A Review and Strategic Approach for the Transition towards Third-Wave Trustworthy and Explainable AI in Connected, Cooperative and Automated Mobility (CCAM)

Alper Kanak[1,2], Salih Ergün[1,2], *Member, IEEE,* Ali Serdar Atalay[3], Stefano Persi[4], Ahu Ece Hartavi Karcı[5]
[1]*ERGTECH Research Center, Switzerland;* [2]*ERARGE - Ergünler Co., Ltd. R&D Center, Turkiye;* [3]*BITNET, Turkiye;* [4]*Mosaic Factor, Spain;* [5]*University of Surrey, UK*
{alper.kanak, salih.ergun}@ergtech.ch; serdar@bitnet.com.tr; stefano.persi@mosaicfactor.com; a.hartavikarci@surrey.ac.uk

***Abstract*—Traditional "2nd Wave (2WAI)" AI algorithms are highly specialized in narrowly-defined tasks in the transportation and mobility domain, however, quite hardly explainable and unable to cover the needs of unbiased and trusted decision-making in the CCAM domain. There is a strong need to accelerate the shift towards human-like "3rd Wave AI (3WAI)" to overcome the challenges of road transport to make autonomous vehicles and driving safe, cleaner and more efficient. This paper aims to shed light on ambitious pathways toward more trustworthy and explainable AI by presenting a strategic approach envisioning the requirements of 5 main audience profiles in the CCAM context (developers, decision-makers, regulators, end-users, and CCAM Service providers). The presented approach envisages a 6-dimensional concept that brings features of i) Transparent and explainable; ii) Fair and impartial; iii) Responsible and accountable; iv) Robust and Reliable; v) Respectful of Privacy; vi) Safe and Secure, transportation. To achieve the transition towards 3WAI a model-based n-sprint V methodology is introduced that is enriched with continuous validation by putting ethics at the centre of all potential innovations. The paper presents a review of the recently funded European projects within the concept of AI-powered CCAM.**

Keywords—Trustworthy AI (TAI), Explainable AI (XAI), autonomous vehicles (AV), Connected, Cooperative and Automated Mobility (CCAM), autonomous driving (AD)

I. Introduction and Current State-of-Play

Cooperative, connected and automated mobility (CCAM) is one of the next big trends in the automotive industry [1] that positions Artificial Intelligence (AI) in the centre for more trustworthy, safer, greener, secure, privacy-preserving, accountable, efficient and adopted mobility practices in both passenger and goods transportation. This paper aims to revisit Europe's CCAM strategy and highlights the new third-way AI trend (3WAI) toward next-generation transportation and mobility systems.

European Commission (EC) has adopted ambitious road safety targets. In continuation with the previous roadmaps for 2010 and 2020, the EC has set a new 50% reduction target for the number of fatalities and serious injuries on European roads by 2030, as a milestone on the way toward the so-called 'Vision Zero' [2].

That is because one thing did not change since then: People. Still, in 90% of all accidents, the driver is a contributing cause [3]. Therefore, there is a real need to develop new technology that will improve future transport systems' safety while reducing their adverse impacts on the environment. In this context, vehicle automation plays a significant role to fulfil these objectives. Consequently, the EC has started working to deliver all elements of its 2019 Road Safety Policy Framework by promoting safe systems and implementing its Sustainable and Smart Mobility Strategy [4].

One of the main expected milestones of this strategy is that "automated mobility will be deployed at a large scale by 2030"[5]. In this context, due to the rapid advancement of AI, connectivity, and self-driving technology, the automotive industry has faced an onslaught of technological and regulatory changes. Mainly, AI has a huge potential to advance the perception, situational awareness, and decision-making processes of autonomous vehicles (AV). However, as AI becomes more advanced than ever, humans are challenged to fully understand and retrace how the algorithm comes to a decision. These aspects have raised some mistrust towards AI and machine learning (ML) algorithms, causing the already existing distrust of autonomous driving (AD) technology to grow (i.e.~75% of US drivers fear AD technology [6]). That becomes a major concern mainly in the automotive sector since AVs have to go through a more rigorous certification process than conventional vehicles to enhance the adoption and acceptance of 3WAI. Therefore, it is crucial to prioritise 'Trust' in AI/ML-powered technologies that are being implemented in automated/self-driving vehicles [7].

