www.rusi.org

UKRI Trustworthy Autonomous Systems Hub

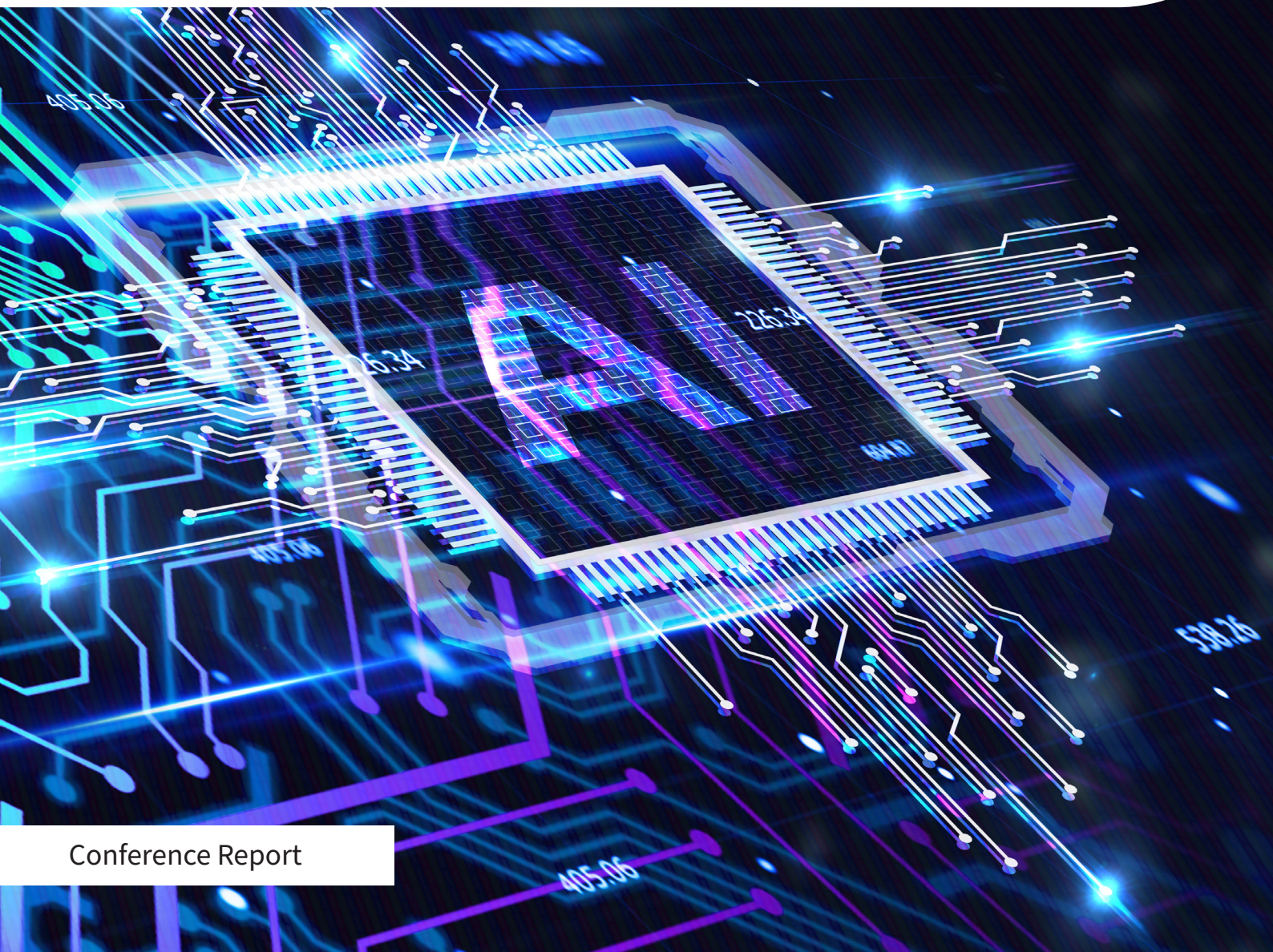
RUSI

Conference Report

Using AI in an Intelligence Context

Future Scenario Workshop

Pia Hüsch



Conference Report

193 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 193 years.

Published in 2024 by the Royal United Services Institute for Defence and Security Studies.

RUSI Conference Report, May 2024.

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org
RUSI is a registered charity (No. 210639)

Acknowledgements

The team of the RUSI Disruptive Technologies Programme is grateful to the Trustworthy Autonomous Systems (TAS) Hub for providing funding for this workshop and report and for their support throughout the research process. A great deal of thanks must go to the team that helped to guide and shape the workshop and report writing. Emma De Angelis, Hugh Oberlander, Michael Boniface, Steven Meers, Ino Terzi, Stuart Middleton, Edward Mortimer and Zenab Hotelwala all provided valuable support, input and editing.

Executive Summary

This workshop report presents a summary of observations and discussions related to an event on the use of AI in intelligence by 2040, held in February 2024. The future scenario workshop was held to examine the use and implications of AI in a hypothetical 2040. It was based on a hypothetical scenario in which the relationship between a technically advanced economic and military power called Roland and the neighbouring island state Islay deteriorates. It unfolded in three parts:

1. The build-up of military exercises by Roland near Islay and the potential for invasion.
2. Roland's full-scale invasion of Islay.
3. Roland establishing civilian authority over Islay.

