Visual Analytics for Human-Centered Artificial Intelligence

Valérie Lavigne
Defence R&D Canada
CANADA
valerie.lavigne@forces.gc.ca

Dr. Margaret Varga
Seetru
UNITED KINGDOM
margaret.varga@seetru.com

Antti Kortemaa
Finnish Defence Research Agency
FINLAND
antti.kortemaa@mil.fi

Georgi Nikolov
Cyber Defence Lab, Royal Military Academy
BELGIUM
g.nikolov@cylab.be

Dr. Carsten Winkelholz
Fraunhofer FKIE
GERMANY
carsten.winkelholz@fkie.fraunhofer.de

Adelica Ndoni
NCIA
NETHERLANDS
adelica.ndoni@ncia.nato.int

Cyril Ray
Naval Academy
FRANCE
cyril.ray@ecole-navale.fr

ABSTRACT

This position paper outlines a framework for integrating Human-Centered Artificial Intelligence, Human-Machine Teaming and Visual Analytics in which information foraging and sensemaking are supported by iterative user feedback. We discuss how this integration can enable safe and efficient AI tools for data analysis and decision support in military applications from a complex system perspective where the dynamic interaction of diverse data, interconnected systems, and emergent characteristics, are too large and complex to analyse and understand without machine support, and the research challenges of this integration.

Keywords: Visual Analytics, Interactive Visualization, Human-Centered Artificial Intelligence, Human-Machine Teaming, Data Analysis, Sensemaking, Trust, Security, Robustness, Cyber Security, Intelligence Analysis, Situation Awareness, Situation Understanding, Explainable AI, Complex System.

INTRODUCTION

The application of Visual Analytics (VA) enablers to Human-Centered Artificial Intelligence (HCAI) can enhance the utilisation of Artificial Intelligence (AI) for data analysis, information foraging, and sensemaking in support of military operations with timely, cost-effective, trustworthy, secure, and relevant information, while accelerating advances and adoption of AI capabilities which maximize a safe and responsible human-machine symbiosis.

In this paper, we first address the rationale for this position and explain the NATO need for responsible and trusted AI. Then, we discuss Human-Centered AI and why VA is a strong proposition to support AI security, robustness, assurance, and explainability. Finally, we discuss the research challenges that remain for the development of Visual Analytics Enablers for HCAI and their assessment in the context of NATO.

RATIONALE: RESPONSIBLE AND TRUSTED AI FOR NATO

AI is recognized as a key enabler for NATO, and the urgency of adopting new AI capabilities is reflected in the recently revisited NATO AI Strategy: "It is vital for NATO to use these technologies, where applicable, as soon as possible" [1]. The NATO AI Review Chief Scientist report [2] states that it will require human-machine symbiosis, which involves addressing the multi-faceted topic of interactions between humans and automated processes, otherwise risking that modes of reasoning developed by AI systems and operators may not be aligned and mutually beneficial.

Shifting Roles and Automation for NATO Stakeholders

The rapidly evolving capabilities of Artificial Intelligence (AI) led scientists [3] to reexamine the roles of humans and machines as teammates (Human-Machine Teaming (HMT)) in information foraging and sensemaking tasks, with the vision where Visual Analytics acts as an interface between humans and machines, enabling tightly coupled teaming for data analysis and informed decision making. Effective teaming requires tight feedback loops. In a spectrum of division of labour that ranges from two extremes, from manual control to full automation, we are especially interested in tasks that fall in the middle. We focus on AI support where users do not delegate a task to AI, but rather perform it jointly in a continuous dialogue, likely involving recommender systems and human approval of actions, where humans continue to be the decision-making authority, and importantly, requiring both human and machine analysis.

Within NATO, there are multiple classes of stakeholders that can be associated with AI systems. In general, besides users employing and operating AI systems, other classes of stakeholders (such as system developers, affected parties, deployers and regulators [4]) can also be associated with them. For this paper and within the context of HCAI and HMT, NATO stakeholders directly interacting with AI systems in any capacity are considered relevant. Today, increasingly more and more systems utilise AI in one way or another, it won't be merely system developers and NATO users in more technical roles (e.g., system operators and analysts) that need to be acknowledged, decision-makers at different levels of NATO can also be expected to employ / interact with AI systems to varying degrees as a part of their duties.

DESCRIPTION: VISUAL ANALYTICS ENABLERS FOR HCAI

Human-Centered AI

HCAI "focuses on amplifying, augmenting, and enhancing human performance in ways that make systems reliable, safe, and trustworthy" [5]. To achieve this, HCAI [6][7] advocates a paradigm where software designs provide users with high levels of understanding and control over their AI-enabled tools to preserve human agency. AI thereby plays a supportive role, amplifying human ability to work in masterful ways. Rather than replacing human decision-making, HCAI emphasizes the importance of designing AI systems that complement and enhance human expertise. This approach prioritizes transparency, interpretability, and user empowerment, ensuring that individuals remain actively engaged in critical thinking and judgment throughout the analytical process.

Visual Analytics Support

Interactive visualization [8], *i.e.* visual representations of data to aid cognition, is a key enabling technology for HCAI which prioritizes human values and agency. Even though visualization cannot address all aspects of trust, it is crucial for human agency in AI [9]. Although visualization concerns and design guidelines exist for general use, there is a need to extend this work to address the unique challenges in the application HCAI

from a military perspective, specifically in improving trust and reducing the impact of trust misalignment. VA is effective in facilitating data integration and reducing the cognitive workload to understand machine learning analysis results, especially for detecting the expected ("known knowns") and discovering the unexpected ("unknown unknowns").

