Swedish Defence University

Comprehensive Shield 2025

An Explorative Study of the Impact of Artificial Intelligence on the Operations Planning Process (Final Report)

Peter Bovet Emanuel, Ørjan Rogne Rise, Stig Tore Aannø, Hanna Forsberg, Anders McDonald Sookermany, Thomas Slensvik, Martin Gurák, Aaron P. Jackson

Swedish Defence University

Comprehensive Shield 2025

An Explorative Study of the Impact of Artificial Intelligence on the Operations Planning Process (Final Report)

Peter Bovet Emanuel, Ørjan Rogne Rise, Stig Tore Aannø, Hanna Forsberg, Anders McDonald Sookermany, Thomas Slensvik, Martin Gurák, Aaron P. Jackson

Peter Bovet Emanuel, *Swedish Defence University*,
https://orcid.org/ 0009-0006-4353-1911

Ørjan Rogne Rise, *Norwegian Defence University College*

Stig Tore Aannø, *Norwegian Defence University College*

Hanna Forsberg, *Swedish Defence University*

Anders McDonald Sookermany, *Norwegian Defence University College*

Thomas Slensvik, *Oslo School of Architecture and Design*

Martin Gurák, *Swedish Defence University*

Aaron P. Jackson, *Swedish Defence University*

SWEDISH DEFENCE UNIVERSITY REPORT SERIES 2025:7

Cover: ChatGPT, "Create a futuristic picture that shows a military staff exercise that engages in exploring human-machine teaming for joint operational planning. The participants have Swedish and Norwegian field uniforms."

Publisher Swedish Defence University
Publication Year 2025
ISBN 978-91-88975-63-8
ISSN 2004-7894
DOI https://doi.org/10.62061/ooom5089

© the authors
This report is licensed under the Creative Commmons CC-BY-NC-ND 4.0 international license. For full terms visit https://www.creativecommons.org.

The views expressed in this report are those of the contributors and do not necessarily reflect the views or policies of any organization with which any contributor is affiliated. While reasonable care has been taken in preparing this report, to the extent permitted by law, the Norwegian Defence University College, the Swedish Defence University, and the contributors disclaim all liability howsoever caused (including as a result of negligence) arising from the use of, or reliance on, this report. By accessing this report, users are deemed to have consented to this condition and agree that this report is used entirely at their own risk

Swedish Defence University
Box 278 05
115 93 Stockholm

www.fhs.se

Table of Contents

Executive Summary **7**

Summary of Recommendations 8

Introduction **11**

Experiment Design **13**

Conditions and Setup **14**

Artificial Intelligence-based Decision Support System 14

Methodology **15**

Data Collection Instruments 15

Ethics, Reliability and Validity Considerations 17

Study Area Analyses **18**

Study Area 1 – Military Planning and Decision-Making 18

Study Area 2 – Comparative Analysis of Group Performance 20

Study Area 3 – Information and Communication Technology and Pedagogical Perspectives 25

Study Area 4 – Attitudes Toward Artificial Intelligence in Military Contexts 29

Forward Outlook **33**

Recommendations for Exercise Comprehensive Shield 2026 33

Suggested Future Research Directions 34

Acronym List 35

About the Contributors 36

Executive Summary

This study marks a historic milestone: the first joint artificial intelligence (AI)-enabled operational planning experiment conducted by the Norwegian Defence University College (NDUC) and the Swedish Defence University (SEDU). Integrated into Exercise Comprehensive Shield 2025 (CS25), the study explored the practical and pedagogical implications of using AI-based decision support systems (DSS) in military planning during professional military education (PME). The study explored AI's influence on planning, time efficiency, cognitive load, educational impact, and user attitudes.

Three Joint Operations Planning Groups (JOPGs) tackled a shared problem within a synthetically developed exercise scenario using two different methodological approaches. The North Atlantic Treaty Organization's (NATO's) Comprehensive Operations Planning Directive (COPD) is the official strategic and operational level guide for the conduct of the Operations Planning Process (OPP). The COPD was used by the first two JOPGs. The third JOPG used an alternative methodology termed Systems-Oriented Design (SOD), which combines design thinking with systems thinking methods. Each JOPG used the AI-based DSS as follows:

- JOPG Alpha (Control): Used COPD but did not use the AI-based DSS
- JOPG Bravo (Treatment 1): Used COPD with the AI-based DSS
- JOPG Charlie (Treatment 2): Used SOD with the AI-based DSS.

Findings indicate that an AI-based DSS can enhance creativity, reduce cognitive load, increase decision confidence, and enable novel planning workflows beyond traditional boundaries. At the same time, this requires staff/a JOPG to overcome some intrinsic obstacles that include workload distribution and needing to develop a basic understanding of, and ability to properly utilize, the technical platform. Importantly, JOPG Bravo and JOPG Charlie came to use the AI-based DSS in two diverse ways. JOPG Bravo applied it throughout the COPD with an increase in their proficiency. Participants in JOPG Bravo also highlighted the pedagogical value of AI as a real-time support tool. For JOPG Charlie, working with a SOD methodology, the

AI-based DSS was available but not heavily used. When applied, its output was not perceived as particularly useful alongside a methodology that emphasized more visual and systemic ways of framing underlying challenges. This suggests that AI's impact may be more pronounced in traditional planning contexts, while its role in supporting alternative methodologies such as SOD requires further exploration.

> "KI kommer med løsninger som vi ikke nødvendigvis selv hadde sett. Og selv om vi kanskje ikke bruker dem, så hjelper det oss til å være kreative og utforske nye muligheter."[1]
>
> Major Nicholas Gran (student)[2]

Crucially, the experiment was embedded within CS25 without disrupting core exercise objectives. Participants reported that the scenario context was meaningful and relevant to future military capability development. Many expressed that the experience deepened their understanding of operational planning and broadened their perspective on AI's role in future defence environments.

The study also offers actionable insights into how AI experimentation can be effectively integrated into complex joint exercises. In the context of growing Nordic PME collaboration, these findings have broader implications—contributing to NATO's evolving concepts, curriculum development, and practical guidance on AI in operational planning and command and control (C2).

External interest has been notable. The experiment attracted senior leadership from both institutions and visiting stakeholders from Norway, Sweden and beyond. Their engagement—through observation and dialogue—underscored that the project is viewed not merely as an educational innovation, but as a forward-looking exploration of AI's role in military decision-making.

In sum, this experiment functioned as more than a pedagogical trial or technical comparison. It established a structured innovation space—one that facilitated student growth, cross-institutional faculty collaboration, and strategic reflection. Importantly, it demonstrated that bounded experimentation with AI can coexist with the integrity and discipline of operational-level exercises. The shared experience and trust developed through this collaboration now provide a strong foundation for continued research and joint publication.

Summary of Recommendations

This study makes a total of 18 recommendations. Twelve of these relate to findings within the experiment's four Study Areas, which are detailed in Section 5 of this

[1] English translation of the Norwegian citation: "AI comes up with solutions that we had not necessarily seen ourselves. And even though we may not use them, it helps us to be creative and explore new possibilities."