The expected milestones and the emerging needs of CCAM stakeholders have triggered new discussions on leveraging up the recent AI knowledge, say SecondWave AI (2WAI), towards the 3WAI by making AI-based solutions more trustworthy and explainable. For this purpose, this paper first presents the CCAM vision in line with the 3WAI and illustrates a strategic overview of ambitious pathways in Section III. Section IV describes a novel verification and validation methodology, called V-Cycle, that can be adapted to make AI-based CCAM systems more trustworthy. Section V concludes the paper.

II. CCAM Vision, IEEE CAS and 3WAI Concept

The CCAM Collaboration is a public-private partnership that brings together the research and development efforts of all stakeholders to expedite the adoption of breakthrough CCAM technologies and services throughout Europe. It intends to maximize the systemic benefits of CCAM-enabled innovative mobility solutions: greater safety, decreased environmental impact, and inclusivity. By bringing together the complex cross-sectoral value chain players with a similar vision, the CCAM vision will establish and implement a unified, coherent, and long-term R&I strategy: "*European leadership in safe and sustainable road transport through automation*". CCAM vision is composed of seven clusters (i. Large-scale demonstrations, ii. Vehicle Technologies, iii.

Validation, iv. Integrating the vehicle in the transport system, v. Key enabling technologies, vi. Societal aspects and user needs, vii. Coordination) [1].

IEEE Circuits and Systems (CAS) vision is aligned with the CCAM Vision as new paradigms governing the computation connectivity in the fields of human-centric technologies, mobility and smart cities are mentioned in the technical strategic areas of the CAS society in the 2020-2024 strategic plan [8]. AI has a special role in cutting-edge CAS research as AI has been implemented either at the integrated circuit (IC) level, GPU and high-level algorithmic level. CAS and AI meet in the CCAM domain, especially at the vehicle level and connected and autonomous system components. The CAS vision still seeks novel solutions for AI computing, neuromorphic computing, accelerators, deep ML techniques and recently trustworthy and explainable AI for AD, AVs and multimodal transportation and mobility solutions in the CCAM area.

The 3WAI as described by DARPA is a new paradigm making AI contextually adapt to changing situations [9]. Aligned with the CCAM strategy there is an increasing need to leverage 2WAI to 3WAI which aims to make algorithms more like a human assistant than just a tool that is trained on a human-curated dataset to solve a specific problem. Note that the 3WAI is still an open research area, especially in the CCAM domain as there exist scientific and technical challenges as well as societal challenges.

As illustrated in Fig 1., the proposed 3WAI concept is based on **a 6-dimensional approach** that is adopted for Trustworthy and explainable AI: **1. Transparent and explainable**: Help participants understand how their data can be used and how AI systems make decisions. **2. Fair and impartial:** Assess whether AI systems include internal and external checks to help enable equitable application across all participants. **3. Responsible and accountable:** Put an organizational structure and policies in place that can help determine who is responsible for the output of AI system decisions. **4. Robust and Reliable**: Confirm that AI systems can learn from humans and other systems and produce consistent and reliable outputs. **5. Respectful of Privacy:** Respect data privacy and avoid using AI to leverage customer data beyond its intended and stated use. Allow customers to opt-in and -out of sharing their data. **6.Safe and Secure:** Protect AI systems from potential risks (including cyber risks) that may cause physical and digital harm. The solutions to these 6 dimensions should be implemented by considering the following four common facilitators: Situational Awareness, Comprehensive multimodal Data Governance, Agile Verification & Validation Process, and Technical and Perceived Trust.