Human in the loop AI couples AI models' efficiency with human knowledge, intuition, and experience to guide / correct the learning process and its resultant outputs (see Figure 1). This is particularly important as visual analytics supports users to make sense of complex data by combining interactive visualization with analytics algorithms.

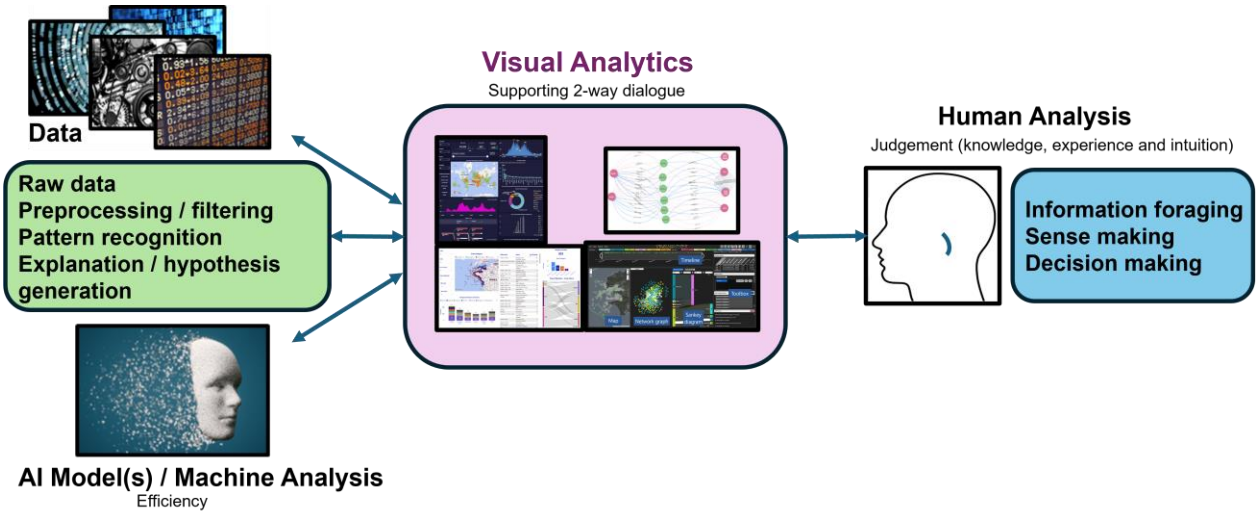

Figure 1: Visual Analytics for Human-Centered AI Framework.

Enabling Robust and Secure AI with Visual Analytics

Ensuring the robustness and security of AI components is a major challenge. Robustness is a traditional core quality of AI components and means that it provides correct and useful statements even in changing environments or with new, unexpected input data. This property requires that these systems do not simply follow rigid rules or are only able to react to known patterns but rather generalize as much as possible from the training data. Security, on the other hand, is a new field that has only recently gained greater focus and means that AI is protected against deliberate manipulation and can operate reliably. There are various attack vectors here, some of which originate from traditional cybersecurity. It is therefore important that the models and training data are protected also from a cybersecurity perspective, and the three main cybersecurity goals of confidentiality, integrity, and availability of the models and training data are ensured. In the context of AI there are additional attack vectors [10], such as adversarial attacks, in which, exploiting knowledge of the functionality of an AI component, deliberately subliminally manipulated input data can be presented, which leads to erroneous behaviour and impacts the decision making process. Such attacks are even more critical in the context of continuous learning, where AI components continue to learn during use in order to adapt to changing environments. In this sense, robustness and security are competing goals in some situations.

Currently, approaches often work with either the developer or the end user controlling the output of the AI components. For this purpose, explainable AI methods are developed and applied. However, this approach is problematic for two reasons. Firstly, the end user ideally wants to be able to rely on the statements of AI components as they consider that the AI components are meant to be tools supporting them in their work or take over subtasks. Even though eXplainable AI (XAI) can increase trust in AI statements, they are often only a necessary crutch for the end user. Of course, in high-stakes, complex decision scenarios (such as a

medical diagnosis or an aeronautics system) having the ability to interrogate the AI's reasoning or engage in a dialogue with the AI is important and can be lifesaving. But in those cases, XAI must be part of the system's design from the ground up, not a Band-Aid on an unreliable model. Secondly, developers of AI-components cannot ensure correct functionality on their own, as they often do not know the ultimate application context and may not be allowed to know it in a security-critical context.

Based on this analysis, there is a strong case for handling aspects of AI robustness and security in line with the processes of classical cybersecurity. Here, end users of information technology and its developers do not communicate directly, nor do developers or users monitor the components of their information system themselves. For this purpose, data is collected, analysed, and evaluated centrally in the company in order to identify unusual and suspicious behaviour, assign causes, and thus detect attacks. The information converges in the so-called Security Operation Center (SOC). A SOC is the central office that continuously monitors an organization's security posture, detects attacks early, coordinates incidents, and continuously improves resilience. It brings together people, processes, and technology. Especially as information systems are equipped with more and more AI components in the foreseeable future, it makes sense to combine AI monitoring and cybersecurity. The objectives and methods are thus subject to similar requirements.