[2] Quoted in: Wettre, S. (2025) 'Et glimt inn i Forsvarets digitale fremtid', Forsvaret, 17 June 2025. Online: https://www.forsvaret.no/aktuelt-og-presse/et-glimt-inn-i-forsvarets-digitale-fremtid, accessed 30 September 2025.

report. An additional six recommendations relate to the design of Exercise Comprehensive Shield 2026, and are detailed in Section 6. Recommendations relating to exercise design are numbered with the prefix ED. All 18 recommendations are summarized in the table below.

No.	**Recommendation**
1.1	AI tools should be strategically integrated into doctrinal planning timelines, especially for early-phase scoping and ideation.
1.2	PME environments should train students in collaborative prompting techniques, enabling teams to leverage AI's full cognitive augmentation potential.
1.3	Future versions of AI-based DSS should incorporate modular COPD-specific guidance, signaling the tool's optimal application points and preventing premature convergence or overreliance.
1.4	The experiment confirmed the value of embedding AI as a conceptual framework, not just a technical assistant, which has implications for both concepts, doctrine, and system development.
1.5	Future experiments should investigate how staff officers and team leaders claim cognitive agency of both the process (methodology) and the progress of planning tasks in a more innovative way (approach).
1.6	Future experiments should investigate how different tools affect reflective thinking and understanding of the problem space.
2.1	Future research should be conducted to further explore the extent to which different tools and methodologies encourage different levels of creativity, for example by employing single/double/triple loop learning systems, or complexity theory.
3.1	Future studies should use Kirkpatrick's four levels of training evaluation to measure student learning and retention in addition to the areas that were measured in this study.
3.2	Students should be provided with the necessary time and support to ensure they fully understand the functionality of AI tools that are used during learning activities. This should occur before activities commence.
3.3	AI should be used more broadly within PME in support of group problem-solving and to augment collective learning, as its use frees time and cognitive capacity for peer-to-peer student discussions and deeper learning.
4.1	Future studies should track how participant attitudes evolve across settings (operational vs educational) and among officers with differing levels and types of prior experience.
4.2	Future studies should also investigate factors such as age, gender, years of service and perceived technological competence, and the impact these have on attitudes towards AI use in military contexts

No.	**Recommendation**
ED.1	Early familiarization and training: Participants should receive structured introductions to the AI-based DSS prior to Exercise Comprehensive Shield 2026. This should include hands-on workshops in collective prompting, interpretation of outputs, and critical validation techniques.
ED.2	Focused experimentation: For Exercise Comprehensive Shield 2026, and in relation to exploring AI-based DSS more thoroughly, all JOPGs included in the experiment must use the same COPD planning methodology, to enable comparisons between them to be made with more precision.
ED.3	Diverse methodologies: Exercise Comprehensive Shield 2026 should continue to explore a diversity of alternative methodological approaches beyond the COPD planning methodology. However, this should occur outside of the scope of this AI-based DSS exploration.
ED.4	Cross-JOPG exchange: Opportunities should be built into Exercise Comprehensive Shield 2026 for deliberate cross-JOPG exchange at both midpoint and conclusion.
ED.5	Pedagogical integration: During Exercise Comprehensive Shield 2026, structured JOPG debriefings should more explicitly address how AI affected decision-making, creativity, and cognitive load. Reflective sessions could be complemented by facilitated dialogues between instructors, observers, and students.
ED.6	Technical and Logistical Support: During Exercise Comprehensive Shield , the AI-based DSS infrastructure should be further stabilized and supported by clear user guidance materials. Redundancy measures should be established to mitigate potential technical disruptions. Technical facilitators embedded in each JOPG may further reduce barriers to effective tool use.

Introduction

This experiment was a joint initiative between NDUC and SEDU to explore and evaluate AI's practical and pedagogical impact in an operational planning context. It occurred within CS25, a staff exercise during which students simulated a combined joint task force (CJTF) headquarters and were tasked with responding to the fictional scenario *Constellatia*. The experiment itself took place within the headquarters' J5 (plans) staff, which was divided into three JOPGs.

The AI-based DSS was positioned not only as a technical tool but also as a catalyst for rethinking decision-making, staff workflow and education in a NATO-style headquarters. As such, *this experiment was not limited to only studying how AI could augment the OPP; it has also helped to develop an understanding of the underlying possibilities and limitations of AI in similar processes*. Moreover, this experiment gave valuable insights into how AI affects internal human cooperation within a team, and how AI can enhance and expound rather than replace human judgement in decision-making.

In contrast to earlier research-integrated course initiatives at our institutions, this experiment does not appear to have generated frustration among students regarding the research presence. None of the JOPGs expressed that the research design interfered with their engagement or learning. On the contrary, students repeatedly articulated appreciation for being part of a clearly framed experimental cell. Moreover, the opportunity for one JOPG to work with a new methodology (i.e., SOD) was seen as a unique exposure to alternative perspectives. Several participants noted that their only regret was the lack of opportunity to later merge approaches with the other groups, which they felt would have deepened their learning.

This report is structured into six sections including this introduction. Subsequent sections proceed as follows.

- Section 2: Experiment Design, summarizes the experiment's focus and its four constituent study areas.
- Section 3: Conditions and Set Up, details the manner in which the experiment was conducted and explains the nature of the AI-based DSS that was used.
- Section 4: Methodology, establishes the research approach, and explains the data collection instruments, and ethics, reliability and validity considerations.
- Section 5: Study Area Analyses, details the research findings in each of the experiment's four study areas. It also makes recommendations in each study area, which relate either to enhancing PME pedagogy, enhancing the use of AI-based DSS in relation to military planning, or constructing future experiments of a similar nature.

- Section 6: Forward Outlook, makes recommendations for the design and execution of Exercise Comprehensive Shield 2026, and suggestions for possible future research directions.

Experiment Design

The focus of this experiment was to explore the use of an AI-based DSS in JOPG Bravo and JOPG Charlie, and to compare the workflow and results between these JOPGs, and between them and JOPG Alpha, which constituted a control group. The experiment thereby adopted a *controlled* format based on three different JOPGs:

- JOPG Alpha (Control): Used COPD but did not use the AI-based DSS
- JOPG Bravo (Treatment 1): Used COPD with the AI-based DSS
- JOPG Charlie (Treatment 2): Used SOD with the AI-based DSS.

In this sense, the experiment was *controlled* because it accepted some constraints imposed by the overall setup of CS25. For example, participants from each JOPG could interact with participants from the other JOPGs throughout the exercise.

The experiment design was framed by four different *study areas*, each of which engaged in different aspects of the phenomena studied.

- Study Area 1 – Military Planning and Decision-Making, investigated how AI-based DSS affected the OPP at the operational level, as contained in the COPD.
- Study Area 2 – Comparative Analysis of Group Performance, compared the performance of the three JOPGs across key dimensions such as plan quality, time efficiency, innovation, and cognitive load.
- Study Area 3 – Information and Communication Technology (ICT) and Pedagogical Perspectives, assessed how AI tools can function as pedagogical assistants during complex staff exercises.
- Study Area 4 – Attitudes Toward AI in Military Contexts, examined participant trust in AI, ethical considerations, perceived risks, and readiness for AI integration in the operational environment.