The proposed 3WAI concept is capable of developing an open-trustworthy conceptual framework, set of innovative methods and human-centric toolsets by enhancing explainability and interpretability for easing the verification and validation (competence check) process of human-like 3WAI algorithms (minimalism and pursuit of excellence) in the CCAM setting. This can be achieved by putting ethics at the centre (character) of all innovations to articulate two main cognitive engineering constructs: trust (competence, character) across the 5 targeted audience profiles (I. Data scientists, II. Decision-makers, III. Regulators, IV. End-users, V. CCAM Service providers) and situational awareness in AVs. The AI tools should take trust into account from two angles:

Trust to a Person=$f\{C_{1P}, C_{2P}\}$f{C_{1P}, C_{2P}}; C_{1P} represents 'Competency' and $\boldsymbol{C_{2P}}$ represents 'Character'; (1)
Trust to an AI-based AD function=$f\{C_{1AI}, C_{2AI}\}$; C_{AI1} and $\boldsymbol{C_{AI2}}$ represent 'Certification' and 'Intent and Integrity' (2)

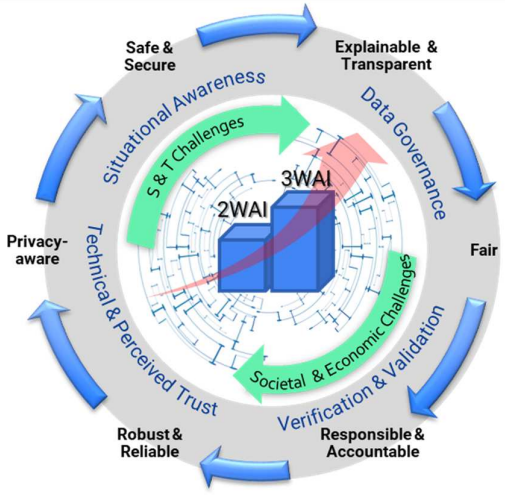

Fig. 1. The proposed 3WAI concept for CCAM

III. Ambitious Pathways

These two angles should be considered as the basis of applications utilising trustworthy and explainable AI. In line with this strategy, the following ambitious pathways are worth getting more focused on improving the current literature.

A. Trust in Automated Vehicles (AV) and Driving (AD)

Perceptions and feelings of future AV users are crucial to increasing the acceptance of the technology. Particularly perceptions of trust, safety, stress and control, are found to be important factors that affect the extent to which users accept AVs. Past literature has identified Trust as a fundamental factor for seamless human-automation interaction. However, there is a tendency for low Trust in AD technology. Furthermore, analyses have shown the link between the GDPR and trustworthy AI systems, including the fairness data protection principle, the regulation of automated decision-making and the assessment and mitigation of any risk of data processing systems to fundamental rights and freedoms of individuals. Although how trustworthy AI is considered by the public is of importance, the knowledge around the public opinion regarding the use and concerns of AI, especially within the CCAM concept, is still restricted and scattered. The recently proposed EU AI Act (COM(2021) 206) [10] introduces more effective legal requirements to prevent or mitigate biases in high-risk AI, through a data management plan, human oversight, risk assessment, transparency measures and regulatory sandboxes. However, these proposals are still highly debated, and their practical implementation is still uncertain given the uncertainty, trust and bias around AI in the CCAM domain. This has raised some **mistrust** towards AI/ML algorithms, causing the already existing distrust of AD technology to grow. Hence, Trust has become central to the future success and implementation of AI-based CCAM services. But informing end-users may not be enough for Trust. Some studies suggest that providing explanations for the actions of AVs can be effective on Trust [11]. Formal methods are also addressed as they increase the trust in the safety of AVs and assist in evidence-based assessment for certification and homologation of the processes [12].

B. Bias Awareness

Bias analysis has been an increasing concern in recent years, after reports that AI models may discriminate against women, minorities and vulnerable groups in different ways [13] but also fail to incorporate environmental and social externalities. A response to this increasing concern has been to develop bias tracking systems such as Fairness 360 [14], or the other 40+ solutions that promise to identify and mitigate bias in AI systems. However, AI bias is not limited to fairness, nor can it be identified by checking for bias at one specific moment in the life cycle of the algorithm.