In the SOC, human operators who make decisions are a central component because they counterbalance the human actors on the attacker side. Even though attackers and defenders are increasingly supported by technology and their own AI components, it is human actors who define the attackers' objectives and represent a large unknown factor. Furthermore, cyber systems especially with AI-components and its interaction with humans should be addressed as complex social technical systems [11]. Defining and implementing resilient countermeasures within such a system needs to consider overarching goals and whether decisions with high responsibility need to be made. SOC operators analyse and prepare data for decision making and escalate to higher authorities if necessary. To provide SOC operators with a comprehensive picture of the situation and the AI components in use, AI-aware Security Incident Event Management Systems (SIEMs) could serve as a central interface to SOC operators, analogous to cybersecurity.

Visual analytics plays a crucial role here, enabling operators to interactively analyse data and understand relationships through informative visualizations. The requirements for the data collected in the SIEM are new and need to be expanded; other metrics that also affect the input data, such as data and concept drift need to be considered. In addition, an analyst must also have basic knowledge of the functionality of the AI models used to assess incidents. The visual interface must also enable this. Visual Analytics should be combined with system-based approaches like Ecological Interface Design (EID) [12]. Following the EID principles, interfaces should reflect the broader work domain: from organizational goals and hazards down to tasks, data streams, and model components [13]. It should also reflect information about the overall system both in the creation of AI models in the MLDevOps (Machine Learning Development and Operations) cycle and in the operational environment and must be prepared in the form of abstraction hierarchies. Consequently, the SIEM-facing VA for operators has different requirements than developer dashboards that focus on training datasets and model internals. After triage and assessment, SOC operators can coordinate mitigations through operations or development: e.g., block or throttle abusive sources, roll back or hot-patch models, rotate keys and secrets, adjust guardrails and policies, quarantine suspect data, schedule targeted retraining, or open a formal incident with post-mortem and hardening actions.

Figure 2 illustrates this perspective and the role of the various VA / User Interface (UI) in the secure operation of information systems with AI components.

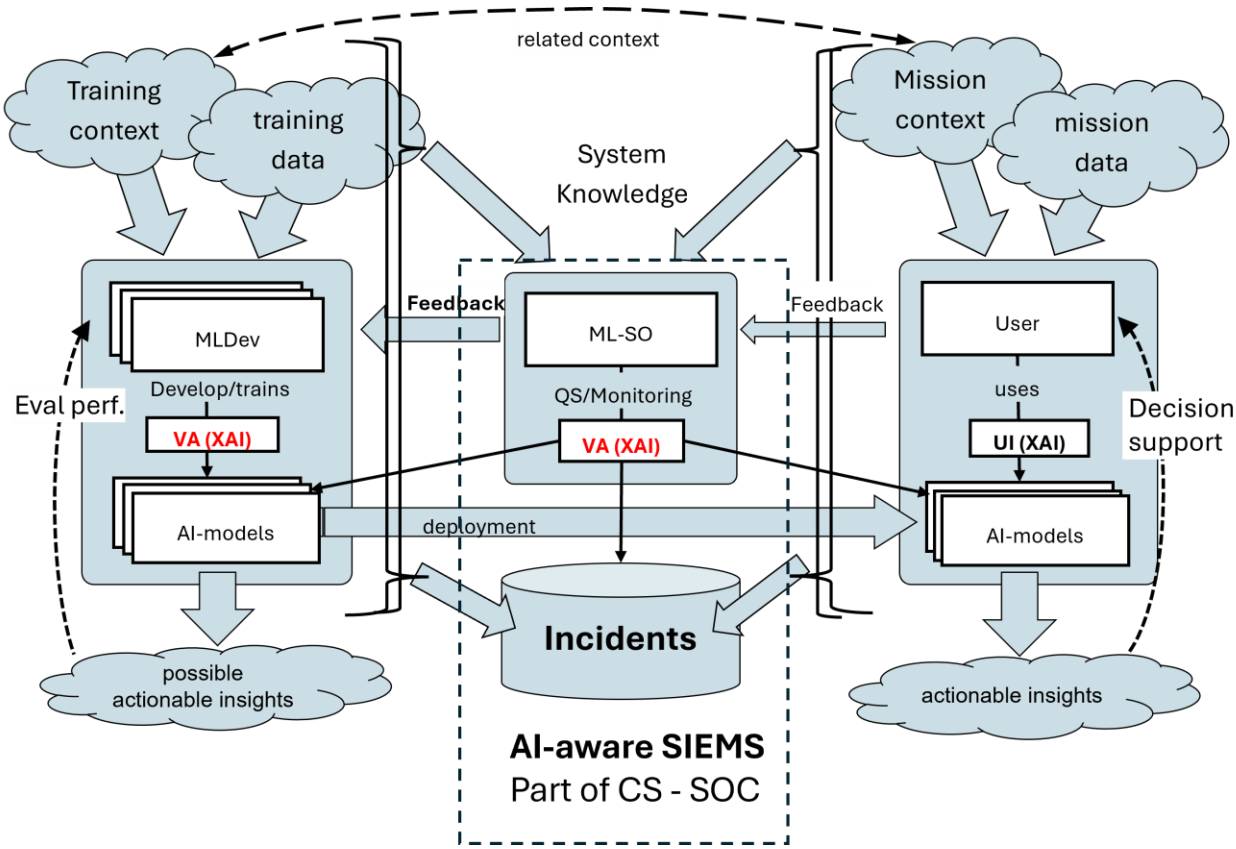

Figure 2: Enabling robust and secure AI through SOC and VA-Interfaces.

Enabling AI Assurance and Explainability with Visual Analytics

Development of VA supported AI capabilities for NATO will be expected to follow the six Principles of Responsible Use (PRUs) of AI in Defence – (1) lawfulness, (2) responsibility and accountability, (3) explainability and traceability, (4) reliability, (5) governability, and (6) bias mitigation [14].