Each of the four study areas and their respective research focus, key research questions, hypotheses, key findings, and recommendations are presented in Section 5 of this report.

Conditions and Setup

All JOPGs worked under equal conditions, which included an identical scenario, receipt of the same planning directives, and a requirement to achieve the same planning milestones and outcomes.

Participants were NDUC or SEDU students who held the rank of OF-2 to OF-4, and who applied for positions in the three JOPGs prior to commencement of CS25. A few participants who applied to join JOPG Bravo were instead assigned to either JOPG Alpha or JOPG Charlie, to make the three JOPGs of equal size. Each JOPG consisted of seven or eight students, including one student who was appointed as JOPG Leader.

It is worth noting that participating students did not have a clear understanding in advance of what the experiment entailed. While some students had a general interest in AI and applied to be part of the J5 staff for that reason, the final distribution across the three groups was largely coincidental. Nonetheless, all students approached the experiment with openness and curiosity, and saw the opportunity to engage in an unconventional planning activity as both motivating and meaningful.

Each JOPG had two designated *Observers/Instructors* who ensured the progression of planning tasks during the exercise. In JOPG Charlie, these Instructors were 'embedded' due to the relative inexperience of the students in the use of the SOD methodology.

Artificial Intelligence-based Decision Support System

This experiment employed a purpose-built AI-based DSS, which was specifically configured to support the functions of a JOPG. A key feature was its access to a comprehensive set of scenario-related data raging from strategic directives and orders to political, military, economic, and social data pertaining to the operational environment and current in-scenario situation. While based on a commercially available model, the AI-based DSS was configured by crafted model instructions that enabled its specialization in support of the JOPGs. Both the generative AI technology and its configuration were designed and developed by NDUC to support planners through functions such as text generation, query resolution, creative prompts, and provision of feedback tailored to OPP and COPD requirements.

Methodology

The study was not methodologically homogenous. Instead, it was intentionally designed as a '*multi-scientific methodological initiative'* that allowed for the employment of different research methodologies to address its four different study areas. A key factor in the success of the experiment was the deliberate decision to not to frame it as a full-fledged scientific experiment with strict control structures, but instead to frame it as a structured, exploratory comparative design. From this perspective, each JOPG was viewed as being subject to an intentionally interventionalist approach—doctrine-based (JOPG Alpha), doctrine-plus-AI (JOPG Bravo), and SOD-plus-AI (JOPG Charlie)—rather than as being components of a conventional control group-versus-treatment group arrangement.

This framing allowed for natural interaction and fluidity between groups, acknowledging the pedagogical and social realities of the exercise. Faculty involved in the research operated simultaneously as facilitators and observers, emphasizing learning and reflection over experimental purity. This helped create an inclusive and low-friction integration of research into the headquarters simulation, which avoided the potential that students may have perceived the research interventions as intrusive or disruptive. In line with this design, the experiment's methodological approach emphasized exploration, seeking to deepen understanding of how different planning logics emerge in operational and educational contexts.

Data Collection Instruments

Data from a combination of sources was collected and later analyzed. From a quantitative perspective, the main data sources were pre- and post-experiment questionnaires, *National Aeronautics and Space Administration-Task Load Index (NASA-TLX) Cognitive Load-survey* workload scores, and time-on-task measurements. The main data sources for the qualitative perspective were Observer/Instructor notes, structured interviews, participant reflections, and daily group debriefs. Each of these instruments is explained below.

- **Pre- and post-experiment questionnaires (surveys)**. Using Microsoft Forms as a survey program, before the experiment the research team collected data from participants on their previous experience and attitudes towards AI. This established a baseline. After the experiment, the research team collected data on participants' considerations during and at the conclusion of the experiment. This included information about any changes in participant attitudes, and about their learning experiences.
- ***Observer Checklists* designed to support daily observations**. The *Observer Checklist* was a structured data instrument designed to capture systematic

observations of each JOPG during the planning sessions. It focused on key dimensions such as teamwork, planning process, use of tools (including AI), creativity and emergence, decision-making quality, and workload or stress. Observers were instructed to neutrally and consistently record their observations for each group and time phase using tick boxes and comment fields. This aimed to ensure that aspects such as group dynamics, creativity, tool interaction, and operational planning quality were documented across all groups, with particular attention given to instances of novelty or methodological innovation. Completed checklists were submitted daily to the Experiment Data Collection Manager, providing a reliable foundation for later analysis.

- ***NASA-TLX Cognitive Load-surveys*** **applied after each task was completed**. The *NASA-TLX Cognitive Load-survey* is a standardized instrument that was used in this study to systematically measure participants perceived cognitive workload after each OPP task. It evaluated six workload dimensions, with this experiment focusing on mental demand, physical demand, temporal demand, and self-rated performance. Each participant rated these dimensions on a 0–20 scale. Their combined ratings yielded an overall workload score for each JOPG. This enabled comparative analysis between JOPGs, offering insights into differences in mental effort, time pressure, frustration, and satisfaction with performance. By applying *NASA-TLX Cognitive Load-surveys* consistently, the study was able to assess how different planning or alternative methodologies, and AI support, influenced perceived workload and operational effectiveness.
- **Structured daily and post-task *interviews*.** Structured daily and post-task interviews served as a qualitative data instrument to capture participants immediate reflections. Conducted at the end of each day's tasks, these short (5–7 minute) interviews provided structured but open responses. For the JOPGs using the AI-based DSS (JOPG Bravo and JOPG Charle), three core aspects were of interest: how the AI-based DSS influenced planning decisions; whether its use led to deviations from COPD processes and methods; and how decision-making was perceived when compared to participants prior experiences without AI. By systematically collecting these reflections across all participants, the interviews enriched the dataset with firsthand insights into the way tasks and problems shaped cognitive processes, team dynamics, and methodological adaptation.
- **Daily group debriefs and *collective reflections*.** An additional qualitative data collection instrument was the conduct of daily group debriefings within each JOPG. These focused on group experiences and enabled collective reflections to be made about the same three core areas of interest discussed during individual interviews (see above). Group reflections were transcribed by research staff, and this data was used to supplement that provided by individual interviews.

The intention to use expert evaluations as an additional measurement of the 'product quality' of each JOPG's output never became operationalized during the

experiment. The reasons for this lie in several unrelated aspects of the simulation. Beyond issues relating to the staffing of the required team of expert evaluators, the main reason for this decision was issues related to lack of clarity regarding what to measure and how to make use of it.

Ethics, Reliability and Validity Considerations

This experiment was approved (i.e., *ethical clearance* was given) by the NDUC Research Committee (*Forskningsnemda ved Forsvarets Høgskole*). This approval process was also made transparent by formally informing the Data Security Office at SEDU prior to the commencement of CS25. The data collection was therefore accepted by both NDUC and SEDU respectively. The most critical aspects of the research ethics for this study were obtaining informed consent from all participants and anonymizing data. These aspects were both achieved, and measures to achieve them were applied throughout the study.

