformer
DARPA	US Defense Advanced Research Projects Agency
Disinformation	Intentionally inaccurate information
DSR	Defence Strategic Review
DSTG	Defence Science and Technology Group
EDTAS	Emerging Disruptive Technology Assessment Symposium
GAN	Generative Adversarial Networks
Generative AI	Artificial intelligence system that can generate content
IRA	Russian Internet Research Agency
ISR	Intelligence, surveillance and reconnaissance
LLM	Large language model
MFM	Multimodal foundation model (a type of generative AI)
Misinformation	Unintentionally inaccurate information
NATO	North Atlantic Treaty Organization
OODA loop	Observe-Orient-Decide-Act loop
R&D	Research and development
S&T	Science and technology
SME	Subject matter expert

Notes