Annex B: The Use of Artificial Intelligence (AI) in Defence Training

Annex version 1.1 (Oct 26)

1. **Direction On The Use Of Artificial Intelligence (AI) In Defence Training**.....2
1.1. Introduction.....2
1.2. Defining AI.....2
1.3. Coherence and future policy development.....2
1.4. Wider Defence policy on AI.....3
1.5. Copyright rules.....3
1.6. The Defence Learning Management Capability (DLMC) ecosystem.....4
1.7. AI use in the generation of DSAT documentation.....4
1.8. Identification of AI use within Performance, Conditions and Standards (PCS).....4
1.9. Use of AI tools in Training Evaluation.....4
1.10. AI in Delivery.....5
1.11. Use of AI tools by students in Assessment.....5
1.12. What is AI use and what are the risks of using it in assessments?.....6
1.13. Refencing AI use within research.....6
1.14. AI misuse.....7
1.15. Detection.....7

Appendix

1 - **Guidance table of Do's and Don'ts**

Annex B: The Use of Artificial Intelligence (AI) in Defence Training

Annex version 1.1 (Oct 26)

1. DIRECTION ON THE USE OF ARTIFICIAL INTELLIGENCE (AI) IN DEFENCE TRAINING

1.1. Introduction

1.1.1. With the rapid advancement of generative AI, it is imperative that the use of AI in Defence Training is developed in a safe, effective and cohered way. This annex provides specific direction on the appropriate application of AI and AI-enabled tools by Defence Systems Approach to Training (DSAT) practitioners, trainers and students across the full range of DSAT activity. It has been aligned with JSP 936[1].

1.2. Defining AI

1.2.1. There is no agreed definition of AI; JSP 936 characterises it as “Machines that perform tasks normally requiring human intelligence, especially when the machines learn from data how to do those tasks[2].” Generative AI is a subset of AI capable of generating text, images, video or other forms of output by using probabilistic models trained across one or more domains.[3]

1.3. Coherence and future policy development

1.3.1. **DSAT AI Coherence Working Group (DACWG).** The DACWG is chaired by the Talent, Skills, Learning and Development (TSLD) Training Policy and Assurance Team. It sits quarterly and is enabled by a Teams Channel[4]. The WG acts as a point of coherence for the emerging use of AI in training pan-Defence and provides the following functions:

- Provides stakeholders with a common picture of innovation, experimentation and research around AI in training across Defence and amongst NATO partners.
- Provides a forum for collaboration and knowledge-sharing pan-Defence, including new use cases and emerging good practice.
- Identifies opportunities for coherence of activity pan-Defence.
- Identifies risks, issues and appropriate mitigations.
- Provides a forum to support the iterative development of Defence policy and guidance on AI in training.

1.3.2. **Military Command Responsibilities.** As key stakeholders, military commands **must** do the following:

[1] JSP 936 Dependable Artificial Intelligence (AI) in Defence
[2] JSP 936 Ch 1 Para 3.
[3] https://www.gov.uk/government/publications/ai-insights
[4] DSAT AI Coherence Working Group

- Provide appropriate representation at the DACWG.

- Keep TSLD Policy and Assurance Team updated on their current governance and coherence arrangements for AI in training (e.g titles and ToRs of relevant SGs/WGs, contact details for desk leads).

- Provide updates at the DACWG on emerging use cases for AI in training and associated lessons or good practice.

1.4. Wider Defence policy on AI

1.4.1. JSP 936 – Dependable Artificial Intelligence (AI) in Defence is the authoritative direction for the use of AI in Defence. JSP 740 Acceptable Use Policy (AUP) for Information and Communications Technology and Services (ICT&S) gives specific direction for the use of AI with ICT:

- You **must not** enter MOD information into public Internet-facing search engines or AI
apps and tools.

- As a DSAT practitioner, trainer or student you **remain responsible** for the **quality and accuracy of your work**, therefore you **must** review and verify that your outputs are accurate and appropriate.

- If using generative AI as a productivity aid for work purposes, only use tools provisioned or authorised by the MOD.

- Service Personnel attending courses run by an external provider **must** ensure they are aware of and comply with the provider's policies covering the use of AI.

