RUSI

Use of Generative AI Policy[1]

Do's and don'ts when using generative AI tools:

Do:

- Take time to understand the technology, how it works, what it is capable of and what are its limitations.
- Review all outputs thoroughly and ensure all claims are verified and confirmed to be based on evidence.
- Ensure you are providing clear input instructions and managing your expectations about the desired output.
- Protect sensitive information and personal data.
- Consider the ethical implications when using generative AI tools (particularly when it comes to perpetuating biases and stereotypes and/or generating offensive content).
- Verify guidelines on use of AI in contracts with funders.

Don't:

- Input proprietary, confidential or sensitive information or personal data.
- Fail to disclose use of AI in outputs.
- Take outputs as fact and publish content without thorough review.
- Use AI to generate misleading or false content.
- Rely solely on generative AI tools to generate a specific piece of content and/or deliverable.

[1] This policy does not form part of any employee's contract of employment, and it may be amended at any time. However, employees will be required to sign this policy upon hiring as a sign of understanding and acceptance. [confirm with Constance]

Foreword

Generative Artificial Intelligence (AI) tools and products offer potential efficiencies for the creation of content and for research tasks, including internal and external communications. While RUSI Staff are permitted to make use of AI tools in the course of their work, the purpose of this policy is to ensure that Staff use generative AI tools in a manner that is consistent with RUSI's values and ethical guidelines, its legal and contractual obligations, and in accordance with our data privacy controls.

Importantly, the field is evolving quickly, so Staff should take general care and attention to the latest developments, and if in doubt, seek advice from the Risk and Compliance Manager if they have questions regarding use of Generative AI and the adequate procedure for approval of its use.

Generative AI has the potential to provide efficiencies in the way that we work but also introduces new legal and commercial risks that RUSI wishes to mitigate. Use of generative AI that is not in accordance with this policy may give rise to a breach of your contract, a breach of funders' and stakeholders' policies or contracts, and/or the following:

- breach of data protection laws;
- breach of legal and/or regulatory requirements or guidance governing the development, deployment or use of generative AI;
- misuse of confidential information belonging to RUSI or to its funders, clients, members and/or suppliers;
- the generation of false or inaccurate information leading to legal liability and/or damage to the reputation of the user, RUSI, its funders, members, clients and/or its suppliers;
- breach of intellectual property rights; and/or
- RUSI's ability to protect works created using generative AI tools may be jeopardised.

To mitigate these risks, generative AI must be used responsibly, and in compliance with RUSI's policies. This policy does not seek to regulate how Staff use generative AI in a purely private capacity, provided that that use has no bearing on RUSI or its activities and provided that no Company materials, information or data is input into generative AI as part of such private use. This policy is intended to ensure that Staff understand the rules governing their use of generative AI in relation to their work for RUSI. It is designed to help you use generative AI responsibly, so as to minimise the risks set out above.

Application

This policy applies to The Royal United Services Institute for Defence and Security Studies (registered charity no.210639, Whitehall, London SW1A 2ET), and its subsidiaries and affiliated companies or organisations controlled by the Royal United Services Institute for Defence and Security Studies (collectively, "RUSI"). It applies to all individuals working at all levels of the organisation, including trustees, employees, consultants, interns and other voluntary workers, casual and agency staff as well as all sub-contractors and partners, together known as "Staff".

To be clear, it applies to all consultants and subcontractors when it comes to their engagements with RUSI. This means that consultants and subcontractors will be required to get authorization from the project officer for any use of generative AI tools throughout their engagement with RUSI or in relation to a project/contract in which RUSI is involved.

Definitions

'Generative Artificial Intelligence' or 'generative AI' are AI systems or models, such as ChatGPT and Bard, that can create new content based on the data that they have been trained on when given an instruction or input Prompt by the user. The difference between generative AI and other AI technologies is that generative AI creates or generates 'net-new' outputs, which could be text, graphics, and music as opposed to other systems which, for example, (i) can group data (identifying common characteristics or properties), (ii) classify or label data: (iii) or use data to come to a decision or determine an action.

'Hallucinations' occur where generative AI believes that it knows a fact to be true, but in reality, it is wrong. Due to the way in which generative AI is trained and operates, it can produce plausible answers which are inaccurate, and in some cases totally made up. As such, you should treat any information created by generative AI with caution and perform your own additional validation checks on any such information prior to using or relying on it. For instance, ChatGPT has been known to invent references and academics, and more worryingly mixes them in with real ones.

'IP' refers to intellectual property which is the ownership over intangible creations of the human intellect such as patents, copyrights, etc.