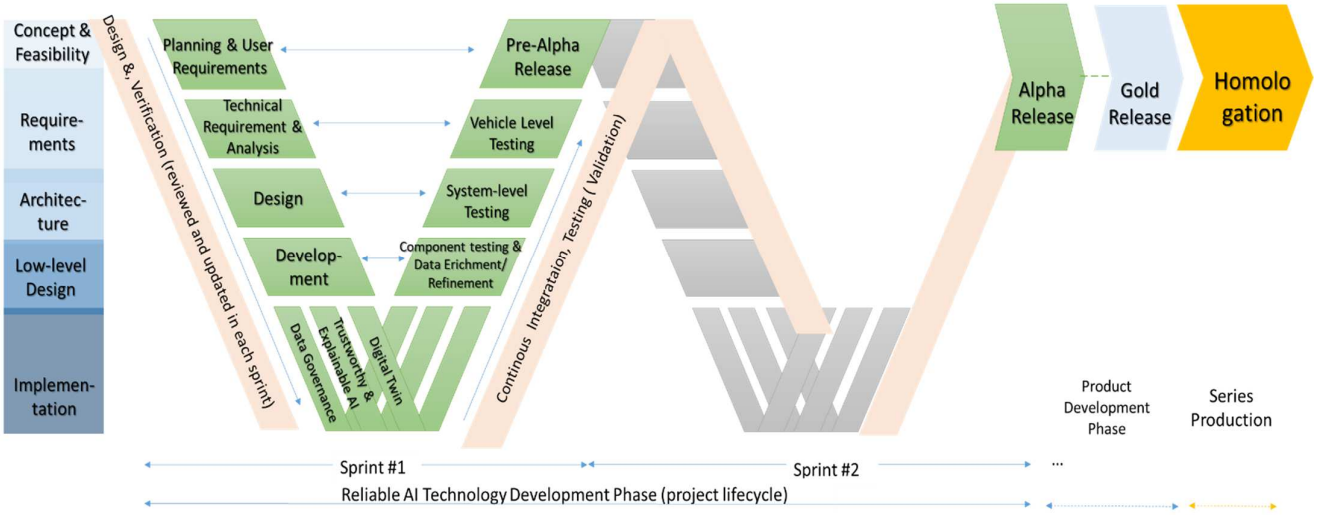

Fig. 2. V methodology based on sprints

C. Explainability & Interpretability

Although ML models can be considered reliable, the effectiveness of these systems is limited by the current inability of machines to explain their decisions and actions to human users. Explainability has been addressed in many capacities but DARPA's approach [9] can be seen as one of the closest approaches to real-life applications as it presents a clear vision to improve the pareto of the explainability vs. learning performance tradeoff for the well-known 2WAI algorithms (Deep learning, RNN, CNN etc.). Since the 2020s, 3WAI has been promoted as the evolved version of 2WAI which is for better contextual adaptation to explain decisions while requiring fewer data and learning with minimal supervision. Different techniques exist for XAI [15]: i) Model-Specific Techniques (white box); ii) Model-Agnostic techniques (black box). Among model-specific algorithms decision trees, DeepLIFT [16] and Supersparse Linear Integer Models (SLIM) [17] are the foremost ones. On the other hand, Model-agnostic techniques present more practical use of 2WAI as black-box models/algorithms, especially within the context of visualization-based approaches for AI explainability. Shapley Additive explanations (SHAP) [18] and (Local interpretable model-agnostic explanations) LIME [19] are the widely adopted black-box techniques where SHAP is slower but more interpretable as compared to LIME. There is a general trend of using black box models encompassing AI for AVs. For instance, in perception, navigation, planning, and control, various visual techniques have been proposed to explain Neural Networks (NN) for scene understanding in AVs [20] (including Class Activation Map –CAM [21], VisualBackProp[22]) for explaining risk analysis and prediction (e.g. saliency maps [23] for accident anticipation, scene graphs for AV risk assessment [24] and building trust in Avs [25], explaining collisions by treeSHAP [26]), for navigation and control via interactive visual interfaces (e.g. natural language generation [27]). GPU implementations of such algorithms as GPUTreeSHAP, are also proposed in research studies [28]. There is also a growing body of research on how AV explanations impact driver-related outcomes, and consequently the adoption of AVs [29].