The workshop was structured around the scenario phases. The following are key observations that participants made in the context of each scenario phase but have relevance to the broader question of what AI use in an intelligence context will look like by 2040 and what implications and questions follow.

Part 1: Pre-Invasion

- The adversary's widespread AI integration in military structures likely constitutes a **strategic advantage**, increasing accuracy and efficiency, especially for data analysis.
- However, AI **dependency and integration also constitutes an additional risk** where access can be achieved to the adversary's systems and data, for example for information on the adversary's movements of troops and equipment.
- The **quality, not just the quantity, of data is decisive** for intelligence purposes. Authenticating information is therefore a key capability.

- **Communication and assessment of information** is likely challenging in this scenario, including the communication of accurate information and analysing available information at fast pace.
- It **is unclear how the UK achieves deterrence** in this context or what deterrence looks like in a zero-trust, AI-fuelled environment. This relates to both the communication of information and the sharing of capabilities for deterrence purposes.

Part 2: Full-Scale Invasion

- AI can be used to **design efficient evacuations**, for example to identify British individuals and their next of kin, or to design and optimise exit plans.
- **UK support to allies** includes sharing intelligence capabilities and assisting integration and coordination of datasets.
- To be prepared for a full-scale invasion scenario, the UK government needs to invest in and support the development of key enablers, **such as computing power, communication infrastructure and critical technologies**.
- The UK government also needs more **credible technology leadership** to implement such aims.
- **Procurement** needs to become more agile to enable shorter innovation cycles.

Part 3: Civilian Authority Established

- **Information operations** may be less effective in a context of information chaos.
- There is a significant need for **authentication methods**.
- It is important that both AI systems and humans are trained in **cultural comprehension**. At the same time, AI can help to understand culture, history and their lessons.

Participants further assessed wider implications, such as the need for the UK to expand AI training for all sectors and skills levels, and identified further research questions based on these observations.

Introduction

This workshop report presents a summary of observations and discussions related to an event on the use of AI in intelligence by 2040, held in February 2024. The future scenario workshop was held to examine the use and implications of AI in a hypothetical 2040. The report highlights two major aims of the workshop.

First, by using a future scenario, it aimed to contribute to the discussion by setting out how the use of AI might enable a profoundly different approach to intelligence by the year 2040. Second, it aimed to improve UK preparedness for this scenario by informing questions for future research, as well as potential challenges, implications and milestones that need to be reached, both for such a scenario to come about, and to enable mitigation of any risks that could arise as a result of such developments.

The hypothetical scenario concerns the deteriorating relationship between a technically advanced economic and military power called Roland and the neighbouring island state Islay. The scenario progresses in three parts: the build-up of military exercises by Roland near Islay and the potential for invasion (Part 1); Roland's full-scale invasion of Islay (Part 2); and Roland establishing civilian authority over Islay (Part 3).