Limitations to reliability and validity include the degree of realism to which a synthetic scenario and an exercise can be compared to real experience. The study's validity also suffers from two other aspects: the constraints mentioned above (see Section 2); and the level of knowledge and prior experience of utilizing the kind of AI-based DSS that was applied in the study (e.g., there was a potential for AI-tool novelty bias). However, triangulation of data sources, along with reflections within the group of researchers and Instructor/Observers during and after the experiment, has helped to ensure the robustness of the findings contained in this report.

Study Area Analyses

This section summarizes each of this experiment's four study areas. These areas are:

- Study Area 1 – Military Planning and Decision-Making
- Study Area 2 – Comparative Analysis of Group Performance
- Study Area 3 – ICT and Pedagogical Perspectives
- Study Area 4 – Attitudes Toward AI in Military Contexts.

For each study area, discussion proceeds using the following structure:

- research focus
- key research questions
- hypotheses to be tested
- key findings
- recommendations
- remark(s).

Discussion in Study Areas 2 and 3 varies slightly from this general structure. In Study Area 2, this variation enables key findings relating to each JOPG to be addressed separately. In Study Area 3, an additional sub-section is added that discusses this area's research framework.

Study Area 1 – Military Planning and Decision-Making

Research Focus
This study area's research focus was to investigate how AI-based DSS affected the OPP at the operational level, as contained in the COPD.

Key Research Questions
The key research questions guiding data collection within this study area were:
(1) How does the AI-based DSS influence planning workflows and decision quality?
(2) Which parts of the COPD are most affected or altered?
(3) Do deviations from the traditional process suggest emergent methodologies?

Hypotheses
Research in this study area tested the following hypotheses:
(1) Hypothesis 1.1: The use of the AI-based DSS increases decision-making speed.

(2) Hypothesis 1.2: AI-based DSS integration enables higher planning accuracy and coherence.
(3) Hypothesis 1.3: AI-supported teams experience lower cognitive load and lower task frustration.

Key Findings

Research in this study area resulted in the following key findings.

(1) **Finding 1.1: Enhanced early-stage planning.** After becoming familiar with the AI-based DSS and reconfiguring how to set up the staff work, participants in JOPG Bravo consistently reported that the AI tool was particularly useful in the initial phases of planning, especially in framing the problem, brainstorming options, and structuring early understanding of the operational environment. Whether or not this aligns with prior research on synthetic cognition and sensemaking acceleration in headquarters contexts has yet to be confirmed.
(2) **Finding 1.2: Cognitive extension, not replacement**. Participants described their interactions with the AI-tool as a 'thinking partner' rather than a decision-maker. Its support expanded their cognitive bandwidth and reduced task saturation during complex OPP activities such as course of action development and political, military, economic, social, information, infrastructure (PMESII) analysis. However, the extent to which AI use affects human judgment was not verified during this experiment.
(3) **Finding 1.3: Workflow deviations with purpose**. In JOPG Bravo, the AI tool's support allowed for intentional divergence from the traditional COPD planning sequence. This included more iterative and exploratory behaviors early in the planning cycle, which were seen as productive rather than disruptive. If there is trust in both AI performance and the staff officer's own ability to give clear instructions to AI (prompting skills), then this finding indicates where there may be a clear temporal advantage. With the right level of staff training and exposure, tasks augmented by an AI-based DSS could theoretically be managed in minutes or hours instead of days or more.
(4) **Finding 1.4: Trust with boundaries**. While confidence in AI support grew throughout the exercise, users remained aware of the need for critical oversight. The tool's recommendations were most valued when paired with peer discussion amongst the group and through cross-referencing.
(5) **Finding 1.5: Team-led prompt engineering.** Success in AI interaction was highest when teams collaborated on prompt design and interpretation of output, suggesting that AI's impact is tightly coupled with intra-team communication strategies.

Given the abovementioned Findings 1.1 to 1.5, it is concluded that Hypotheses 1.1 and 1.3 are supported. However, the findings from this experiment cannot clearly indicate any specific support to Hypothesis 1.2.

Recommendations

As a result of the key findings in this study area, the following recommendations are made.

(1) Recommendation 1.1: AI tools should be strategically integrated into doctrinal planning timelines, especially for early-phase scoping and ideation.
(2) Recommendation 1.2: PME environments should train students in collaborative prompting techniques, enabling teams to leverage AI's full cognitive augmentation potential.
(3) Recommendation 1.3: Future versions of AI-based DSS should incorporate modular COPD-specific guidance, signaling the tool's optimal application points and preventing premature convergence or overreliance.
(4) Recommendation 1.4: The experiment confirmed the value of embedding AI as a conceptual framework, not just a technical assistant, which has implications for both concepts, doctrine, and system development.
(5) Recommendation 1.5: Future experiments should investigate how staff officers and team leaders claim cognitive agency of both the process (methodology) and the progress of planning tasks in a more innovative way (approach).
(6) Recommendation 1.6: Future experiments should investigate how different tools affect reflective thinking and understanding of the problem space.

Remark

JOPG Charlie did not use the AI-based DSS extensively. During two interviews done post-experiment, participants expressed that AI did not contribute to their understanding. They interacted with AI to determine factors and collect data about the situation; this data was printed and transcribed into digital platforms that supported the SOD methodology (e.g., Miro). This interaction seemed to be disconnected from the overall process of immersion into the problem space. Still, it is likely that the use of and engagement with the AI-based DSS is relational to its perceived value.

Study Area 2 – Comparative Analysis of Group Performance

Research Focus

This study area's research focus was to compare the performance of the three JOPGs (i.e., groups with and without access to the AI-based DSS) across key dimensions such as plan quality, time efficiency, innovation, and cognitive load.

Key Research Questions

The key research questions guiding data collection within this study area were:

(1) Do AI-supported teams produce better operational plans?
(2) Are AI users more confident in their decisions?
(3) Does AI support enable faster or more iterative planning?

Hypotheses

Research in this study area tested the following hypotheses:

(1) Hypothesis 2.1: JOPG Bravo and JOPG Charlie outperform JOPG Alpha in plan quality and innovation.
(2) Hypothesis 2.2: Groups using AI require less time to complete planning tasks.
(3) Hypothesis 2.3: Groups using AI report higher satisfaction and lower cognitive load.

Key Findings

Research in this study area resulted in the following key findings, which are discussed separately for each of the three JOPGs.

JOPG Alpha (Control): Used COPD but did not use the AI-based DSS. JOPG Alpha simulated a standard JOPG found in J5 branches in previous activities. It used the OPP contained in the COPD, and tools such as PowerPoint and Excel, thereby reflecting contemporary practices within the NATO alliance.

Within the framework of this experiment, we did not make any restrictions on what tools JOPG Alpha could use, but at the same time we did not provide the specific AI-based DSS that the other JOPGs were encouraged to use. As the COPD was the planning tool that the students were familiar with from their previous PME, we expected that they would use the COPD and adapt it to the set of problems they were asked to solve. The students were tutored by experienced officers in a traditional manner.

As a result of JOPG Alpha's role as the control group, no key observations have been made regarding this JOPG.