The integration of HCAI and VA for information foraging and sensemaking with feedback loops represents a transformative approach towards understanding complex data and interconnected systems, and serves as the interface for human-AI dialogue. It will enhance human situation understanding and decision making by combining the computational power of AI with intuitive visualizations and human cognitive insights. HCAI places human intent, value and oversight at the core of AI system design. It emphasizes transparency, interpretability, ethical alignment, and agency. This ensures that the systems are efficient as well as accountable and inclusive, where VA can empower HCAI along different dimensions of trustworthy AI [15], supporting the six Principles of Responsible Use of AI in Defence.

Key to this integration is having VA, information foraging, sensemaking processes, and feedback mechanisms that keep humans in control while supported by the performance, interpretability (XAI), and reliability of AI systems.

XAI supported by Visual Analytics can provide an effective and intuitive means for data explainability via data exploration and even post-hoc explainability to review the results and enrich them through visualizations for better understanding of why and how results are generated by AI algorithms.

With the rapid progress of AI-based solutions, systems utilising AI are becoming more potent andprogressive but also opaquer and more complex as well as harder for humans to comprehend. XAI is "concerned with developing approaches to explain and make artificial systems understandable to human stakeholders" [4].

In order to be able to confidently use and employ an AI system (or any other highly advanced system) in a critical setting such as the military domain, one needs to be able to trust the system. The system should also deserve trust, i.e., it should be trustworthy [5]. Within the context of XAI, pursuing trustworthiness involves more dimensions than just being able to comprehend why and how an AI system makes its decisions. For instance, according to Ali et al. [15], there are several interrelated concepts (explainability, interpretability, transparency, interactivity, fairness, robustness, satisfaction, stability and responsibility) contributing to trustworthiness.

Within XAI, the focus of explainability may vary with many possible dimensions such as data explainability, model explainability and post-hoc explainability [15]. For HCAI, HMT and other concepts discussed above, several opportunities to apply XAI in an AI system can be identified. XAI can be applied among other things to data, models and results, and the requirements for XAI will vary according to user roles, tasks and expectations.

RESEARCH CHALLENGES

In this section we discuss the research challenges that need to be addressed for achieving Human-Centered AI with Visual Analytics in a way that benefit NATO operations.

Problem Characterisation for Visual Analytics Enablers Identification and Pairing

It is essential to characterize roles, processes and information requirements characterization for human-machine teams in a data analysis and sensemaking context involving frequent, dynamic and interdependent interactions with feedback. Hybrid teams composed of humans and machines add an additional layer of complexity where mental models about the situation need to be shared between humans and machines, with each having different requirements in meeting the same objective. Beyond the roles defined in the Pirolli and Card sensemaking model [16], we need to consider the humans involved in the development and validation of AI tools.

Adopting a complex systems perspective to VA has been explored across multiple application domains (such as Cyber [17][18], COVID [19][20], and Maritime [21][22]), where identifying visualizations at micro, meso, macro and meta levels can highlight different facets of the systems involved. Similarly, this approach could inform a human-centered perspective that integrates VA as the primary communication channel between humans and AI systems.

In parallel to characterizing roles and tasks, it is necessary to adopt a VA approach that identifies appropriate VA enablers to facilitate data analysis and to overcome the challenges of information / mental model sharing among human and AI team members, coordination, and explainability. The challenge is to assess the extent to which current AI capabilities can contribute to the roles involved in data analysis and sensemaking. Finally, there is a need to develop guidelines to inform which enablers are relevant and appropriate for the categories of tasks and roles in question.

Effective Integration of AI and VA

With regards to integrating VA and AI, there is a growing trend in recent work to leverage AI capabilities within VA applications (referred to as Deep Visual Analytics) [23], as well as to leverage generative AI capabilities to offer insights and produce views on complex data, or to improve interactive elements of visual interfaces [24]. AI capabilities can also be integrated into VA solutions to augment them, and Generative Artificial Intelligence (GenAI) can be leveraged to create elements of VA solutions. It is essential to have VA representations for communicating AI-generated insights from data, as well as explain intent, process and results.

However, existing AI / Machine Learning (ML) methods were designed independently from visualization and similarly existing visualization methods were designed independently from AI / ML methods. There is a need to explore native methods and tools integrating AI and Visualization [25] to create capabilities that move beyond that. These approaches are far from being mature and require further research to assess and realize their potential for defense applications.

The main challenges of integrating VA and AI systems (see Figure 3) include:

- developing efficient visual representations capable of making AI models and results transparent and interpretable, for example, developing appropriate and meaningful representations of the decision-making process, thus increasing their understandability;
- designing VA with adaptable interaction elements that enable the user to give feedback on AI results in a virtuous feedback loop with the AI system, thus developing a continuous dialogue and realising an effective interface between humans and AI;
- developing visual elements that clearly communicate uncertainty in AI processes and reliability of information and models, increasing human confidence in the system results and building trust; and
- facilitating multimodal integration.

As mentioned, addressing human concerns principles [5] promote end user's acceptance, and dedicated visualization design guidelines for HCAI are being developed to include recommendation on improving balance / fairness by reducing biases, advancing transparency of both AI models and information, advancing explainability to increase understandability of results and models. Specific visual elements could be designed to transparently communicate these and other aspects, such as provenance of information to the end users who can be empowered with enhanced support for decision-making. Also, dedicated elements could be developed to clarify aspects such as accountability and facilitate the identification of the decision-making authority.