1.5. Copyright rules

1.5.1. To copy or use a copyrighted work, you **must** obtain permission from the copyright owner, for example via an appropriate licence. Under UK law, permission is required to use copyrighted material as prompts, for data summarisation or data extraction in generative AI[5] or as training data for new AI models. Users are permitted to use materials in which they own the copyright – for example, where MOD staff generate prompts independently as part of their duties, these will be Crown copyright and may be used for all official purposes. Copyright material that is owned by third parties may only be used where the MOD has been granted a specific licence to use / upload the material to AI. Whilst material which is out of copyright is free to use, it is not safe to assume that copyright protection has expired without obtaining confirmation, as copyright can often subsist for 100+ years. TPs **must** provide clear direction on the copyright licences held and the permissions they

[5] https://cla.co.uk/ai-and-copyright/

allow. If users are uncertain whether copyright restrictions apply to the material they wish to use, advice **should** be sought from the MOD's Intellectual Property Rights Group[6].

1.6. The Defence Learning Management Capability (DLMC) ecosystem

1.6.1. The DLMC is a significant programme aimed at modernising the management of learning and development across Defence. Whilst there is the freedom to establish local contracts with software suppliers for AI-enabled tools, Defence training organisations must first consider the need for coherence within the wider DLMC ecosystem and the possibility of utilising existing contracts or in-house solutions. JSP 822 Chapter 3 Annex A lays out the need for coherence, compatibility and interoperability with current and planned DLMC services.

1.7. AI use in the generation of DSAT documentation

1.7.1. **All AI-generated DSAT documents must undergo review by a qualified Subject Matter Expert (SME).** Any DSAT content produced with the assistance of AI tools, whether partially or fully generated, **must** be formally validated by an authorised SME before it is shared, stored, or used for training or assessment purposes. The SME is responsible for confirming the accuracy, relevance, and compliance of the material, ensuring that AI-generated outputs meet organisational standards and do not introduce errors, bias, or misinformation into DSAT resources.

1.7.2. **Only authorised DSAT creators are permitted to generate new DSAT documents.** To maintain the integrity, accuracy, and security of training and assessment materials, the creation of any new DSAT content—whether produced manually or with the support of AI tools—must be carried out exclusively by individuals formally designated as DSAT creators. Staff and learners **must** not use AI systems to generate, modify, or replicate DSAT documents unless they have explicit authorisation. This ensures consistency with organisational standards, protects sensitive information, and prevents the unintended distribution of unverified or non-compliant materials.

1.8. Identification of AI use within Performance, Conditions and Standards (PCS)

1.8.1. During the analysis and design stages of the DSAT process, TRAs **must** consider whether the conditions detailed within training objectives' Performance Conditions and Standards (PCS) **should** stipulate whether a task is to be completed with or without the use of generative AI.

1.9. Use of AI tools in Training Evaluation

1.9.1. The analysis of training evaluation data, particularly when dealing with both qualitative and quantitative survey data, can be a time-consuming and resource-intensive process. The integration of AI tools offers an efficient solution by speeding up the processing, analysis, and reporting of survey data. Training establishments should use AI

[6] DIPR-Secretariat@mod.gov.uk

as a complement to, rather than a replacement for, traditional methods of analysis. Guidance on the use of AI in evaluating training data is as follows:

- **Quality of data input.** AI analysis requires good data. Where data is incomplete (due to low return rates) or biased (due to poorly designed surveys), AI may provide inaccurate or misleading responses. It is crucial to ensure that survey questions and response options are clear, well-structured, and designed to minimise bias, enabling AI tools to accurately interpret and analyse the data.

- **Combine AI analysis with human validation.** While AI can generate valuable insights, it cannot substitute the critical thinking, contextual understanding, and expertise of human analysts. Human validation is essential to ensure the accurate interpretation of results, particularly when addressing complex or sensitive issues. AI should be used to enhance, not replace, human judgement in the decision-making process.

- **Ethical and privacy concerns.** Analysing student feedback **must** remain compliant with GDPR and data protection regulations. Where AI tools are employed in survey response analysis, this must be transparently communicated to participants and reflected in the consent notice provided to them. Training Establishments **must** ensure that data is handled ethically and securely, avoiding public / open generative AI tools when evaluating sensitive data.

1.10. AI in Delivery

1.10.1. **Equity of access to AI tools by students.** Where generative AI is used to support learning, assessment preparation, or skill development, appropriate guidance, training, and access should be provided so that all learners can benefit fairly and safely. Measures should be taken to identify and mitigate barriers—technical, financial, or educational—to uphold an inclusive learning environment.

1.11. Use of AI tools by students in Assessment

1.11.1. TDAs **must** state their permitted use of AI within AStrats and ASpecs, ensuring that this information is communicated to students.

1.11.2. To provide clear and consistent direction on the appropriate use of AI in ASpecs the following categorisation system **should** be used:

- **Category 1:** Prohibits the use of generative AI, enabling the demonstration of cognitive development, independent knowledge and understanding.