JOPG Bravo (Treatment 1): Used COPD with the AI-based DSS. JOPG Bravo provided a compelling case study in hybrid planning, combining traditional doctrinal sequencing with adaptive use of AI support. Compared to the control group (JOPG Alpha), JOPG Bravo showed clear signs of increased tempo, creative exploration and a lower cognitive load from the given tasks. Key findings relating to JOPG Bravo were:

(1) **Finding 2.1: Higher planning tempo and confidence.** Observers and participants noted a smoother transition through planning milestones once initial tool familiarization was complete. Decision confidence increased in

tandem with tool fluency, and the group met all deliverables ahead of schedule.

(2) **Finding 2.2: Cognitive load reduction.** *NASA-TLX Cognitive Load-survey* indicators suggest that JOPG Bravo experienced lower cognitive load, particularly during information synthesis and the mission analysis phase. The AI tool acted as a 'pre-processor' of complex information, reducing friction. Moreover, JOPG Bravo's planning products were perceived as the results of a team effort in which participants viewed the AI tool as a team member. The proverb '*do not marry your plan'* was never experienced. Rather, the opposite occurred, suggesting that the effort invested to produce each product was less strenuous than the effort that would have been required without AI supplementation.

(3) **Finding 2.3: Comprehensive understanding and creativity.** JOPG Bravo's planning products showed a greater internal consistency and a broader, more comprehensive understanding of the problem compared to JOPG Alpha. This was particularly evident in how the AI supported the team in connecting the Commander's intent and the overarching objectives to form a branch plan that met the specified criteria.

(4) **Finding 2.4: Collective prompting dynamics.** JOPG Bravo's group culture fostered team-based prompting, where members co-developed queries and reflected critically on the AI's outputs. This collective practice enhanced AI tool effectiveness and created a shared mental model of the problem space.

(5) **Finding 2.5: Minimal disruption to COPD flow.** Although some deviations occurred, JOPG Bravo maintained alignment with the COPD's overarching OPP progression. When deviations did occur, they were often tied to AI-enabled rethinking, rather than resulting from confusion or disorganization.

(6) **Finding 2.7: Situational awareness and digital literacy.** Participants quickly realized that the AI tool's responses required doctrinal triangulation and factual scrutiny. This led to an emergent peer review-type behavior where outputs were validated through group deliberation and feedback from prompting in a secondary thread (cross-validating).

JOPG Charlie (Treatment 2): Used SOD with the AI-based DSS. Participants in JOPG Charlie demonstrated a marked evolution in their approach to planning. Initially faced with considerable uncertainty, particularly due to the lack of a predefined method or template, they gradually embraced the principles of SOD. Working with tools such as *giga mapping* and *iterative stakeholder mapping*, they began to understand and visualize complexity as something to be preserved and deepened, rather than simplified. Key findings relating to JOPG Charlie were:

(1) **Finding 2.8: Staying in the problem space.** JOPG Charlie's SOD methodology emphasized staying in the problem space rather than rushing toward solutions, a dynamic that was at times frustrating but ultimately illuminating.

Over time, students expressed that this methodology not only helped them make sense of the scenario but also retain and build upon earlier insights, especially when visualized using digital platforms like Miro.

(2) **Finding 2.9: Collective learning.** JOPG Charlie's approach fostered not just individual insight but also a strong sense of collective learning. The collaborative nature of SOD meant that learning occurred through shared exploration, dialogue, and co-construction of meaning. Participants gradually shifted from seeing learning as something individual and internal to understanding it as emergent, distributed, and socially embedded.

(3) **Finding 2.10: Divergence vs convergence.** Seen from the workspace of JOPG Charlie, there were clear cultural signals worth noting. Experimental work related to SOD gained legitimacy over time, particularly due to previous interventions connected to the joint operations planning course that occurred in the lead up to CS25. Many personnel expressed a deeper understanding of—and value in—exploring SOD as well as AI-supported approaches. Still, some underlying skepticism persisted. SOD was occasionally perceived as incompatible with military planning due to its divergent and exploratory nature, which appeared to challenge the perceived need for doctrinal convergence within a headquarters. Concerns included that SOD might be time-consuming, misaligned with the battle rhythm, or may have produced plans that deviated from organizational expectations. However, these concerns often stemmed from unfamiliarity with the methodology.

(4) **Finding 2.11: Inviting curiosity.** Interestingly, JOPG Charlie acquired a unique identity within CS25's simulated CJTF headquarters. Often humorously referred to as the 'Steiner school' or the 'arts and crafts group', JOPG Charlie developed a visible and somewhat mythical status—being both different and intriguing. This outsider positioning invited curiosity, prompting others to visit and engage with the group. A blend of mild skepticism and genuine interest resulted in a flow of informal interactions that enriched the experiment's relational and cultural dimensions. Leadership figures—including the CJTF Commander—described JOPG Charlie in briefings as a 'think tank', recognizing its role in generating alternative perspectives as being valuable to their own conceptualizations of the problem and resulting decision-making. While some external visitors unfamiliar with the experiment expressed surprise at JOPG Charlie's presence within a military headquarters, this was a minority view. On the whole, the experiment was met with curiosity, strategic interest, and an emerging institutional acceptance of experimentation as a legitimate educational and operational activity.

(5) **Finding 2.12: More creative.** Members of JOPG Charlie expressed that they felt like they had been more creative than the two other groups. This sparked a deep debate with members of JOPG Bravo, who felt they had been equally creative. This raised an interesting question: have the different JOPGs been

more or less creative than each other? Corollaries of this question include, have different JOPGs been creative in different ways, or have the tools (e.g., in JOPG Charlie's case using more visual and design methods) created a false sense of creativity? Identification of these questions indicates a potential area for future research.

Given the abovementioned Findings 2.1 to 2.12, it is concluded that Hypotheses 2.2 and 2.3 are partly supported. However, the findings from the experiment do not indicate any clear support for Hypothesis 2.1. In light of Study Area 2's focus on individual JOPGs, some qualifiers are required regarding these conclusions.

- Qualifier regarding Hypothesis 2.1 (not supported): The findings provide no clear support for this hypothesis because no expert panel was assembled, and no blind test was conducted to impartially compare planning products produced by each JOPG. That said, JOPG Bravo's products demonstrated internal consistency, comprehensive understanding, and integration of the Commander's intent. Their hybrid use of COPD and an AI-based DSS yielded higher tempo without sacrificing doctrinal coherence. Similarly, JOPG Charlie demonstrated creativity and innovation through SOD, emphasizing problem exploration, collective learning, and divergent thinking. Therefore, while Hypothesis 2.1 is not supported by this experiment, it has not been refuted either. This presents an opportunity for future experiments to test this hypothesis.
- Qualifier regarding Hypothesis 2.2 (partly supported): JOPG Bravo clearly experienced time efficiency gains. Their higher tempo, earlier completion of deliverables, and smoother progression through planning milestones demonstrated that AI integration reduced delays and accelerated planning. However, this acceleration only occurred following an initial period of familiarization with the AI tool. On the other hand, JOPG Charlie did not plan faster. Their commitment to staying in the problem space and embracing SOD meant that their process was more time-intensive and exploratory. This divergence highlights that AI can increase tempo when paired with familiar methods (JOPG Bravo) but may not yield time savings when integrated into exploratory or divergent approaches (JOPG Charlie).
- Qualifiers regarding Hypothesis 2.3 (partly supported): The evidence again shows differentiated outcomes. For JOPG Bravo, *NASA-TLX Cognitive Load-survey* data and observer feedback both confirmed reduced cognitive load, higher confidence, and stronger satisfaction compared to the control group (JOPG Alpha). The AI-based DSS functioned as a cognitive aid, reducing friction in information processing. JOPG Charlie, while reporting higher levels of creativity and collective learning, also experienced frustration and uncertainty early in the process, reflecting the challenges of SOD's divergent methodological structure. Satisfaction increased over time as the group embraced SOD, but no systematic data indicated reduced cognitive load

compared to COPD Alpha. Instead, the value of their approach seems to lie more in conceptual richness and cultural impact than in reducing workload.