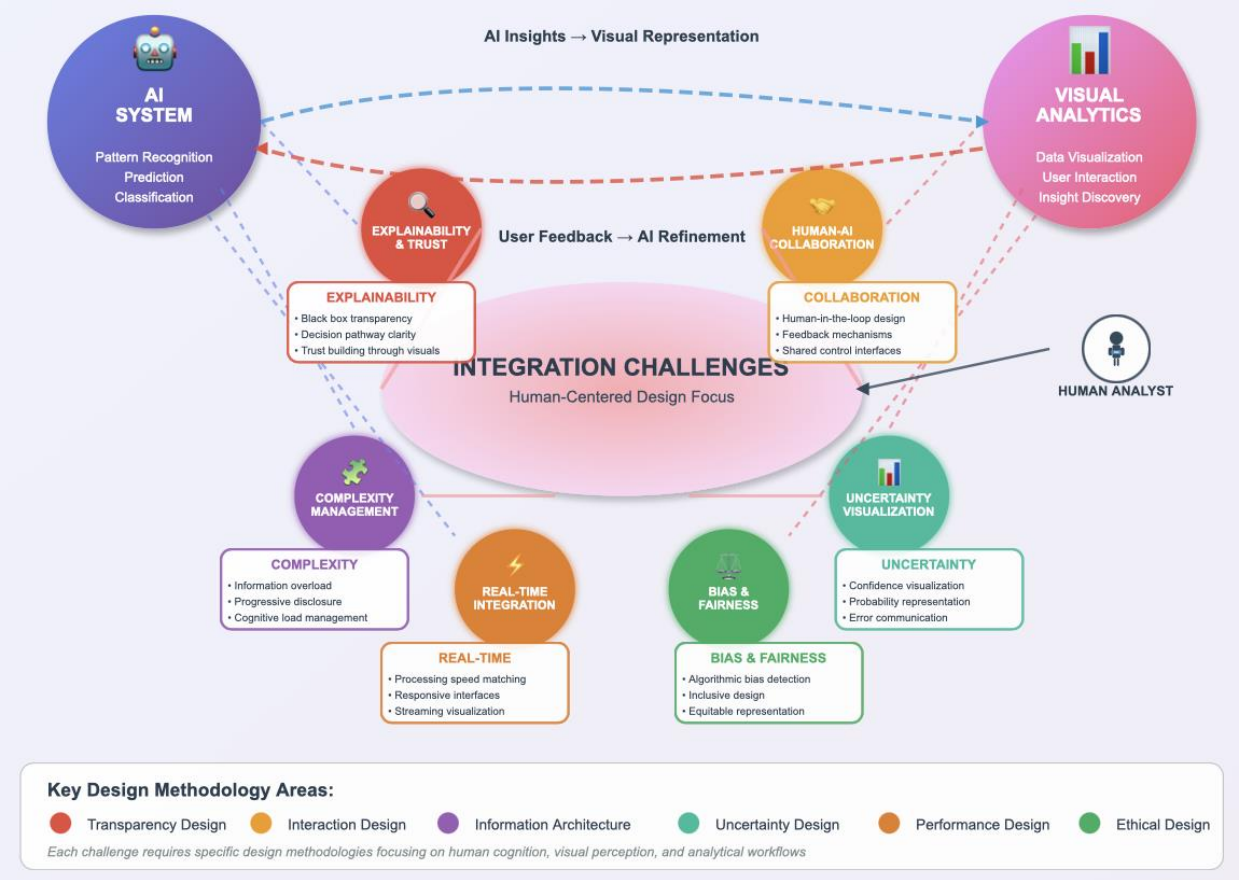

Figure 3. VA and AI integration challenges.

NATO-Relevant Assessment

Methodologies that embody the Visual Analytics approach and enablers need to be assessed on NATO relevant military data and scenarios. For this discussion, we consider operational applications for Cyber Situation Awareness and Understanding, and for Intelligence Analysis. Both represent military applications that involve complex systems where the dynamic interactions of diverse data and interconnected systems are too large and complex to analyze and understand without machine support.

Cyber Situation Awareness and Understanding

An often-encountered concern in the Cyber domain is the vagueness of results generated by AI tools. More often than not, AI powered detection systems will generate alerts and include only the detected indicator, excluding any relevant context. This is detrimental for building strong Cyber Situation Awareness as it hinders the comprehension of the state of the environment and decision making. As the Cyber domain is an abstract environment, humans need to rely on sensors inside the network to perceive possible threats and abnormal behaviour. The integration of AI tools bestows certain level of confidence on their ability to correctly detect suspicious behaviour and flagging it accordingly. These two factors, the role of AI systems in the perception and comprehension, need to be explored deeper as the blind integration without input from actual users is detrimental to the human-machine interaction. Better understanding in how AI detection tools detect and analyse network events leads to higher confidence and enhances the Cyber Situation Awareness and informed decision making (see Figure 4).