'Guardrails' are the rules given to generative AI technology requiring it to avoid certain topics or answers. For example, due to regulatory concerns, this might include financial advice. This may also include avoiding topics or answers around potentially unlawful matters such as the creation of dangerous objects. It may also be programmed to avoid subjects that it is not well-trained to answer. Staff should take this into account when assessing any outputs created using generative AI technologies.

'Prompt' is a question or request that you write for the generative AI tool to answer or solve.

'GPT' is short for "generative pre-trained transformer", which is a type of LLM (see below) that uses deep learning to produce natural language texts based on information requested in the input. ChatGPT is an example of a GPT model which can be used to generate text.

'Large Language Models' or 'LLMs' are a type of generative AI that can generate human like text in response to a Prompt. They use deep learning techniques and massive data volumes to generate a response.

Roles and responsibilities

- All Staff are responsible for ensuring that their use of generative AI is in accordance with this policy, and must make themselves aware of, and comply with, their responsibilities, as outlined in this policy, to protect confidential and sensitive information when using generative AI.

- Line managers are responsible for use of AI approval, as well as ensuring that their teams are aware of and comply with this policy. They will also be responsible for getting funder approval prior to using generative AI tools for research project outputs. Line managers must report any violations of this policy to the Risk & Compliance Manager.
- RUSI's Risk and Compliance Manager is responsible for documenting an approved list of generative AI systems for use by Staff.
- The Director of HR is responsible for handling any complaints concerning violation of or noncompliance with this policy, including any allegations of harassment, discrimination, or bias that may be raised by employees, customers or other third parties.

Principles underlying RUSI's Generative AI Policy

Adhering to the principles set out below will reduce risks associated to the use of generative AI in RUSI's work:

- Staff's use of generative AI in the workplace must be limited to use for business-related purposes and must, at all times, be in accordance with, but not limited to RUSI's policies.
- Staff must not use generative AI in any way that could be considered discriminatory, or could amount to defamation, harassment, intimidation or bullying, or in any way that could harm the reputation of another.
- Staff must not use generative AI to create illegal content or for illegal purposes.
- Staff must not use offensive, obscene or abusive language, graphics or imagery when inputting content into generative AI and must not attempt to create content which is offensive, obscene, or abusive through the use of generative AI tools.
- Staff must not input RUSI materials, information, or personal data into generative AI, including any client, funder, prospective funder, member or supplier materials, information or data and including any materials, information or data relating to Staff, whether via the input of such data as training data to a generative AI technology or in any instruction or Prompt. Uploading information to public AI tools, including that entered through Prompts or queries into tools like ChatGPT, is a form of release of that information to a third party.
- Staff must not enter protectively marked information (proprietary, confidential, classified or export controlled) into the generative AI tool.
- Staff must not input any personal or sensitive information into generative AI, including usernames, passwords, or security tokens.
- Generative AI tools must be used as a starting point only, meaning all outputs must be thoroughly reviewed and their reliability questioned. As Generative AI tools cannot currently discern between real and false information found in public sources, Staff must always assume outputs include false claims and/or bias and review each and every statement made.
- All content made using generative AI must ensure there are no issues with IP and whenever generative AI is used, this must be acknowledged prominently in the output and disclosed to all relevant third parties, such as funders, members, etc.
- Staff must not in any way provide or suggest any endorsement or recommendation by RUSI of any third-party generative AI technology.

- Wherever Staff do use generative AI, they must protect their login credentials and ensure that any generative AI accounts that they hold are not accessible to unauthorised third parties. The use of multi-factor authentication is advised in respect of any generative AI tools and technologies used.
- Staff use of generative AI should be in compliance with all applicable laws and regulation, including data protection and privacy laws.

Using generative AI

RUSI recognises that the use of generative AI within your daily work may help with brainstorming ideas or producing content for internal use. However, outputs produced by generative AI tools should never be considered a finalized product.

Before using generative AI technologies, you must obtain prior written approval of your Line Manager. Approval must be in accordance both with this policy and RUSI's Ethics Policy for Research Projects and should be granted on a case by case basis If a generative AI tool is to be used in outputs linked to a research project, funder approval is also required.

RUSI's Risk and Compliance Manager will keep a list of approved generative AI tools, which can be found on SharePoint. Only generative AI tools from this list may be used to generate new content. AI tools which do not create net new outputs will be excluded from this policy. If you would like a tool to be added to the approved list, your request should be made to the Risk and Compliance Manager.