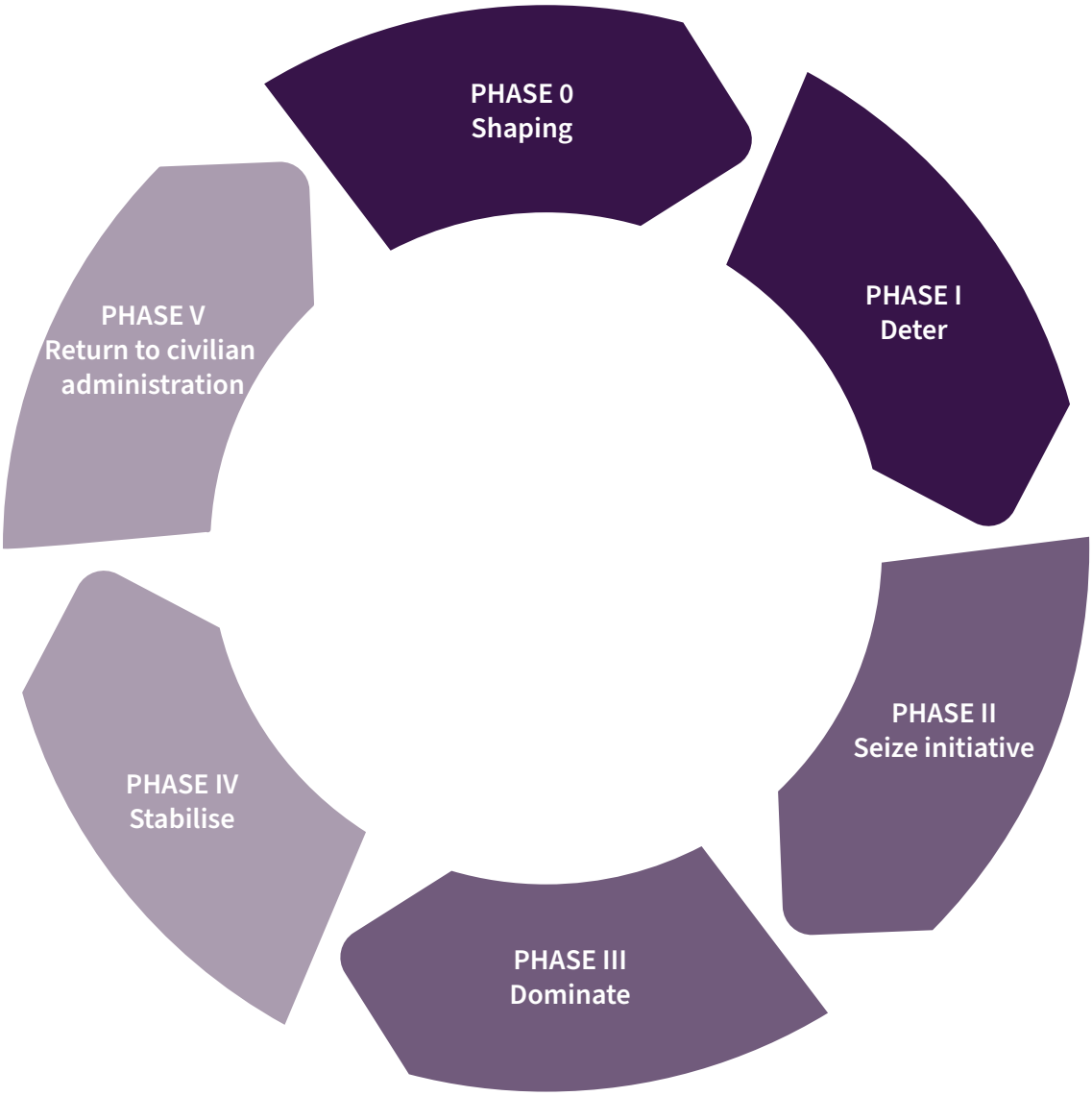
Workshop Methodology

This report is based on a two-part workshop with practitioners, representatives from the public and private sectors, and academics. The workshop was co-organised by the Trustworthy Autonomous Systems (TAS) Hub and RUSI.

In the first part of the workshop, participants discussed a hypothetical future scenario illustrating the use of AI in a 2040 intelligence context. They received several prompts for a future scenario (listed below as Parts 1–3). These mirrored the six phases of conflict to structure the discussion.[1] The scenario evolved in three parts, each representing two phases of the conflict cycle (colour-coded below on Figure A). For the second part of the workshop, participants were split into breakout groups reflecting the stages of the scenario.

1. Paul Scharre, American Strategy and the Six Phases of Grief', *War on the Rocks*, 6 October 2016, <https://warontherocks.com/2016/10/american-strategy-and-the-six-phases-of-grief/>, accessed 30 April 2024.

Figure 1: Six Phases of Conflict, Grouped in Line with Workshop Discussions



Source: Author generated.

Limitations

Workshop discussions were held at an unclassified level. Access to information, particularly in an intelligence context, and participants' ability to share their knowledge, was therefore, at times, limited.

The scenario is hypothetical, and participants did not receive background information on Roland's or Islay's capabilities or additional context. Given the lack of details about the cultural, historical or political context of Roland, discussion remained speculative at times.

Structure

This report's structure follows the three phases of the fictional scenario. Each section of the report introduces the specific setting of each part of the scenario. It then offers reflections from the plenary and the respective breakout groups that were charged with examining each part of the scenario in further depth. Finally, a fourth section addresses wider implications that follow from the overall scenario, including legal, ethical and workforce implications.

The Scenario

The information boxes in each of the following subsections describe the initial scenario input participants received. The boxes are divided into three progressive phases of the escalation and are followed by the participants' reaction to each.

Scenario Part 1

Description

Over the next decade and a half, a large and economically powerful adversary – a state called Roland – expands its technological advantage in AI and uses its technological and economic power to advance its geostrategic ambitions. Through a combination of industrial cyber espionage operations, innovation, widespread exports of its technologies and internal implementation strategies, the country manages to roll out the wide use of AI, including in its defence, intelligence and police sector. Roland has access to some of the most advanced AI systems. Roland has different cultural and ethical norms from the UK and is willing to use AI technologies in ways the UK would not.

It is now 2040. Over the past few months, Roland has repeatedly held large-scale military exercises. It is now feared that Roland will invade a neighbouring country, Islay, an island state. Out of fear of such invasion, Islay has forged close bonds with regional allies and NATO countries seeking to deter conflict.

Figure 2: Lead-up to Invasion

Source: Author generated.

AI integration as a (dis)advantage? Participants considered whether Roland's widespread integration of AI is a strength or a weakness. Initial assumptions implied that widespread AI integration would lead to greater efficiency and high technological advancement. However, there was disagreement and uncertainty about what AI-enabled weapons and systems would be available by 2040, and whether they might introduce new vulnerabilities. Participants reflected that, while in the medium to long term, military adoption of AI is likely to confer strategic advantage, early adopters may inadvertently introduce new, difficult-to-predict weaknesses. Such weaknesses can be exploited by technologically adept adversaries.