Taken together, these findings related to Hypotheses 2.1, 2.2 and 2.3 suggest that AI support improves planning effectiveness, confidence, and efficiency when embedded within familiar doctrinal processes, but yields more experimental, creative, and exploratory benefits when combined with a novel methodology (SOD). This underscores that AI's impact is contingent on broader circumstances, in this case choice of planning or alternative methodology. It can accelerate established processes or deepen exploratory methodologies, but it cannot necessarily achieve both acceleration and increased depth simultaneously.

Recommendation
As a result of the key findings in this study area, the following recommendation is made:

(1) Recommendation 2.1: Future research should be conducted to further explore the extent to which different tools and methodologies encourage different levels of creativity, for example by employing single/double/triple loop learning systems, or complexity theory.

Remark
As discussed above in relation to Study Area 2's hypotheses, the intended comparison of the respective JOPGs' solutions to the problem had to be cancelled due to several different reasons, including staffing (no ability to stand up an expert panel), and minor interventions in the tasking to each JOPG leading to different interpretations early in the activity. However, a blind test of JOPG outcomes should be pursued during Exercise Comprehensive Shield 2026.

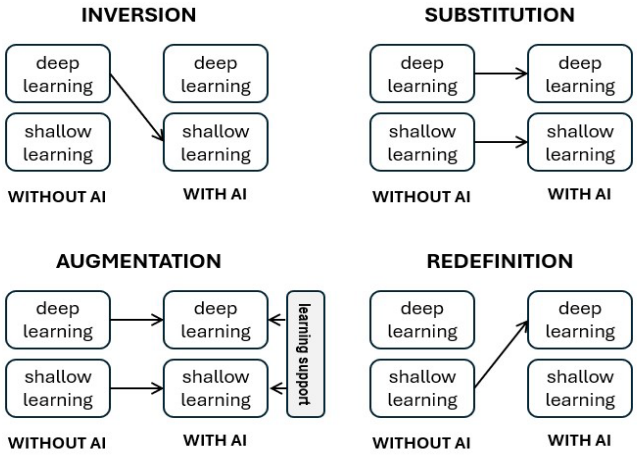
Study Area 3 – Information and Communication Technology and Pedagogical Perspectives

Research Focus
This study area's research focus was to assess how AI tools can function as pedagogical assistants during complex staff exercises.

Research framework
The use of AI in education can have various effects, from hindering learning to completely transforming it for the better. The ISAR model is a framework that categorizes AI's effects into four main types: *Inversion*, *Substitution*, *Augmentation*, and *Redefinition* (see Figure 1).

- ***Inversion (I, negative impact)*:** This occurs when the use of AI, even though it is meant to help, makes learning worse or more superficial. For example, if a student uses an AI tool like ChatGPT to write an entire report, they might not engage as intensely with the source material. This can lead to 'cognitive laziness', where the student avoids deep and critical thinking and thereby undermines basic prerequisites for learning.
- ***Substitution (S)*:** This occurs when AI replaces an existing teaching method or task but does not change the core learning process. For example, AI could generate seminar questions that are just as good as those a teacher would create. The learning content is similar, but AI makes the creation process faster or easier, and the teacher thereby gets more time to support students.
- ***Augmentation (A)*:** This occurs when AI enhances learning by providing extra support that was not available without it, or by making existing support better, thereby improving the learning experience without changing the fundamental task. An example of this is an AI system that can identify patterns in large amounts of data or adapt the difficulty of practice problems in real-time based on students' previous results. This kind of cognitive support can help the student understand and remember things more efficiently, freeing cognitive power that can be used to solve more complex problems.
- ***Redefinition (R)*:** This occurs when AI enables completely new types of learning tasks that are very difficult or impossible without it. These new tasks encourage much deeper, more creative, and interactive learning. Examples include complex simulations and conversations with AI-driven personas.

Figure 1. *The ISAR model of effects of AI in education*[3]



[3] Figure 1 derived from information contained in: Bauer, E., Sailer, M., Greiff, S., Graesser, A.C. and Scheiter, K. (2025) 'Looking Beyond the Hype: Understanding the Effects of AI on Learning', Educational Psychology Review, 37(45), pp. 1–27.

In this study, we have used the ISAR model in combination with research on cognitive science to support hypotheses development and discussions. In very brief terms, students must be active in their learning and engage with the content that they are meant to learn to be able to develop and consolidate knowledge. Thus, if the use of AI reduces the student's engagement with the material they are supposed to learn (*I*), this can result in less learning and retention. On the other hand, if AI can be used to reduce students' extrinsic and intrinsic cognitive load, and enable deeper studies of content, it can instead support learning (*A/R*).

Key Research Questions

The key research questions guiding data collection within this study area were:

(1) Does the use of AI-based DSS foster learning and knowledge retention, and how does it affect students' engagement with the COPD methodology?
(2) How does the use of AI-based DSS affect students' cognitive workload and stress levels when working with complex tasks?
(3) To what degree could AI support complement PME curricula and teaching methods for enhanced learning?

Hypotheses

Research in this study area tested the following hypotheses:

(1) Hypothesis 3.1: The use of the AI-based DSS will augment learning due to cognitive support that enables deeper thinking, but there is also a risk for inversion regarding the COPD methodology itself.
(2) Hypothesis 3.2: Teams working with complex questions will experience lowered cognitive workload when using AI to augment their inquiries.
(3) Hypothesis 3.3: If the AI-based DSS is exploited in an augmentative way it will complement current PME curricula and teaching methods in a way that enhances learning.

Key Findings

Research in this study area resulted in the following key findings.

(1) **Finding 3.1: AI-based DSS as a learning enhancer and its impact on methodology**. After an initial period of confusion regarding the AI system's functionality, students reported increased learning outcomes and deeper reflection in their work. The AI tool helped them analyze large datasets and identify relevant focus areas, which strengthened knowledge retention and understanding. However, it was observed that, especially in the beginning, attention shifted from the COPD methodology to understanding the AI tool, which in some cases led to the AI tool overshadowing the methodology itself. This highlights the importance of balancing AI use with conscious engagement in core methodological principles.