D. Predictive Situational Awareness (SA)

Situational awareness in CCAM is frequently used to model, analyse, and understand the behaviour of drivers in different contexts [30], but also how AVs use AI techniques to view and interpret their driving environment [31], and emulate how humans perceive and reason. To make predictions, AI-based reasoning techniques are employed to anticipate the development of critical situations and to make predictions. Several knowledge-based inferencing, heuristic algorithms, Bayesian reasoning, fuzzy logic, NN, and high-end processing technologies have already enabled some aspects of AI-based SA, such as basic traffic sign recognition and lane detection, detection and intent estimation of pedestrians and road-users, detection and motion modelling of near-by vehicles, risk assessment and accident avoidance [32].

E. Verification and Validation (V&V) of AI-based Systems

Perception and decision-making deals have the capability of overcoming complex and uncertain problems that need to be assessed through realistic models relying on critical traffic scenarios. The problematic nature of AI/ML in CCAM, along with the difficulty of finding suitable solutions, calls for research on the V&V of AI systems. VERIFAI [33] has been proposed as a software toolkit for the formal design and analysis of systems that include AI/ML. V&V approaches have been addressed for the safety of AVs [34] in recent years but cyber-physical security and privacy-aware V&V towards the homologation of vehicles have not been presented [35]. Moreover, Digital Twin (DT) has become the key technology bridging SA, V&V and TAI/XAI in intelligent transport systems. In recent years blockchain-enabled consensus mechanisms are integrated to enhance the security and efficiency of multi-stakeholder collaborations [36].

IV. V-Cycle Methodology

An effective development methodology is proposed to unlock the potential of AI through the development of explainable and trustworthy human-centric tools, and methods

for the CCAM community, from the viewpoint of 5 targeted audience profiles for different types of transport in 6 dimensions (Section II). This is required for accelerating the shift from the 2WAI towards the 3WAI through model-based n-sprint V-cycle, to overcome the challenges of road transport (safe, clean and efficient) (Fig.2). The proposed methodology follows time-boxed iterations throughout the development cycle to overcome the complexity of targeted technologies and challenges. In this view, the implementation methodology is based on an iteration-based n-V cycle, consisting of n V-type sprints for V&V of the AI outputs continuously. Therefore, each sprint becomes a complete V including development and integration and testing. The approach outputs in 5 interlacing layers (left blue layers): i) Concept and feasibility; ii) Requirements specification; iii) Architecture design; iv) Low-level design; v) Implementation. The left arm of each V model is followed for understanding and (re)elicitation of needs, concept building and refinement and verification of AI solutions and tools. Within the initial definition phase, the requirements on the concept are collected and all boundary conditions targeting legislative, user-related aspects together with functional requirements are considered. Hence, the low-level stakeholders such as development teams, data scientists, technical architectures, psychologists (human-like design), and automotive engineers play a critical role with high-level stakeholders. Once the concept phase is passed successfully, the design phase starts leading to further analysis, detailed design models, and the simulation of functions. The final step includes the development phase, which fulfils the realisation of TAI and XAI solutions. The right arm of n-V deals with the continuous integration, testing and validation of the components at various levels from different perspectives (explainability, trustworthiness, bias etc.) for the selected use-cases. The process continues with system-level and vehicle-level testing where the black-box testing strategy is applied. The first V is then followed by the second and further sprints to reach the alpha, beta and future releases until the Gold version is released and the homologation process starts.

In the proposed methodology the gained experiences and knowhow can be transfered from one V-cycle to another through effective communication among developers and also by using online project management tools. Data and process management is also crucial in iterative methodologies similar to the proposed n-sprint V-cycle approach. Especially for large-scale projects there is a srong need to define the roles, specify the requirements and interlink the developments, achievements, tests, verification and validation procedures with each other. An ontological semantic interlinking is needed to model the relations between users, beneficiaries, developers, and all other stakeholders as well as requirements and functions of systems, subsystems, modules, services and data.

V. Discussion on Positioning of EU Projects

CCAM initiative is designed to support EU countries and the European automotive industry in their transition to connected and automated driving while ensuring the best mobility environment for the public. This initiative has been built on the previous research framework programmes (FP) starting from FP6 and continued in FP7, Horizon 2020 (H2020) and recently in Horizon Europe. Building on the previous 'Europe on the Move' of May 2017, the European Commission (EC) put forward a strategy to carry Europe to a leader position for automated and connected mobility [37].