Wider AI integration could therefore also constitute a weakness if access to Roland's datasets can be achieved. For example, if Roland is widely using AI systems in its military operations, it would also do so for its troop mobilisation and exercises. If access to such data can be secured, Islay and its allies could accurately monitor equipment and troop movements. A comparison was drawn with the fitness data collected by the app Strava, popular among military personnel tracking their exercise. In 2018, data shared by athletes as part of a public social network revealed base locations and patrol routes.[2] Similar information can draw a telling picture for Islay and allied intelligence services if access to the adversary's data can be secured. In that sense, the widespread integration of AI could constitute a weakness for Roland as it possibly leads to new vulnerabilities. High standards of operational security are therefore likely to be an even more critical factor in future military conflict.

Communication and decision-making in a disinformation environment. Observations also addressed implications of sharing genuine information and

2. Jeremy Hsu, 'The Strava Heat Map and the End of Secrets', *Wired*, 29 January 2018.

disinformation. In view of expected domestic but also international disinformation campaigns, enabled by deepfakes and tailored to individuals, experts considered that there is a real risk that any information acquired might be manufactured. One expert described this scenario as an 'arms race for who can manufacture and identify fake information more effectively'. In such a context, communicating accurate information also poses a challenge. Technologies such as cryptographic watermarking to verify the provenance of authentic media such as governmental communications will be increasingly important. Experts were concerned about building communication bridges but also the speed at which decision-makers must rely on, assess and judge available information. It is expected that time pressure for data assessment and decision-making is likely to increase by 2040. This enhances the risk that decision-makers do not have sufficient time or inclination to assess the quality of data underlying their decision-making process. AI assistance could be helpful in this context, to address the scale and pace of this challenge, although it might limit the involvement of meaningful human oversight and understanding.

Quality of data. Participants anticipated that more data will be available by 2040, but expressed concern about the quality of data fuelling AI technologies and intelligence work more generally. How do you ensure the data collected or used to train your own AI systems is accurate and useful? Experts argued that the decentralised nature of the UK intelligence community can constitute an advantage, as different institutions would likely use different datasets, thus making them more resilient to false information and enable checks and balances. However, data analysts pointed out that any available data is only useful if it can be turned into information that is useful – a process that currently still requires human input, and that human input is unlikely to scale to meet the data needs in 2040. Testing the information pace prior to such a scenario is thus a key element of preparation. The participants also pointed out that by 2040 it is likely that a large proportion of data available for machine learning models may, in fact, be AI generated. The implications of this remain to be clarified.

How to communicate for deterrence. In this zero-trust context, disinformation and the validation or reliance on false information determine a pre-conflict scenario of the kind given here between Roland and Islay. This includes potential communication between conflict parties and their allies. A fundamental concern is how to deter the adversary without being misinterpreted. Even communication to de-escalate a situation may not be trusted. It is even unclear what deterrence looks like in such scenario, raising questions such as what kind of deterrence signals Islay, the UK and allies would want to send, and what channels to use to send them. Traditional deterrence relies on hardware such as nuclear missiles to signal capabilities and willingness to use them. For some experts, AI technology's deterrence is much more communication-based, with doubt and

need to interpret the adversary's capabilities remaining. The adversary would have to rely on its own intelligence to assess allied AI capabilities. Other participants considered that deterrence in this context remains of a physical nature, in that AI would be implemented in weapons and other systems. One expert added that AI enabled more opportunity for deterrence without putting humans at risk, for example where subsurface weapon systems could patrol the waters around Islay.

Capability sharing as deterrence. Similarly, it was unclear what kind of AI-enabled military capabilities the UK and other allies can provide to Islay to ensure deterrence. Some were sceptical about providing AI technologies to Islay which might be an ally in that context but not in the future. They drew attention to the risk that the UK and its allies might regret providing technologies that could be used against them in the future. This concern reflects experiences of the Cold War, when Western allies provided capabilities to states for deterrence against the Soviets, but which were later used against them. This was the case with training and weapons supplied to the Mujahedeen in Afghanistan, who were equipped by Western powers to withstand the Soviet Union but later turned against former supporters.[3] Providing capabilities to Islay may, however, also become easier with the expansion of AI technologies. One practitioner pointed out that in Ukraine, the delivery of capabilities is restricted due to concerns that weapons may be used for escalation.[4] AI systems, however, could be programmed to only be used for predefined purposes, eliminating the risk of escalation and enabling earlier capability sharing. A technologist supported this assessment, explaining that if Islay were able to circumvent the programmed restrictions by reverse engineering the delivered AI systems, it would have the necessary skills to make them in the first place. However, one participant pointed out that it might not be within allies' power to decide what AI technologies to share if they are increasingly open source or off-the-shelf technologies. In this scenario, the bar to diffusion may be quite low. The assumption that the UK and its allies can determine who has access to what technologies and capabilities might be a false premise, according to this expert.

The relationship between AI, deterrence and implications for (de-)escalation were identified as areas of great interest for future research.

