

FCAS Ethical AI Demonstrator

Florian Osswald^{1,*†}, Roman Bartolosch^{1,†}, Torsten Fiolka^{1,†}, Engelbert Hartmann^{2,†},
Bernhard Krach^{2,†}, Jan Feil^{2,†} and Martin Lederer^{3,†}

¹Fraunhofer FKIE, Wachtberg, Germany

²Airbus Defence and Space GmbH, Manching, Germany

³Data Machine Intelligence Solutions GmbH, Munich, Germany

Abstract

While artificial intelligence (AI) has become part of more and more areas of daily life – both private and business – this development has not yet progressed as far in the military sector. This is just changing with the development of new projects, such as the Future Combat Air System (FCAS) – a highly ambitious European defense project planned as a replacement of systems such as the Eurofighter from 2040 onwards. To facilitate and accelerate discussions on the ethical implications of the use of AI in the military domain, we developed the FCAS Ethical AI Demonstrator. We chose the Target Detection, Recognition, and Identification as one highly probable use case and implemented a simulation to showcase the ethical implications of the collaboration between the operator and an AI-assisted system in that application. To help the operator understand and assess the classifications of the used automatic target recognition, explanations of the AI results are computed with an Explainable AI (XAI) method and then provided in the user interface. With this hands-on demonstrator, we are pleased to contribute to the discussions on the ethical implications of the use of AI in military applications.

Keywords

Ethical AI, Explainable AI, Future Combat Air System (FCAS), Targeting Cycle

1. Introduction

Artificial intelligence (AI) has become a technology that influences social and economic life in many ways – ChatGPT shows this vividly [1]. The military domain is not excluded from this trend [2, 3, 4, 5, 6, 7]. AI has the potential to help operators make decisions in situations with ever decreasing time and ever more information available [8]. The amount of information available to the operator is also extensive in the Future Combat Air System (FCAS), so the support of an AI seems necessary. Set out to be the most ambitious European defense project for the upcoming years, FCAS is a joined effort of European nations. From 2040 onward FCAS is planned to integrate gradually with the current systems such as the Eurofighter, which it finally will replace.

Late-breaking work, Demos and Doctoral Consortium, colocated with The 1st World Conference on eXplainable Artificial Intelligence: July 26–28, 2023, Lisbon, Portugal


*Corresponding author.

†These authors contributed equally.

✉ florian.osswald@fkie.fraunhofer.de (F. Osswald); jan.roman.bartolosch@fkie.fraunhofer.de (R. Bartolosch);
torsten.fiolka@fkie.fraunhofer.de (T. Fiolka); engelbert.hartmann@airbus.com (E. Hartmann);
bernhard.krach@airbus.com (B. Krach); jan.feil@airbus.com (J. Feil); ml@datamachineintelligence.eu (M. Lederer)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

AI will be utilized in FCAS to help operators focus on the information relevant to the situation at hand. However, the use of AI in such a sensitive domain poses serious ethical and legal questions. To investigate the responsible use of new technologies in FCAS and to determine necessary guidelines for such a system, the *FCAS Forum* [9] was founded – an interdisciplinary commission with experts from fields such as political science, history, and theology, next to technical specialists that have already published on this matter [4, 3, 6, 10]. As a basis for discussions about the responsible use of AI in the military domain we developed the FCAS Ethical AI Demonstrator. It showcases an exemplary scenario of collaboration between the operator and an AI.

Using Explainable AI (XAI) methods is imperative here, since it is crucial for the operator to understand and be able to assess the AI's assistance. Extracting the information on which neural networks base their decision, thus making the decision process explainable resp. interpretable, is the focus of the research field XAI. It was heavily influenced by the XAI program the Defense Advanced Research Projects Agency (DARPA) launched in 2017 [11, 12]. In the subsequent years, various approaches were proposed as surveyed in [13, 14, 15, 16, 17, 18, 19, 20].

In the following, we first elaborate on the motivation of developing the FCAS Ethical AI Demonstrator. This requires a rough introduction into the operational context of a military situation. After that, we explain the actual scope and content of the demonstrator. We then highlight the technical details including the relevance of XAI in it. Finally, the necessity of XAI for developing ethical AI is discussed.

2. Context and Motivation

The use of AI is expected to have an immense impact on the conduct of military operations [21]. Performing operational activities in a dynamic environment requires a quick adaptation of decisions. This is formalized in the OODA-Cycle [22], short for Observe, Orient, Decide, and Act. It describes four stages of decision-making in fast changing environments. A key question in relation to the use of AI in these stages of decision-making is the degree of human involvement. In this context, a machine's level of authority can be described by its dependency on human actors in the execution of the OODA-Cycle activities, especially in light of operational uncertainty [23]: *Human-in-the-loop* (human makes decisions and acts), *Human-on-the-loop* (systems make decisions and act, human is monitoring and can intervene), or *Human-out-of-the loop* (systems make decisions and act, no human intervention possible)[24].