Once approval has been granted to make use of a generative AI tool by the Line Manager and the funder where applicable, Staff must abide by the following guidelines:

- You must never input trade secrets, confidential, valuable, or personally identifiable information (information that makes it possible to work out who a person is) into AI generative tools;
- To avoid copyright infringement issues, outputs produced by generative AI tools should only be a small fraction of the finished product and acknowledgement that these outputs are not the property of the author will always be required.
- Before circulating internally or externally, publishing or otherwise making available any output created with the help of generative AI, Staff should review and edit it for proper context and accuracy. Staff should also treat any information created by generative AI with caution and perform their own additional validation checks on any such information prior to using or relying on it due to Hallucinations and Guardrails.
- When circulating, publishing or otherwise making available the output, Staff should clearly identify that the output has been created with the assistance of, generative AI technologies. Acknowledgment of generative AI must be prominently featured in the output.
- Staff must maintain good information security practices as per RUSI's Data Protection Policy, Cyber Security Guidance for Line Managers and Staff and the Cyber Requirements for Contractors.
- Staff must abstain from inputting RUSI's trademarks, works subjected to copyright, brands, logos, or other identifying material as well as RUSI email addresses or other contact details into generative AI tools.

- All uses of generative AI within RUSI must be from a properly licensed generative AI service and must have been reviewed and approved by RUSI's Risk and Compliance Manager, based on whether or not they offer sufficient protection of confidential information, data privacy and security, and include sufficient safeguards around accuracy, bias detection, intellectual property rights and protection, and other pertinent risks.

If generative AI is going to be used in the production of RUSI Public Facing Outputs, Staff must be particularly careful and aware of the following:

  - Norms and requirements surrounding the citation of AI-generated output, as well as disclosure of the use of AI technologies, are complex, rapidly evolving, and often unclear. In addition, these norms and requirements may vary substantially depending on the source; for example, publishers, journals, professional organizations, and funding organizations may all have (or be in the process of developing) policies surrounding the use of AI that need to be navigated.
  - Those who are involved in proposing, internally reviewing, performing, or disseminating research bear the responsibility for familiarizing themselves with the policies and standards governing the use of generative AI in their research, and are ultimately responsible for the work that they produce and disseminate. They are also responsible for seeking approval from RUSI's Ethics and Integrity Committee.
  - This responsibility includes properly attributing ideas and credit, ensuring the accuracy of facts, relying on authentic sources, and appropriately disclosing the use of AI in research.
  - The use of generative AI should be clearly and transparently disclosed and documented as part of the research methodology. Documentation requirements will vary substantially from research context to research context and from discipline to discipline, but researchers should err on the side of caution by explicitly and prominently disclosing any material use of generative AI in their research activities, as well as clearly highlighting which parts of an output were made using generative AI tools.
  - Moreover, many research contracts or grant agreements will explicitly establish that the client or funder will own the intellectual property rights of the resulting research or outputs. These types of contracts will effectively prohibit the use of generative AI tools in the production of any outputs as RUSI can't guarantee sole IP rights to the funder if generative AI tools have been used in the production of a deliverable.
  - Researchers should also be aware of or anticipate conflicting views surrounding generative AI coming from journals, publishers, professional associations/societies, and funders/sponsors regarding the use generative AI in the creation of new content.

Personal use of generative AI

You must not use RUSI's computers, networks, or systems (including via smartphones or tablets) to access generative AI tools for personal use at any time. Any unauthorised use of generative AI is strictly prohibited.

Monitoring

- RUSI's Cyber Security policies in particular in relation to RUSI's right to monitor, intercept and read communications, applies equally to use of generative AI technologies via RUSI's systems or network.
- RUSI's AI policy should be read in conjunction with the following policies: Data Protection Policy, Data Protection Guidance for Researchers, RUSI's Cyber Security Guidance for Line Managers and Staff and RUSIs Cyber Security Requirements for Contractors.

Breaches of this policy

Because of the importance of this policy, an employee's failure to comply with any requirement of it may lead to disciplinary action and this action may lead to dismissal for gross misconduct. If you are not an employee, breach of this policy may result in termination of RUSI's contract with you with immediate effect.

You are also reminded that, in certain circumstances, an act that breaches this policy may also constitute a criminal offence. You should note, in particular, that inputting the Institute materials, data or information (including commercially sensitive or confidential information), into generative AI tools may amount to misconduct even if it takes place:

a) on a personal account with appropriate privacy settings.
b) outside normal working hours; and/or
c) without using the Institute's computers, software, systems, and networks.

If, while working for RUSI, you become aware of any misconduct or wrongdoing by any member of Staff in breach of this or related policies, you must report it to the Risk and Compliance Manager.

RUSI will review and update this policy regularly to take account of changes in technology, legal obligations and best practices.

Version control

Author	First drafted	Approval date and approving body	Latest update
Andrea Plazas	25/10/2023	14-12-2023 approved by Research Committee	07-12-2023