AI to enhance cultural understanding. Socio-technical approaches that combine AI with behavioural and social science understanding to improve cultural understanding featured in many additional observations participants made on

3. Martin Beckford, 'National Archives: Britain Agreed Secret Deal to Back Mujahideen', *The Telegraph*, 30 December 2010.
4. For example, see Phil Stewart and Idrees Ali, 'U.S. and Ukraine Discuss Danger of Escalation as New Arms Extend Kyiv's reach', *Reuters*, 26 May 2022.

the scenario. AI was perceived as helpful in some instances but less so in others. For example, some considered that AI technologies could help understand adversaries and their culture, history and ethical norms, including their online security culture. AI's perceived role in fostering understanding of other cultures was considered particularly helpful given that it has been challenging for the UK and its allies to understand how the adversary thinks.

Back to human intelligence? Others, however, found that in this convoluted space of misinformation and big data, the importance of human intelligence stood out to provide situational awareness. One human task includes developing indicators and warnings that ensure the UK and its allies are not being deceived. While AI systems can help identify when these thresholds are met, turning these variables into programmable logic in the first place is a human task. This is especially challenging to develop when determining the adversary's intent to launch an invasion.

Against this backdrop, participants in the breakout group focusing on this scenario discussed three questions that they considered required further attention.

Figure 3: Key Questions for Breakout Group Participants

1. **What limitations should be imposed on AI's decision-making authority?**
While participants acknowledged AI's advantages in data processing and analysis, they also agreed there was a need to set limitations for AI's use and autonomy in decision-making processes. However, there was uncertainty as to what exactly those limitations could and should look like. What checks and balances do we create to structure the human–AI relationship? Human-centric AI systems were identified as an area of interest for future research as participants recognised that failures of human–machine interactions is one of the risk areas for AI adoption.

2. **How can it be ensured that AI is implemented in a dynamic fashion and algorithms regularly reviewed and updated?**
Participants voiced concern over the risk that once programmed and no longer updated, the gap between an AI system and an ever-evolving work and institutional culture can widen. The fear is that the more static the programming of AI is, the more likely it is that there will be an increasingly larger gap between the once programmed understanding of an organisation that is reflected in the AI systems and the continuously developing organisational culture that is reality, for example due to a change of structures, employee profles or organisational priorities. In contrast, humans interacting with one another can adjust and bridge differences in organisational culture. Experts therefore argued that it is important to find a way for AI to be dynamic. This is especially important for any elements in its coding that are inherently static which need to be continuously reviewed and updated to avoid bias. Such dynamic approach is also important for decentralised intelligence institutions such as those in the UK. Otherwise different intelligence institutions might find it hard 'speaking' to one another, since each of them will have its own built-in assumptions and biases.

3. **What are the different standards for predictive versus descriptive use of AI?**
Experts in this group pointed out that not all AI technologies and all use cases are the same or bear the same risks. For example, they drew a distinction between predictive and descriptive uses of AI. In this context, experts would like to see further clarification on what technical, ethical and regulatory standards apply to such use cases.

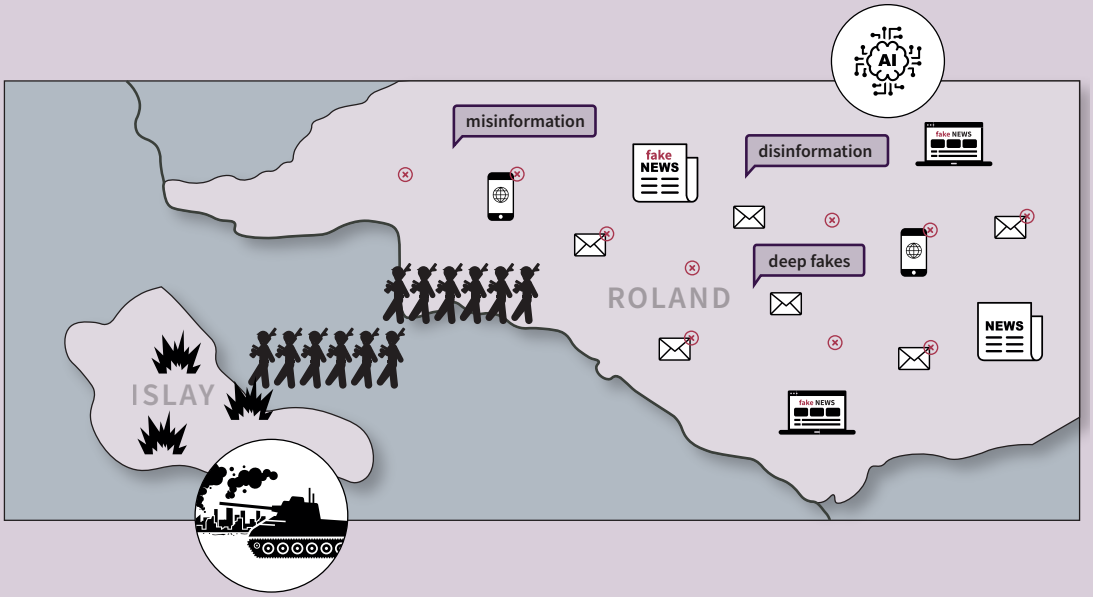
Source: Author generated.