(2) **Finding 3.2: AI-based DSS and cognitive workload**. JOPG Bravo encountered initial challenges in developing an understanding the AI tool, resulting in time spent figuring out how it worked. However, they eventually experienced lower cognitive workload and reduced stress compared to other JOPGs, after organizing training and receiving support in use of the AI tool. Guidance from both the Instructor and an AI expert was crucial in reducing their cognitive load. Students reported that AI helped them manage complex questions efficiently and that they felt less overwhelmed as they became more familiar with the AI tool. Peer encouragement and supportive leadership helped alleviate stress and foster collaboration. Overall, the experiment underscores the importance of teamwork and adaptive strategies in educational settings involving new technology.

(3) **Finding 3.3: AI-based DSS as a complementary tool to PME curricula and methods.** When the AI-based DSS was used to supplement existing educational programs, students experienced enhanced learning outcomes. The AI tool contributed to complementing and strengthening current pedagogical methods, especially in addressing complex and interdisciplinary problems. The use of AI brought advantages such as reduction of emotional bias and the provision of data-driven insights, which improved the JOPG Bravo's efficiency. Participants successfully created and refined three operational designs, leading to quicker workflows and expanded thinking. As their knowledge of AI grew, they developed a nuanced appreciation for its capabilities and limitations and recognized the essential role of human oversight in validating AI outputs. Members of JOPG Bravo also expressed confidence in the AI's recommendations regarding the task, having meticulously tracked all inputs and outputs. This contributed to a high-quality product delivered in a short timeframe, and their Instructor/Observer acknowledged the group's rapid progress and quality improvement. Members of JOPG Bravo also proposed the implementation of AI across the headquarters to further optimize its operations.

(4) **Finding 3.4: Identification of areas for future study.** Participants expressed that AI provided advantages such as reducing mental load and enabling deeper analysis (*A*), but concerns arose regarding dependency and the risk of losing foundational knowledge (*I*). They felt that AI can support learning activities, but that a balance between its use and human oversight is crucial for fostering creativity and effective problem-solving. In future studies, it will be important to also measure learning and retention among students. Kirkpatrick's four levels of training evaluation could be suitable in this context, as they take into consideration students' reactions, learning, and behavioral changes, as well as organizational performance.

(5) **Finding 3.5: Overall impact on learning and retention.** In the beginning of the exercise, learning about the AI system distracted students from the task

and increased their cognitive load. But when they had figured out its functionality, they used the AI tool to analyze large amounts of data and find areas to focus on and study more deeply.

Given the abovementioned Findings 3.1 to 3.5, it is concluded that Hypotheses 3.1, 3.2 and 3.3 are all supported.

Recommendations

As a result of the key findings in this study area, the following recommendations are made.

(1) Recommendation 3.1: Future studies should use Kirkpatrick's four levels of training evaluation to measure student learning and retention in addition to the areas that were measured in this study.
(2) Recommendation 3.2: Students should be provided with the necessary time and support to ensure they fully understand the functionality of AI tools that are used during learning activities. This should occur before activities commence.
(3) Recommendation 3.3: AI should be used more broadly within PME in support of group problem-solving and to augment collective learning, as its use frees time and cognitive capacity for peer-to-peer student discussions and deeper learning.

Remark

Observations, questionnaires and interviews conducted in JOPG Bravo were used to collect data and study factors that are known to have an impact on learning and memory retention. As JOPG Charlie never fully used the AI-based DSS as an integral support to their SOD methodology, they are not included in Study Area 3's analysis.

Study Area 4 – Attitudes Toward Artificial Intelligence in Military Contexts

Research Focus

This study area's research focus was to examine participant trust in AI, ethical considerations, perceived risks, and readiness for AI integration in the operational environment.

Key Research Questions´

The key research questions guiding data collection within this study area were:

(1) How do users assess the reliability and trustworthiness of AI in planning?
(2) What concerns (ethical or practical) emerged?
(3) Do participants envision future roles for AI in C2 structures and PME?

Hypotheses

Research in this study area tested the following hypotheses:

(1) Hypothesis 4.1: Members of JOPG Bravo and JOPG Charlie will show greater enthusiasm and trust towards AI due to their exposure to it during the study.

(2) Hypothesis 4.2: Members of JOPG Bravo and JOPG Charlie will perceive a higher impact of AI on military operations and education in the future.

Key Findings

Research in this study area resulted in the following key findings.

(1) **Finding 4.1: Comfort and enthusiasm.** Overall, participants were moderately comfortable using AI, with notable variation across groups. JOPG Bravo demonstrated the highest enthusiasm and confidence, even when encountering technical challenges, while JOPG Alpha and JOPG Charlie showed more cautious engagement.

(2) **Finding 4.2: Trust and reliability.** Trust in AI varied widely—from skepticism (minimal reliance) to strong confidence (perceiving AI as a 'digital partner'). Many participants assumed the role of 'peer reviewers', cross-checking AI outputs for accuracy. This indicates that trust was conditional and linked to both AI tool performance and the team's prompting skills.

(3) **Finding 4.3: Ethical concerns.** The most common concern was the *loss of human control*, with strong emphasis on maintaining human responsibility in decision-making. Participants highlighted accountability challenges if AI were to assume a larger role in planning or execution. An example of this is indicated by the following statement made by a participant in JOPG Bravo: *"Om AI tar øver planeringen och genomførandet av militæra operationer i før høg utstræckning ær ansvarsfrågan en utmaning"*.[4] The need for humans as the final decision authority was repeatedly stressed – at least a need for humans to make the final decision and act as a filter. The need to be critical and evaluate the information the AI brings forth was also a recurring theme.

(4) **Finding 4.4: Perceived future impact.** Across all JOPGs, participants expressed a strong belief that AI will have significant impact on future military operations and education (see Figure 2). Post-exercise questionnaires indicated an upward trend (see Figure 3) in perceived impact on PME (mean increase from 4.19 to 4.50 on a 5-point scale). While JOPG Alpha's view of the utility of AI in operations slightly declined, JOPG Bravo and JOPG Charlie indicated stable or increased expectations.

[4] English translation of the Norwegian citation: "If AI takes over the planning and execution of military operations at a high level, then the issue of responsibility will become a challenge."

Figure 2. *Change in perceived impact of AI on future military operations*

Pre-exercise questionnaire

Group	Mean value	Mode value
A	4,58	5
B	4,78	5
C	4,40	5

Post-exercise questionnaire

Group	Mean value	Mode value
A	4,14	4
B	4,78	5
C	4,60	5

Figure 1 notes. A comparison between the answers (before and after) of all JOPGs on the question: How much impact will AI have on military operations in the future? (1 = low and 5 = high). The data in Figure 2 shows that the general belief is that AI will have a great impact on military operations in the future. JOPG Alpha showed the greatest difference between pre- and post-exercise with a 0.44 lower mean value. JOPG Charlie showed a 0.2 higher mean value post-exercise whereas JOPG Bravo was the group with the highest mean value both pre- and post-exercise.

***Figure* 3.** *Change in perceived impact of AI on future PME*

Pre-exercise questionnaire

Group	Mean value	Mode value
A	4,00	4
B	4,11	4
C	4,80	5

Post-exercise questionnaire

Group	Mean value	Mode value
A	4,00	4
B	4,67	5
C	4,80	5

Figure 2 notes. A comparison between the answers (before and after) of all JOPGs on the question: How much impact will AI have on military education in the future? (1 = low and 5 = high). The overall values indicate that the participants perceived that AI will have a greater impact in military education in the future.