- **Category 2:** Permits the use of generative AI to support student learning and development. Students **must** reference all AI usage. The focus remains on evaluating students' independent understanding and application of knowledge, while still permitting AI as a supplementary resource.

- **Category 3:** Use of generative AI tools is mandated and assessed, aiming to build critical AI literacy and ethical awareness.

1.11.3. This tiered approach seeks to balance the increasing need for AI proficiency with the imperative of maintaining academic integrity and genuine skill development.

1.12. What is AI use and what are the risks of using it in assessments?

1.12.1. It is important that trainers and students alike are aware that the range of AI tools and their capabilities is expanding quickly, and that there are limitations to their use (such as producing inaccurate or inappropriate content known as hallucinations[7] [8]). Where the use of AI is authorised to obtain information and content which might be used in work produced for assessments, careful consideration must be taken. AI tools often produce answers which may seem convincing but contain incorrect or biased[9] information. Some AI tools have been identified as providing answers to questions that can prompt inappropriate actions, and some can also produce fake references to books/articles[10] [11].

1.13. Refencing AI use within research

1.13.1. There are two levels of attribution. The first level is formal citation, required when specific, unique content generated by AI (image, diagram, text, code, or paraphrased ideas) is directly incorporated into the submission. The second, is an overarching acknowledgement, mandated for process-based assistance, such as using the tool for brainstorming, grammar refinement, or outlining, which should typically be placed in an acknowledgments section or, for technical work, the methodology section.

1.13.2. **In-text citations.** Traditional citation relies on the concept of a recoverable source; a stable document or dataset that a peer can access and verify. LLM outputs fail this standard because responses vary over time, the prompt itself is often proprietary to the user, and a persistent, public URL for the specific output is typically unavailable, all making referencing a challenge for the researcher. To ensure verifiability and align with the rigorous expectations of academic partners, a robust LLM citation **must** capture specific metadata necessary for future audit. Where TDAs/TPs allow the use of AI in research the guidance below provides a comprehensive framework that may be adopted:

- Prompt Description/Output Title: A description of the prompt is necessary to identify the generated content for review.

[7] Reducing Hallucinations and Trade-Offs in Responses in Generative AI Chatbots for Cancer Information: Development and Evaluation Study – PMC
[8] GPTZero uncovers 50+ Hallucinations in ICLR 2026
[9] Measuring gender and racial biases in large language models: Intersectional evidence from automated resume evaluation
[10] HalluCitation Matters: Revealing the Impact of Hallucinated References with 300 Hallucinated Papers in ACL Conferences
[11] Fabricated citations in the age of AI: A wake-up call for editors, reviewers, and authors - PMC

- Tool Name and Version: The specific model used (e.g., OpenAI ChatGPT 3.5 or 4.0) **must** be listed, as model architecture and training data influence output.

- Date of Interaction: The exact date (day, month, year) the prompt was entered and the response received is crucial to account for rapid model updates and output drift over time.

1.13.3. An example of in-text citation would be: When given the prompt "*explain who Clausewitz is to a GCSE level student*" the LLM generated text indicating that "*Carl von Clausewitz (1780–1831) was a Prussian (German) soldier and military thinker. He is most famous for writing a book called On War, which is still one of the most influential works on warfare, strategy, and military leadership.*" (Microsoft Copilot (GPT5.0 model), 20 Mar 2026).

1.13.4. **Acknowledgement.** Students **must** acknowledge any use of AI, even if they didn't directly use the AI generated content. For example, if they have used AI in the development of ideas and concepts of the work, they must still acknowledge that usage. A thorough disclosure statement **must** include the name and version of the generative AI system used, the company, the URL, the date of content generation, and a brief, yet specific, description of how the tool was applied[12].

1.13.5. An example of an acknowledgement would be: "Utilised ChatGPT (v 5.1- www.chatgpt.com, 13 Mar 26) to assist with creation of my draft essay plan".

1.14. AI misuse

1.14.1. AI misuse refers to any use of generative AI tools that breaches organisational policy, undermines learning integrity, or creates risks to accuracy, fairness, or security. Examples include (but are not limited to):

- Unauthorised use of generative AI tools to complete assessed work.

- Using generative AI tools to complete assessed work without disclosure. This includes using generative AI tools to create ideas, examples, essay plans or correct draft work without adequate acknowledgement.