Scenario Part 2

Description

A sophisticated propaganda campaign enabled by AI technologies pushes the message of a historical claim to Islay. A purported mass exercise concurrently masks a full-scale invasion of Islay.

Figure 4: Roland Invades



Source: Author generated.

AI to design efficient evacuations. In case of a full-scale invasion, the UK and other countries would seek to evacuate their citizens located in Roland. Experts argued that collected datasets and AI technologies to analyse them can help to identify not only UK citizens and their next of kin but also their locations. Furthermore, AI technologies can be used to identify and optimise exit plans. One participant suggested creating a foreign office evacuation app that leverages AI. An app such as this could disseminate targeted instructions to respective citizens, identifying their best route to evacuation and sending instructions about when to go and where. For example, it could identify an airfield, as well as provide packing instructions. A challenge here is that the app would depend on reliable communication systems – which may be switched off or disrupted in a conflict scenario. However, it was countered that current practices imply that communication systems are too valuable to both parties to a conflict. As a result, it may be unlikely that one party would attempt to switch them off completely. Another challenge was raised by a military practitioner, who pointed

out that UK citizens are not normally targetable abroad. In these circumstances, however, exactly that is required. The underlying AI systems would need to be sufficiently flexible and fed by appropriate data. This raises both questions about how and when this data is collected in the first place, and whether it is done so in a way that complies with GDPR and other privacy laws. A practitioner in the private sector saw this as less of a problem, based on recent experience of identifying UK citizens for evacuation in Sudan, where some of these capabilities were already implemented. The commercially available information they relied on was indeed GDPR compliant.

UK intelligence support for Islay to maintain situational awareness. Experts were generally positive about the ability to upskill the intelligence capability of UK allies based on recent experience. While it comes at the risk of losing some of the UK's control, upskilling Islay by handing over considerable capability of the UK and its allies would improve both Islay's and the UK's intelligence-gathering abilities. Data integration is currently still a challenge and takes weeks. By 2040, it will be vastly quicker. One data expert estimated timelines to integrate datasets could move from weeks to hours or days. One expert pointed out that in this scenario, a UK military response is unlikely, as Islay is not part of NATO or another military alliance. Instead, UK support likely plays out in other ways, such as coordination of classified and non-classified data resulting in shareable intelligence. Allies benefit from the UK's decentralised intelligence institutions, with data collected by one agency verified by another. Assured quality of data allows the UK and its allies to maintain situational awareness and informs other steps, for example the designing of particularly effective sanctions.

Risk of centralised technology capabilities. The concentration of power in the technology field was considered a risk in this scenario. While militaries retain some autonomy, much of technology capability is developed and implemented by the private sector. The concentration of power is particularly stark for AI technologies, where a few large technology companies hold considerably more knowledge and capability than the public sector. While currently, much of this power is in a few allied hands, this could change by 2040, particularly with the growing capability of open-source AI models. Much depends on the relationship between public and private sectors. Experts warned that it was prudent to assume that when a company resides in a specific country, its alliance also sits with said country in a scenario such as this. While commercial concentration is a risk that is difficult to address, possible geostrategic implications are considerable for this future scenario. While some steps have already been taken to diversify and de-risk supply chains, for example for semiconductors, experts were sceptical that these would be significantly decentralised by 2040.

A breakout group discussed these issues further, focusing on three requirements considered necessary to put the UK in a strong position by 2040.

Figure 5: Key Requirements Identified by Breakout Group in Response to Part 2

1. **Government needs to systematically invest in, and support development of, key enablers needed for this scenario.** Participants in this breakout group stressed the importance of several key technologies and related infrastructure that would be fundamental to a future scenario such as this one. These key enablers could include edge computing, communication infrastructure and semi-conductors, and the need for de-risked supply chains and production scalability. To identify which technologies are needed, participants suggested the simulation of battlespace and the surrounding environment, for example with the help of AI-generated digital twins. However, participants also stressed that while cooperation with the private sector is key, no organisation can achieve such technological advancement on its own. Instead, an ecosystem approach is needed to realise these aims. Here, the UK government should show leadership in setting the vision for technology through policy, and leverage its purchasing power to signal strategic priorities and stimulate emerging behaviour in the market.

2. **The UK needs credible technology leadership from government and within national security and defence structures.** Participants agreed on the need for stronger leadership from government, both in terms of setting a vision but also for defining smaller steps for implementation. If the aim is to become a 'global AI superpower' as announced in the UK National AI Strategy, participants would like to see clearer, practical milestones set by government. Setting this strategic intent is crucial, including for a timeline of the next 5–10 years. At the same time, participants acknowledged that a stronger vision for AI implementation also requires more technology leadership positions within government and defence structures. More leadership positions need to be filled by technicians, or those with a stronger understanding of technology. This requires both flexible promotion structures that do not overlook those with a more technology-focused profile, and the opportunity for technology experts from the private sector to enter civil service at a later stage and develop trust with colleagues.