Given the abovementioned Findings 4.1 to 4.4, it is concluded that Hypotheses 4.1 and 4.2 are supported.

Recommendations

As a result of the key findings in this study area, the following recommendations are made:

(1) Recommendation 4.1: Future studies should track how participant attitudes evolve across settings (operational vs educational) and among officers with differing levels and types of prior experience.

(2) Recommendation 4.2: Future studies should also investigate factors such as age, gender, years of service and perceived technological competence, and the impact these have on attitudes towards AI use in military contexts.

Remark

Due to limited research resources, the measurement of attitudes was restricted to analyzing questionnaires post-experiment. More rigorous scales and longitudinal tracking will be considered to strengthen future analyses.

Forward Outlook

Recommendations for Exercise Comprehensive Shield 2026

Building on the experiences and findings from CS25, several practical recommendations can be made for the design and execution of Exercise Comprehensive Shield 2026.[5]

(1) **Recommendation ED.1: Early familiarization and training.** Participants should receive structured introductions to the AI-based DSS prior to the exercise. This should include hands-on workshops in collective prompting, interpretation of outputs, and critical validation techniques. Such preparation will reduce initial tool-related friction and allow groups to engage more effectively in planning tasks from the outset.

(2) **Recommendation ED.2: Focused experimentation.** A more focused research approach requires streamlining the experimental setup. For Exercise Comprehensive Shield 2026, and in relation to exploring AI-based DSS more thoroughly, all JOPGs included in the experiment must use the same COPD planning methodology, to enable comparisons between them to be made with more precision.

(3) **Recommendation ED.3: Diverse methodologies.** Exercise Comprehensive Shield 2026 should continue to explore a diversity of alternative methodological approaches beyond the COPD planning methodology. However, this should occur outside of the scope of this AI-based DSS exploration.

(4) **Recommendation ED.4: Cross-JOPG exchange.** Opportunities should be built into the exercise for deliberate cross-JOPG exchange at both midpoint and conclusion. A dedicated 'fusion phase' could allow teams to compare, merge, or critique outputs, thereby deepening both learning and methodological reflection.

(5) **Recommendation ED.5: Pedagogical integration.** Structured debriefings should more explicitly address how AI affected decision-making, creativity, and cognitive load. This will help participants develop not only operational competence but also meta-cognitive awareness of AI's role in planning. Reflective sessions could be complemented by facilitated dialogues between instructors, observers, and students.

[5] To ensure deconfliction from recommendations made in relation to each of the four study areas discussed in Section 5, these recommendations are numbered using the prefix 'ED'. In this context, ED stands for 'exercise design'.

(6) **Recommendation ED.6: Technical and Logistical Support.** The AI-based DSS infrastructure should be further stabilized and supported by clear user guidance materials. Redundancy measures (backup servers, pre-loaded datasets) should be established to mitigate potential technical disruptions. Technical facilitators embedded in each JOPG may further reduce barriers to effective tool use.

Suggested Future Research Directions

The CS25 experiment has raised several important questions that merit further investigation in future cycles of experimentation and PME research. Longitudinal studies should be conducted to explore how officers *trust* in AI evolves over repeated use. Particular attention should be given to *calibrated trust*, i.e., conditions under which officers appropriately rely on or reject AI-generated outputs.

Future research should employ *systematic evaluation frameworks*—such as blind expert panels to assess plan quality and Kirkpatrick's four levels of training evaluation—to produce more robust assessments of both operational and pedagogical effects.

The comparative dynamics between doctrinal (COPD) and systemic design (SOD) approaches deserve further exploration. A key research question is whether different methodologies foster distinct forms of creativity and problem framing, and how these differences affect perceived plan quality. However, this is a different research question than that of this study, and it may therefore require a different research design and setup. As declared in this report, JOPG Charlie never really came to use the AI-based DSS, as they emerged into another form of thinking that never allowed AI-based DSS to intervene.

Further testing should be undertaken during staff exercises and, where appropriate, in operational headquarters contexts. This will help determine to what extent insights from educational settings scale to real-world C2 structures.

Finally, future studies should examine the ethical and doctrinal implications of AI-enabled planning, including issues of accountability, responsibility, and human oversight. Such studies should aim to inform NATO doctrine and PME curricula with evidence-based recommendations.

"Om AI tar øver planeringen och genomførandet av militæra operationer i før høg utstræckning ær ansvarsfrågan en utmaning."[6]

(Participant in JOPG Bravo)

[6] English translation of the Norwegian citation: "If AI takes over the planning and execution of military operations at a high level, then the issue of responsibility will become a challenge."

Acronym List

AI	artificial intelligence
C2	command and control
CJTF	combined joint task force
COPD	Comprehensive Operations Planning Process
CS25	Exercise Comprehensive Shield 2025
DSS	decision support systems
ED	exercise design
ICT	information and communication technologies
ISAR	inversion, substitution, augmentation, and redefinition
JOPG	Joint Operations Planning Group
NASA-TLX	National Aeronautics and Space Administration-Task Load Index
NATO	North Atlantic Treaty Organization
NDUC	Norwegian Defence University College
OPP	Operations Planning Process
PME	professional military education
PMESII	political, military, economic, social, information, infrastructure
SEDU	Swedish Defence University
SOD	Systems-Oriented Design

About the Contributors

Dr Peter Bovet Emanuel *is a Military Teacher in War Studies at Swedish Defence University. He is also a Lieutenant Colonel (OF-4) in the Swedish Special Forces. His research focuses on AI for future military operations and military innovation.*

Ørjan Rogne Rise *is a Military Teacher at Norwegian Defence University College. He is also a Lieutenant Colonel (OF-4) in the Norwegian Army, and he holds a master's degree in the Study of Professions. His research interests are in PME and developing pedagogics and didactics.*

Stig Tore Aannø *is a Military Teacher at Norwegian Defence University College. He is also a Lieutenant Colonel (OF-4) in the Norwegian Air Force, and he holds a master's degree in military sciences. His research interest lies in military innovation, AI and multi-domain operations.*

Dr Hanna Forsberg *is a Senior Advisor in ICT at Swedish Defence University. She holds a PhD in cell-biology and genetics. Her research interest is in ICT pedagogics.*

Professor Anders McDonald Sookermany *is Professor in Pedagogics at Norwegian Defence University College. He is also a Lieutenant Colonel (OF-4) in the Norwegian Air Force. His research focuses on future PME concepts.*

Thomas Slensvik *is a PhD Candidate at Oslo School of Architecture and Design. He is also a Commander (OF-4) in the Norwegian Navy. His research interest is in Systemic Operational Design.*

Martin Gurák *is a Teaching Librarian at Swedish Defence University. He holds a bachelor's degree in library and information science.*

Dr Aaron P. Jackson *is Senior Lecturer in War Studies at Swedish Defence University. His research focuses on operational art, military planning methods, and military design thinking.*