1.15. Detection

1.15.1. **AI Detection Tools**. AI detection tools must not be relied upon as the **sole** source of evidence for unauthorised AI use. AI detectors are inherently biased, and it is widely acknowledged that these detectors consistently misclassify Non-Native English Speakers' (NNES) writing samples as AI-generated[13]. This algorithmic bias is

[12] A primary function of a reference list is so the reader can refer to the original source, which is not possible with AI generated content; most academic publishers stipulate GenAI systems should not typically be cited as a source in the reference list (https://library-guides.ucl.ac.uk/referencing-plagiarism/acknowledging-genAI).
[13] GenAI detection tools, adversarial techniques and implications for inclusivity in Higher Education

also evident in the incorrect identification of content written by students with autism spectrum disorder. Deploying tools that unintentionally penalise or exclude neurodiversity or NNES students contradicts fundamental academic equity goals. Until such time that detection software can demonstrate no bias towards neuro-diverse or NNES student, TPs must look at alternative approaches to AI misuse detection within academia or risk violating procedural fairness and eroding institutional trust.

1.15.2. **Developing an Evidence-Based Investigation Protocol.** When suspicion of unauthorised assistance or fabrication arises, the investigation protocol must rely on definitive, verifiable markers:

- **Verbal Défense (Viva):** The student is unable to verbally explain, expand upon, or defend the submitted work. The requirement to defend one's work verbally is a nearly perfect defence against both cheating and sophisticated AI use, as it assesses the student's *understanding* (process) over the text's *fluency* (product). Only the authentic author can reliably explain the reasoning and intellectual trajectory of the work.

- **Non-Verifiable Content:** The submitted work includes citations or references to content that does not exist, cannot be verified, or relies on data that is fabricated or implausible.

- **Inconsistency with Process:** The final work is inconsistent with preparatory steps (outlines, drafts) previously submitted by the student, or the inclusion of advanced approaches exceeds the known scope or demonstrated ability of the student.

Appendix

1. Guidance Table of Do's and Don'ts

Appendix 1 to Annex B: The Use of Artificial Intelligence (AI) in Defence Training

Guidance table of Do's and Don'ts

Area	Do	Don't	Reason
Academic Integrity & Attribution	**Do** cite AI-use clearly in academic work following institutional and subject-specific guidelines.	**Don't** submit AI-generated content as your own original work without proper attribution.	**Reason:** To prevent **academic misconduct** (plagiarism) and maintain transparency, ensuring the work genuinely reflects the user's skills and knowledge.
	Do critically evaluate all AI outputs for accuracy and bias before reliance or incorporation.	**Don't** assume AI-generated content is accurate, unbiased, or a substitute for core learning.	**Reason:** AI systems are known to "hallucinate" (produce false information) and can perpetuate **bias** present in their training data. You are ultimately responsible for the work.
Copyright & IPR	**Do** only input content into AI tools that you are certain you have the **right to use** (e.g. your own original content).	**Don't** input content from copyrighted materials (e.g. textbooks, journal articles, licensed images) for the purpose of **training** a public or open AI model.	**Reason:** Most AI tools' terms and conditions state that input data may be used to further train the model, which is a use **not covered** by the CLA education licence and constitutes **copyright infringement** (using material to train AI is generally considered copying that requires a specific licence).
	Do use AI output as a starting point, then critically revise, re-write, and fact-check it to ensure originality.	**Don't** use AI to generate entire pieces of text or code for submission without substantial human transformation.	**Reason:** To mitigate the risk of inadvertently reproducing material from the AI's training data (**copyright infringement**) and to adhere to academic integrity rules.

Appendix 1 to Annex B: The Use of Artificial Intelligence (AI) in Defence Training

Annex version 1.1 (Oct 26)

Area	Do	Don't	Reason
GDPR & Data Protection	**Do** use anonymised or aggregated data when conducting research using AI tools.	**Don't** input **any personal data** (names, addresses, student IDs, sensitive research data, etc.) into public/open generative AI tools (e.g. free versions of ChatGPT, Gemini, etc.).	**Reason:** Public AI models typically do not offer **commercial data protection** and may store, share, or use the input data for further training, causing a **GDPR breach** by transferring personal data to an unauthorised third party outside of your control.
Audit	**Do** retain sufficient records of how AI tools have been used.	**Don't** use AI in ways that prevent explanations or try to hide that you have used it to assist you.	**Reason**: Defence requires **transparency** and trying to hide use could result in disciplinary/administrative action and possible RTU due to AI misuse.
Decision making and Use	**Do** retain a human in the loop for all outputs informed / created by AI.	**Don't** rely on AI generated outputs to justify decisions without verification from a human.	**Reason**: AI systems **are fallible.**