3. **Procurement needs to become more agile to enable shorter innovation cycles.** Participants in this group voiced concern that contemporary, slow procurement processes are inadequate to secure a strategic advantage from key emerging technologies. A future scenario in 2040 was considered a short timeline in light of current procurement practices. Experts argued in favour of a more agile procurement process to allow for quicker innovation cycles – not just in war scenarios such as in Ukraine. Participants considered that the involvement of smaller, innovative companies should play a bigger role, but also acknowledged that such involvement requires a trade-off between newness of capability and the scalability of production. While scalability and cutting-edge capabilities were considered desirable, experts stressed these developments must also be able to guarantee the safety and the security of the product. For example, where a product lands in the adversary's hands, the individual components must be useless to the adversary and the system must possess a kill switch that can be remotely operated. Here, AI safety and security standards developed in the civilian ecosystem could also drive such developments in the defence context.

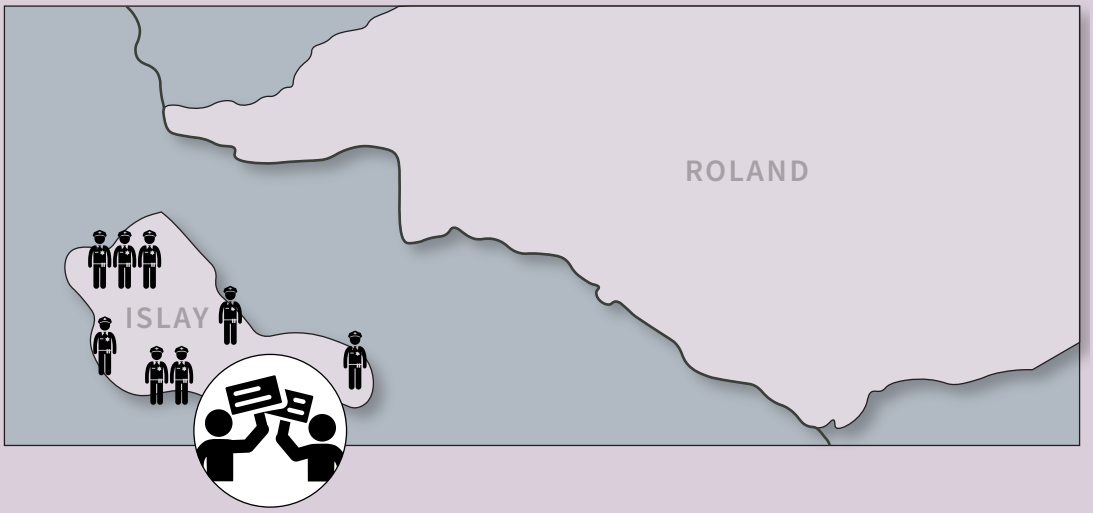
Source: Author generated.

Scenario Part 3

Description

After months of occupation, protests in Islay City, one of Islay's largest smart cities, are gathering momentum in support of resistance. Roland's regime seeks to suppress these protests.

Figure 6: Occupation of Islay



Source: Author generated.

Information chaos limiting impact of information operations. Experts were sceptical about the impact of information operations in this scenario, as large parts of the population may have become desensitised to fake news and other misinformation. It is therefore questionable whether propaganda campaigns by Roland would be impactful. On the flip side, even if Islay's resistance were to produce evidence of atrocities or other unlawful behaviour from occupying forces, there would be questions about whether the wider population would accept them as such, possibly hindering further resistance. One expert argued that if disinformation campaigns continued to proliferate, there would be a risk that, by 2040, people might have become jaded by the overload of genuine information and disinformation. Instead, they might end up believing what they want, closing themselves to genuine information and rational evidence.

Need for authentication methods. In this scenario, authentication methods become particularly important for reaching Islay's as well as Roland's population. One way to authenticate information is by watermarking it, for example to make fake news or AI-generated information easily identifiable. However, it was unclear who watermarks information and how effective watermarking is. Social

media platforms were considered to play a special role in this context – both as potential authorities to regulate or watermark information but also as platforms for intelligence competition, especially prior to an invasion. Another suggestion was that end-users' technology, such as smartphones, might provide a scoring of the information the user consumes to indicate its accuracy or potential fakeness – a valuable tool that governments might be interested in funding.

Wider technology and communication infrastructure. How the UK can support opposition to the occupation and reach Islay's population also depends on how general communication structures and related technologies will evolve to 2040. One expert referred to the possibility of bypassing government or commercially controlled networks, and the availability of quantum computing to bypass encryption or AI's capability to break encryption.

Against this backdrop, the breakout group discussion focused on the following three aspects:

Figure 7: Key Requirements Identified by Breakout Group in Response to Part 3

Source: Author generated.

Wider Implications

A fourth breakout group remarked the wider implications following from the scenario, particularly focusing on legal, ethical and workforce implications. Its three main takeaways are as follows.