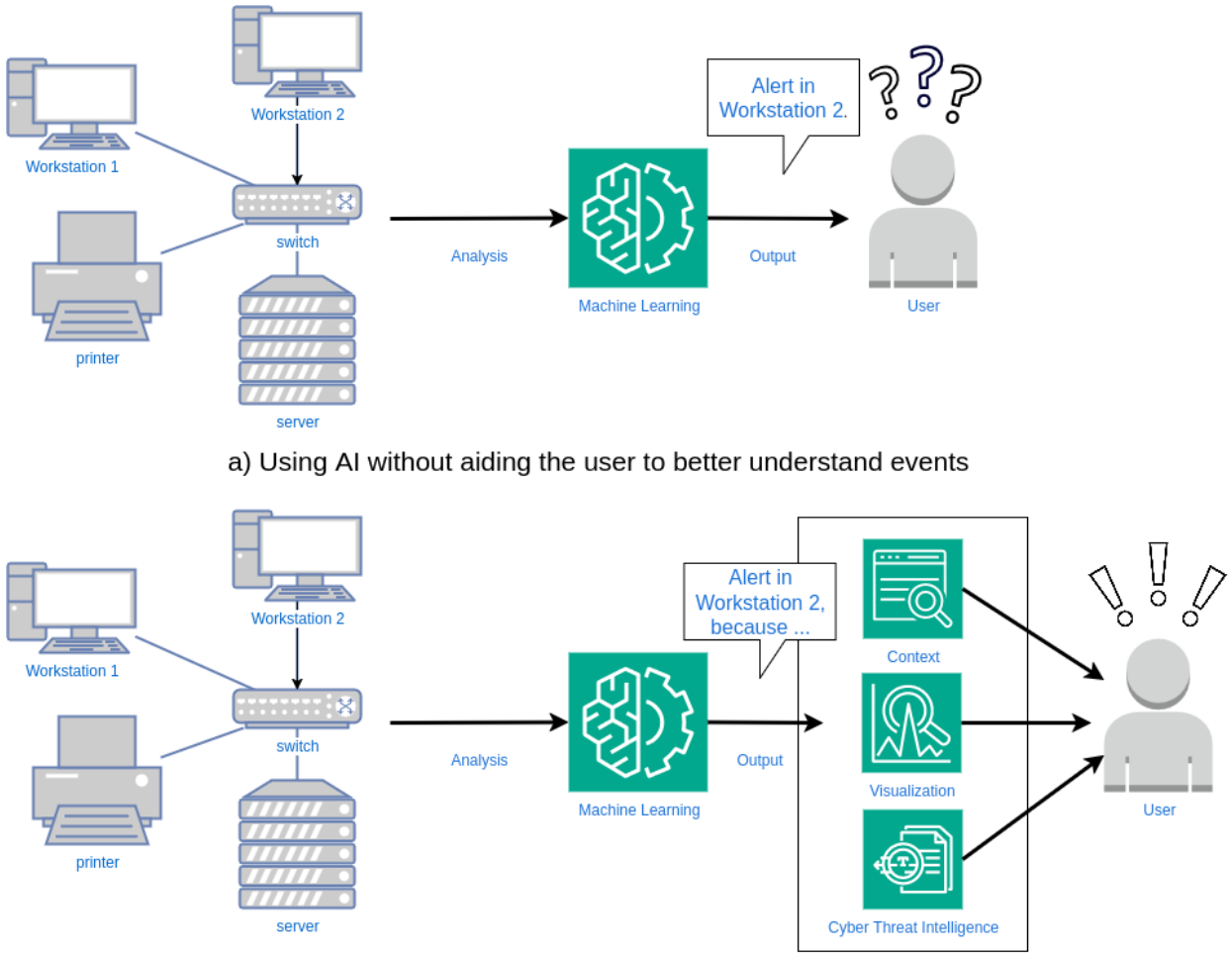

Figure 4. Examples of incorrect and correct integration of AI tools in the Cyber Domain, a) integration as done currently, b) enhancing Cyber Situation awareness through the integration of AI tools.

To achieve this desired outcome, it is necessary to understand, define and evaluate the role of AI tools, how they handle information and transition from the black box paradigm to a grey box implementation through the implementation of XAI techniques. It is also important to consider the intended use and position of such AI tools in a system as well as the intended users. The Human-AI interaction is governed by the human aspect as it is highly dependant on the target audience and the message that needs to be conveyed. A forensic analyst and a system administrator will require different information and it is important for the AI tools to adapt and tailor their results, abstracting information where needed, adapting to the different levels of knowledge and competence. Adaptation and abstraction of information is very important in the Cyber Domain as seen in the context of Advanced Persistent Threats (APTs) and the difficulty of detecting and understanding their impact [18]. When combining AI tools with Visual Analytics in the Cyber domain, it is vital to take into consideration the ultimate goal of the visualization and answer four important questions: (1) what information needs to be shown, (2) what the user is specifically looking for, (3) what the user wants to achieve, (4) what conclusions can be made using the visualization.

Another aspect of the industry wide push for the integration of AI tools, is the security concerns linked to the wide adoption of Large Language Models (LLMs) and other ML technologies. Often, adoption of these tools is rushed, without much concern for establishing correct security measures and adversarial attack mitigation strategies. The OWASP Gen AI security Project publishes annually multiple documents detailing risks and mitigations linked to the use of emergent AI technologies, such as last years "2025 Top 10 Risks & Mitigations for LLMs and GEN AI apps" [26], detailing different attack vectors for currently used LLM and AI tools. As such technologies are often public facing, releasing them without adequate security, can lead to disastrous outcomes. For example, through the use of Prompt Engineering and Prompt Injection [27], adversaries can get hold of sensitive personal information or better understanding of the models used for training and deployment. That does not mean that LLMs should not be used, as through their natural language understanding, they can be used for large data summarization and even the creation of appropriate visualizations of complex information. Only recently, the correct way to create cyber-oriented LLMs has been discussed [28], but a major issue is the potential to target such tools with sophisticated attacks by, for example, using backdoor attacks and manipulation of the models [29] or data poisoning attacks [30]. The aforementioned techniques can also target more sophisticated ML algorithms, such as Neural Networks, to manipulate their features and produce erroneous results or obfuscated malicious behaviour. To better implement and protect such tools, a higher degree of understanding is needed, which will lead to higher trust in the Human-AI interaction.