The role of the human in the application of AI in military operations and the allocation of responsibility for machine executed authorities appear central for the development of regulatory frameworks and the discussion of ethical implications [2, 4, 25]. A starting point for a detailed discussion is a concrete use case. For FCAS in [24], an initial but still incomplete set of AI use cases was identified and preliminarily assessed. Among others this includes Mission Planning and Execution (MPE), Target Detection, Recognition and Identification (DRI), and Cyber Security and Resilience (CSR). The FCAS Ethical AI Demonstrator focuses on DRI covering the use of



Figure 1: Screenshot of the Ethical AI Demonstrator. In the video, four objects are detected by the ATR as air defense systems. With three buttons below the video stream, the user can confirm or reject a detection or mark it for further investigation. In the list on the left, the detected vehicle type is selected. On the right, the heatmap generated by LIME is shown to explain the ATR detection.

AI technology to detect and identify potential targets with an Automatic Target Recognition (ATR). The following section provides a detailed overview of its scope.

3. Scope and Content

The FCAS Ethical AI Demonstrator showcases the collaboration between an operator and an AI performing DRI. One exemplarily implemented scenario shows an unmanned aerial system (UAS) flying ahead of the main forces to detect and identify hostile air defense systems (see Figure 1). On the ground, military vehicles are intermixed with civilian infrastructure. The user of the demonstrator, i.e. the operator of the UAS, has to decide upon the AI's detections. They must verify or reject a detected target or mark it for further investigation.

To keep the operator with *meaningful control in the loop* the target detection is presented together with an explanation. For this, the well-established XAI method LIME (Local Interpretable Model-Agnostic Explanations, [26]) is used to generate heatmaps visualizing the features of the targets which were essential for the detection. The user of the demonstrator is thus set in a possible real-life situation where they can experience the impact an AI-based assistant could have on the process of target selection. This especially facilitates discussions about ethical issues which might arise.

4. Technical Details

The FCAS Ethical AI Demonstrator is implemented as a web-based application system. Correspondingly, the synthetic video, ATR data and LIME explanation are combined and controlled in a web-based user interface. In the following, we first briefly describe the synthetic video. Then, we describe the used ATR and how its detections are explained with LIME. Finally, the overall application architecture and user interface (UI) are described. As a data basis for the demonstrator, we use a 4K video generated within the high fidelity Airbus pilot training environment. It is based on a scriptable aircraft model carrying a controllable video pod and observing a manually defined automated ground scenario. The necessary metadata for geolocalization is embedded into the recording.

The ATR AI model used on the synthetic video is based on the Airbus proprietary *CeMoreDeep* architecture. It combines a feature extractor of a convolutional neural network with a highly optimized support vector machine to detect and classify objects. Furthermore, the Airbus proprietary software *AI Engine RT* stabilizes the generated tracks and transfers the predicted position into latitude and longitude for geolocalization and visualization on a map. As this ATR model is provided as a black box, a model agnostic XAI method is necessary to explain its classifications. While several approaches like SHAP [27] and Ablation-CAM [28] are suitable, we have chosen LIME [26], a well-established surrogate based XAI model, for this purpose. LIME works by sampling input images and creating locally interpretable models around each given image. By doing so, it identifies the critical areas of the image that the ATR model uses to make its predictions. This allows users to better understand the decision-making process of the ATR model and evaluate the quality of its classifications.

The demonstrator is implemented as a distributed architecture using web-based technologies. After logging in into the web UI, the user can play back the video; ATR and LIME data are loaded from a web server and displayed accordingly. The UI guides the user through the experience and additionally collects his or her reactions.

5. Explainable and Ethical AI

Jobin et al. [29] analysed the guidelines and principles that were issued to constitute ethical AI. Five dominant principles were identified: transparency, justice and fairness, non-maleficence, responsibility, and privacy [29, 30]. XAI results, here the visualization of relevant image parts, help to make the decision-making process of the AI more transparent. While ethics provides principles that should be met by an AI in order for it to be considered ethical, XAI can be an enabler to meet the requirements and uncover other ethical issues that may arise in a particular context. With the FCAS Ethical AI Demonstrator, we contribute to the discussions on the ethical use of AI in a specific context, the military domain.