1. **The UK needs to increase AI training for all sectors and skills levels.** Participants argued in favour of widespread AI training to develop the UK's AI literacy. This ensures that people are ready to use AI technologies in their everyday work. More specifically, it is necessary to train people in AI-related critical thinking skills and human–AI teaming methods. In a defence context, that might mean training intelligence analysts so they can prompt AI bots/tools in the right way and optimise results from AI in human–AI interactive engagements.
2. **The UK needs to develop its AI assurance capacity.** Participants suggested creating centres of excellence for AI model auditability and the quantification of AI model uncertainty and predictability. However, they argued that these should be managed by the government, regulators or academia. The involvement of AI vendors was viewed critically, due to a perceived conflict of interest. Importantly, assurance should take place prior to the deployment of AI technologies. This is not the case in current AI models: large language models, such as ChatGPT, are released fast and problems fixed later – an approach deemed undesirable by participants.
3. **The UK needs to actively build resilience into UK data and compute capacity for AI.** Participants agreed on the vital role of data needed to train AI systems. Dependency risks arise where much of this data is sourced from outside the UK. Data access might be contested in the future – for example in a scenario such as the one examined here. Participants therefore agreed that maintaining a UK data archive for AI training could become even more important. Furthermore, GPU clusters are needed to train AI models. Again, clusters outside the UK might be contested, so more local capacity – such as the UK national AI compute cluster in Bristol – is needed. Participants also agreed that there needs to be more UK sandboxing capacity to evaluate AI models, including to test models that are ethically challenging to UK's stance. Experts pointed out that competitors are often not as ethically constrained as the UK, but considered there is a need to ensure that the UK's AI models work effectively within and outside the country's current ethical stance. This would also allow AI models to operate if the UK's ethical stance changed in the future, particularly if there were an existential threat to the UK. This was considered necessary as it takes years to develop AI models.

Therefore, readiness – including sandboxing – needs to happen years before any such changes.

Conclusion

This report has set out the use case and implications of AI technologies in an intelligence context taking place in 2040. The hypothetical scenario underlying this analysis concerned the deteriorating relationship between the technically advanced economic and military power Roland and its neighbouring island state Islay. The scenario progressed in three parts: the build-up of military exercises by Roland near Islay and the potential for invasion (Part 1); Roland's full-scale invasion of Islay (Part 2); and Roland establishing civilian authority over Islay (Part 3).

Some of the observations made in this analysis are specific to a particular phase of the conflict. For example, the role of deterrence and potential ways in which the UK can communicate or demonstrate deterrence clearly relate to a pre-full-scale invasion scenario. Similarly, some specific use cases, such as the AI-enabled data analysis for evacuation purposes, are context specific and not equally relevant to all scenario stages.

The majority of observations about the UK's ability to secure a strategic advantage from AI technologies and how to implement and benefit from these technologies by 2040 applied across all stages of the scenario. Participants repeatedly returned to several core themes. From a human perspective, these included the need for more skills development at all levels and training in human–machine interactions, as well as the need for clear technology leadership among senior UK government officials. On the more technical side, there was uncertainty about the exact use cases and advances that could be expected by 2040. Similarly, questions remained as to the legal, ethical and technical restrictions that should be applied to any AI systems in use. However, there was wide agreement on the urgent need for stronger collaboration with the private sector and the wider technology and innovation ecosystem to harness AI's full capabilities. This also includes need for stronger governmental leadership on technology investment and development of AI and other related technologies, such as compute and future telecommunication infrastructure. Participants also repeatedly stressed the urgent need for a more agile procurement process to enable shorter innovation cycles.

The nature of the exercise and the limited information on the fictional states involved left room for speculation and uncertainty. However, this also allowed the conversation to identify interesting questions that can guide future research projects. These included:

- What does deterrence look like in an age of zero trust and disinformation enabled by AI?
- How can procurement become more agile to enable shorter innovation cycles and include more start-ups?
- What does it mean to be 'a global AI superpower'? How does this translate to areas such as compute, data capture, assurance regimes and effective test environments?
- How can the UK's aim of being a global AI superpower be translated into more concrete short- and medium-term goals?
- How can the UK fill the skills gap that arises at all levels, including at a leadership level where more technology expertise is needed to be able to set a concrete vision for UK technology leadership?
- How can the UK achieve effective human–machine teaming while setting adequate limitations (for ethical, technical and legal aspects) and implementing appropriate regulation?
- What data environment is required to effectively use AI-enabled intelligence capabilities by 2040? What does adequate data governance look like that effectively protects individual freedoms while allowing the UK to secure a strategic advantage through AI-enabled analysis? And what data does the adversary use to train its AI systems? Does it rely on data on its own past military practices? What are the implications?
- What is the historical and cultural context, and how can the UK use AI to understand this context?

Given the considerable amount of uncertainty and the fast-moving technological developments that underpin any analysis of the geopolitical and national security implications of AI in an intelligence context, further research into these areas is valuable to the public, private and third sectors of the UK and its allies.

About the Author

Pia Hüsch is a Research Fellow in cyber, technology and national security at RUSI. Her research focuses on the impact, societal risks and lawfulness of cyber operations and the geopolitical and national security implications of disruptive technologies such as AI. Prior to joining RUSI, Pia conducted her doctoral research on the lawfulness of offensive cyber operations in international law. Pia's other research interests include cybercrime, the governance of cyberspace, election interference, cyber warfare and the relationship between law and technology. Pia holds a PhD and an LLM in International Law Security (with distinction) from the University of Glasgow and an LLB in European Law from Maastricht University.