Intelligence Analysis

In support of NATO operations, the responsibility for anticipating potential threats typically lies with the intelligence community. This task is increasingly complex, as analysts must interpret dynamic situations shaped by numerous rapidly evolving factors and elements. Intelligence analysis involves the examination of diverse sources, ranging from vast volumes of raw sensor data to nuanced human-generated reports, each contributing to a multifaceted understanding of unfolding events. Despite advances in automation, intelligence analysis remains fundamentally reliant on human judgment. Analysts must apply contextual awareness and domain expertise that current AI systems cannot replicate. Strategic decision-making, in particular, demands human oversight, as it involves interpreting ambiguous or incomplete information and assessing potential consequences within broader geopolitical and operational frameworks.

The generation of Indications and Warnings (I&W) through data analysis presents several challenges. Sensor data streams are not only voluminous but also growing in scale and complexity. Relying solely on human analysts to process raw data is both inefficient and prone to errors. Integrating AI technologies can significantly alleviate this burden by automating the detection of anomalies, patterns, and correlations across large datasets. However, the true value of AI lies not only in its processing capabilities but in its synergy with Visual Analytics (VA) tools.

By integrating AI with VA, analysts gain powerful means to explore, interpret, and communicate insights. AI can detect relevant patterns and anomalies, while VA enables users to interact with these findings visually by drilling down into details, tracing data provenance, and contextualizing results with broader operational narratives. This integration supports both sensemaking and decision support, allowing analysts to move fluidly between high-level summaries and granular data views. Importantly, while AI systems are adept at identifying statistical regularities from raw data, they lack the capacity to explain why these patterns matter. Human interpretation remains essential for understanding causality, assessing intent, and evaluating implications, as well as the contextual significance of the insights revealed. Once analysts have formed a coherent understanding of a situation, they must translate their insights into actionable intelligence for stakeholders. Here again, VA plays a critical role by facilitating the clear and compelling communication of complex information.

A further concern in the use of AI for intelligence analysis is the overly agreeable nature of most current AI chatbots [31] where language models tend to exhibit sycophantic behaviour, agreeing with user inputs

regardless of their validity. This poses risks in intelligence contexts, where information is often by its very nature incomplete, uncertain, contradictory and subject to multiple interpretations. Effective analysis requires the ability to consider competing hypotheses and evaluate evidence critically. Systems that reinforce user assumptions without challenge may inadvertently result in biased assessments or obscure alternative explanations.

Providing a visual explanation of the AI system's recommendations, analytical process, and the underlying goals it is pursuing, can help users critically assess the validity of the outputs. By making the reasoning behind AI-generated insights transparent, such as highlighting which data sources were prioritized, what patterns were detected, and how conclusions were drawn, analysts are better equipped to verify that key aspects have been considered. This not only fosters trust in the system but also ensures that results are grounded in reliable evidence rather than biased assumptions. VA, in this context, serves as a bridge between automated analysis and human judgment, enabling users to interrogate the AI's logic and refine interpretations based on operational realities.

The integration of AI and VA offers transformative potential for intelligence analysis but must be approached with careful consideration of human-AI collaboration dynamics. AI can enhance efficiency and scalability, while VA ensures transparency, interpretability, and effective communication. Ultimately, human expertise remains central to making sense of complex data and guiding strategic decisions in support of NATO operations.

CONCLUSION

To build effective, ethical, and secure AI systems, we must move from automation to collaboration. Integrating Visual Analytics, information foraging, sensemaking, and feedback into HCAI and HMT pipelines will create systems that are not only "intelligent", but are aligned with human intent, are ethical, safe and trustworthy. We believe this framework is one of the possible directions for future research, operational tools, and policy development. The development of Visual Analytics enablers for HCAI would be beneficial across multiple military application domains and could be exploited to inform NATO AI adoption.

ACKNOWLEDGMENTS

The authors would like to thank Elena Camossi for her insightful and invaluable contributions to this work.

REFERENCES

[1] Summary of NATO's revised Artificial Intelligence (AI) strategy. (2024, July) https://www.nato.int/cps/en/natohq/official_texts_227237.htm

[2] Wells, B. (2022, September). Artificial Intelligence (AI) Review. NATO Chief Scientist's Report. NATO UNCLASSIFIED.

[3] Wenskovitch, J., Fallon, C., Miller, K., & Dasgupta, A. (2021, October). Beyond visual analytics: Human-machine teaming for ai-driven data sensemaking. In 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX) (pp. 40-44). IEEE.

[4] Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 1–24.

[5] Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

[6] Shneiderman, B. (2022). Human-Centered AI. Oxford University Press.

[7] Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding Human-Centred AI: a review of its defining elements and a research agenda. *Behaviour & Information Technology*, 1-40.

[8] Hoque, M. N., Shin, S., & Elmqvist, N. (2024). Harder, Better, Faster, Stronger: Interactive Visualization for Human-Centered AI Tools. Preprint at arXiv:2404.02147.

[9] Beauxis-Aussalet, E., Behrisch, M., Borgo, R., Chau, D. H., Collins, C., Ebert, D., *et al*. (2021). The role of interactive visualization in fostering trust in AI. *IEEE Computer Graphics and Applications*, *41*(6), 7-12.

[10] Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., .& Li, K. (2021). Artificial intelligence security: Threats and countermeasures. ACM Computing Surveys (CSUR), 55(1), 1-36.