References

- [1] OpenAI, GPT-4 Technical Report, Technical Report, OpenAI, 2023.
- [2] I. Verdiesen, F. S. de Sio, V. Dignum, Accountability and control over autonomous weapon systems: A framework for comprehensive human oversight, *Minds and Machines* 31 (2021) 137–163. URL: <https://doi.org/10.1007/s11023-020-09532-9>. doi:10.1007/s11023-020-09532-9.
- [3] W. Koch, On digital ethics for artificial intelligence and information fusion in the defense domain, *IEEE Aerospace and Electronic Systems Magazine* 36 (2021) 94–111.
- [4] W. Koch, On ethically aligned information fusion for defence and security systems, in: 2020 IEEE 23rd International Conference on Information Fusion (FUSION), IEEE, 2020, pp. 1–8.
- [5] W. Koch, What artificial intelligence offers to the air C2 domain? NATO allied command transformation (ACT), 2022. URL: https://issuu.com/spp_plp/docs/what_artificial_intelligence_offers_to_the_air_c2_fr=sNzFiMzQ4MjEzNTc.
- [6] U. Franke, Harnessing artificial intelligence (2019). URL: <https://www.fcas-forum.eu/publications/Harnessing-artificial-intelligence.pdf>.
- [7] H. Meerveld, R. Lindelauf, E. Postma, M. Postma, The irresponsibility of not using AI in the military, *Ethics and Information Technology* 25 (2023) 14.
- [8] W. Koch, Perspectives on AI-driven systems for multiple sensor data fusion, *tm - Technisches Messen* 90 (2023) 166–176. URL: <https://doi.org/10.1515/teme-2022-0094>. doi:10.1515/teme-2022-0094.
- [9] W. Koch, FCAS forum. mission, 2023-04-17. URL: <https://www.fcas-forum.eu/en/mission/>.
- [10] E. Rosert, F. Sauer, How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies, *Contemporary Security Policy* 42 (2021) 4–29.
- [11] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), *IEEE Access* 6 (2018) 52138–52160. doi:10.1109/ACCESS.2018.2870052.
- [12] D. Gunning, E. Vorm, J. Y. Wang, M. Turek, DARPA’s explainable AI (XAI) program: A retrospective, *Applied AI Letters* 2 (2021) e61. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ail2.61>. doi:<https://doi.org/10.1002/ail2.61>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/ail2.61>.
- [13] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Information Fusion* 58 (2020) 82–115. URL: <https://www.sciencedirect.com/science/article/pii/S1566253519308103>. doi:<https://doi.org/10.1016/j.inffus.2019.12.012>.
- [14] R. Guidotti, A. Monreale, F. Turini, D. Pedreschi, F. Giannotti, A survey of methods for explaining black box models, *CoRR* abs/1802.01933 (2018). URL: <http://arxiv.org/abs/1802.01933>. arXiv:1802.01933.
- [15] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, Machine learning interpretability: A survey on methods and metrics, *Electronics* 8 (2019). URL: <https://www.mdpi.com/2079-9292/8/8/832>. doi:10.3390/electronics8080832.

- [16] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, K.-R. Müller, Explaining deep neural networks and beyond: A review of methods and applications, *Proceedings of the IEEE* 109 (2021) 247–278.
- [17] D. Minh, H. X. Wang, Y. F. Li, T. N. Nguyen, Explainable artificial intelligence: A comprehensive review, *Artif. Intell. Rev.* 55 (2022) 3503–3568. URL: <https://doi.org/10.1007/s10462-021-10088-y>. doi:10.1007/s10462-021-10088-y.
- [18] C. Molnar, *Interpretable Machine Learning*, 2 ed., 2022. URL: <https://christophm.github.io/interpretable-ml-book>.
- [19] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, W. Samek, *Explainable AI Methods - A Brief Overview*, 2022, pp. 13–38. doi:10.1007/978-3-031-04083-2_2.
- [20] U. Schmid, B. Wrede, What is missing in XAI so far?: An interdisciplinary perspective, *KI - Künstliche Intelligenz* 36 (2022). doi:10.1007/s13218-022-00786-2.
- [21] P. Svenmarck, L. Luotsinen, M. Nilsson, J. Schubert, Possibilities and challenges for artificial intelligence in military applications, in: *Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting*, 2018, pp. 1–16.
- [22] R. Coram, *Boyd: The fighter pilot who changed the art of war*, Hachette+ ORM, 2002.
- [23] M. Firlej, A. Taeihagh, Regulating human control over autonomous systems, *Regulation & Governance* 15 (2021) 1071–1091.
- [24] M. Azzano, S. Boria, S. Brunessaux, B. Carron, A. Cacqueray, S. Gloeden, F. Keisinger, B. Krach, S. Mohr dieck, *The responsible use of artificial intelligence in FCAS—an initial assessment* (2021). White Paper. Available online at <https://www.fcas-forum.eu/articles/responsible-use-of-artificial-intelligence-in-fcas>.
- [25] F. E. Morgan, B. Boudreaux, A. J. Lohn, M. Ashby, C. Curriden, K. Klima, D. Grossman, *Military applications of artificial intelligence: ethical concerns in an uncertain world*, Technical Report, RAND PROJECT AIR FORCE SANTA MONICA CA SANTA MONICA United States, 2020.
- [26] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?" explaining the predictions of any classifier, in: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.
- [27] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Advances in neural information processing systems* 30 (2017).
- [28] H. G. Ramaswamy, et al., Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 983–991.
- [29] A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines, *Nature Machine Intelligence* 1 (2019) 389–399.
- [30] H. Vainio-Pekka, M. O.-o. Agbese, M. Jantunen, V. Vakkuri, T. Mikkonen, R. Rousi, P. Abrahamsson, *The role of explainable AI in the research field of AI ethics*, *ACM Trans. Interact. Intell. Syst.* (2023). URL: <https://doi.org/10.1145/3599974>. doi:10.1145/3599974, just Accepted.