[11] Winkelholz, C., Träber-Burdin, S., Flemsich, F. (2015) 1st International Conference "Human Systems Integration Approach To Cyber Security", Sofia, Bulgaria.

[12] Varga M., Winkelholz C., Träber-Burdin, S. (2018) An Exploration of User Centered and System Based Approaches to Cyber Situation Awareness, 15th IEEE, Symposium on Visualization for Cyber Security (VizSec), 22nd October 2018, Berlin, Germany.

[13] Vicente, K. J., & Rasmussen, J. (2002). Ecological interface design: Theoretical foundations. IEEE Transactions on systems, man, and cybernetics, 22(4), 589-606.

[14] NATO Responsible AI Assessment and Toolkit, November 2023, NATO UNCLASSIFIED.

[15] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. Information Fusion, 99, 1–52.

[16] Pirolli, P., & Card, S. (2005, May). Sensemaking processes of intelligence analysts and possible leverage points as identified through cognitive task analysis. In Proceedings of the 2005 International Conference on Intelligence Analysis, McLean, Virginia (Vol. 6).

[17] Nikolov, G., Varga, M., Panganiban, A. R., Kullman, K., & Lavigne, V. (2025, August). Enhancing Cyber Situation Awareness: Visualizing Advanced Persistent Threats as Complex Systems. In International Conference on Availability, Reliability and Security (pp. 90-107). Cham: Springer Nature Switzerland.

[18] Nikolov, G., Varga, M., Panganiban, A.R., Kullman, K., Lavigne, V. (2025). Enhancing Cyber Situation Awareness: Visualizing Advanced Persistent Threats as Complex Systems. In: Coppens, B., Volckaert, B., Naessens, V., De Sutter, B. (eds) Availability, Reliability and Security. ARES 2025. Lecture Notes in Computer Science, vol 15995. Springer, Cham. https://doi.org/10.1007/978-3-032-00633-2_6.

[19] Varga, M., Ndoni, A., Träber-Burdin, S., Panganiban, A. R., & Lavigne, V. (2024, July). Interactive Visual Analysis of COVID-19. In *2024 28th International Conference Information Visualisation (IV)* (pp. 7-12). IEEE.

[20] Varga, M., Panganiban, A. R., Ndoni, A., Lavigne, V., Träber-Burdin, S., & Lem, M. (2025). Interactive data driven exploration of COVID-19. *Information Visualization Journal*, August 2025 (TBP).

[21] Camossi, E., Lavigne, V., McKenzie, M., Ray, C., Träber-Burdin, S., & Varga, M. (2025, June). Visual Analytics for Maritime Complex Systems. In OCEANS 2025 Brest (pp. 1-7). IEEE.

[22] Camossi, E., Lavigne, V., McKenzie, M., Ray, C., Träber-Burdin, S., & Varga, M. (2025, August). Advancing Maritime Situational Awareness with Visual Analytics and Complex Systems. In *2025 29th International Conference Information Visualisation (IV)* (pp.32-37). IEEE.

[23] Islam, R., Akter, S., Ratan, R., Kamal, A. R. M., & Xu, G. (2021). Deep visual analytics (DVA): applications, challenges and future directions. Human-Centric Intelligent Systems, 1(1), 3-17.

[24] Ye, Y., Hao, J., Hou, Y., Wang, Z., Xiao, S., Luo, Y., & Zeng, W. (2024). Generative ai for visualization: State of the art and future directions. Visual Informatics.

[25] Kovalerchuk, B., Andonie, R., Datia, N., Nazemi, K., & Banissi, E. (2022). Visual knowledge discovery with artificial intelligence: Challenges and future directions. In Integrating artificial intelligence and visualization for visual knowledge discovery (pp. 1-27). Cham: Springer International Publishing.

[26] OWASP Gen AI security Project, (2025). 2025 Top 10 Risks & Mitigations for LLMs and Gen AI Apps, https://genai.owasp.org/llm-top-10/

[27] Pedro, R., Castro, D., Carreira, P., & Santos, N. (2023). From prompt injections to SQL injection attacks: How protected is your llm-integrated web application?. Preprint at arXiv:2308.01990

[28] Zhang, J., Bu, H., Wen, H., Liu, Y., Fei, H., Xi, R., et al. (2025). When LLMS meet cybersecurity: A systematic literature review. Cybersecurity, 8(1), 55.

[29] Shi J, Liu Y, Zhou P et al (2023) BadGPT: Exploring security vulnerabilities of ChatGPT via backdoor attacks to InstructGPT. Preprint at arXiv:2304.12298

[30] Alber, D. A., Yang, Z., Alyakin, A., Yang, E., Rai, S., Valliani, A. A., *et al.* (2025). Medical large language models are vulnerable to data-poisoning attacks. *Nature Medicine*, 31(2), 618-626.

[31] Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., *et al.* (2024). Towards understanding sycophancy in language models. In *12th International Conference on Learning Representations, ICLR 2024*.

ACRONYMS

AI: Artificial Intelligence

APTs: Advanced Persistent Threats

EID: Ecological Interface Design

GenAI: Generative Artificial Intelligence

HCAI: Human-Centered Artificial Intelligence

HMT: Human-Machine Teaming

I&W: Indication and Warnings

LLMs: Large Language Models

ML: Machine Learning

MLDevOps: Machine Learning Development and Operations

NATO: North Atlantic Treaty Organization

SIEMs: Security Incident Event Management Systems

SOC: Security Operation Center

UI: User Interface

VA: Visual Analytics

XAI: eXplainable AI