AI has been comprehensively covered in many projects funded by the EC in the recent H2020 programme. In this study, we present an overview of the front-runner projects that deliver AI-enabled technologies for the benefit of smarter CCAM solutions. The projects were analysed by considering their focus on the concepts like explainability (E), trustworthiness (T), Robustness and Reliability (R), Fairness and accountability (F), Privacy-awareness (P), Security-awareness (S) and finally safety-awareness (SF). A subjective methodology is applied to score the level of E, T, R, F, P, S, and SF by analysing the (recent) project deliverables, patents, products and the papers published as the outputs of dissemination and exploitation outputs of the projects. The keywords related to the factors (E, T, R, F, P, S, and SF) and the techniques applied have been identified by the authors and scored as Low (L), Medium (M) or High (H) depending on the level of coverage. The results of the expert opinions about the selected projects (but not limited to) are presented in Table 1.

Table 1. Analysis of projects funded under the H2020 programme according to the scope of AI in terms of the factors E, T, R, F, P, S, and SF

Project	AI Scope	E	T	R	F	P	S	S
PLANET	Predictive Analytics/AI-based models for last-mile logistics and ocean shipping	L	M	H	L	M	H	M
ICT4CART	Dynamic adaptation of vehicle automation level based on infrastructure information, smart parking, etc.	L	H	M	M	H	H	M
ASSURED	AI-enabled smart charging optimisation for full-size Urban Heavy-Duty vehicles	L	M	M	L	M	M	H
LEVITATE	Connected and autonomous vehicles on Traffic, Safety, & Emissions optimisation and societal impact analysis	L	M	M	H	H	H	H
HADRIAN	characterization of novel driver roles and safer AD enhanced with AI	M	M	M	M	H	H	H
MEDIATOR	Corrective and preventive driver mediation, Integrating driver states into fitness values	H	M	M	M	H	M	H
SAFE-UP	Proactive safety systems and tools for a constantly upgrading road environment	H	H	M	M	H	H	H
RESIST	AI-enabled road infrastructure monitoring in case of extreme weather conditions	L	H	H	M	M	H	H
PONEPTIS	AI-enabled multisensor fusion and micro-climate monitoring for road status	L	M	M	M	M	H	H
SAFEWAY	GIS-based infrastructure management system for optimized response to extreme events of terrestrial transport networks	L	M	M	M	M	H	H
OSSCAR	AI-based understanding of future mixed traffic accident	H	H	H	M	H	H	H

VIRTUAL	Open-access virtual testing Protocols for enhanced road user safety	H	M	M	M	M	M	H
PIONEERS	safety-critical accident scenarios and multi-sensor fusion	H	M	M	M	M	M	H
GREEN CHARGE	Smart energy management and charging in sustainable urban mobility plans	L	M	M	L	M	H	H
MOISTER	AI-enabled EV sharing and smart parking & charging	L	M	M	L	M	H	H
INCIT-EV	Decision support systems for EV charging and mobility planning	L	M	M	M	M	M	M
VISION-xEV	Digital twin and vehicle models for optimised EV performance	L	M	M	M	M	H	M
Project	**AI Scope**	**E**	**T**	**R**	**F**	**P**	**S**	**S**
LEAD	Low-emission adaptive last-mile logistics through Digital Twins	L	M	M	M	M	M	M
ULaaDS	AI-enabled micro-logistics at the urban scale	L	M	M	M	H	H	M
DOMUS	AI-enabled user-centric design optimisation for efficient EV in urban trans.	M	M	M	H	H	M	M
Multi-Moby	Pre-emptive trail braking and tractor control	L	M	M	M	H	H	H
Park4SUMP	AI-enabled parking management in sustainable urban mobility planning	L	M	M	M	H	H	M
INDIMO	AI-supported inclusive and user-centric digital mobility	M	H	M	H	H	H	L
TRIPS	Intelligent AI and Augmented Reality UX interfaces for disabled	H	H	M	H	H	M	L
SPROUT	Co-created city-specific future urban mobility scenarios	H	M	L	L	M	M	L
MORE	AI-enabled Option generation, Co-created street designs and simulation of street activities	H	M	M	H	H	M	L
HARMONY	Harmonised spatial and multimodal transport planning tools	H	M	M	H	H	M	L
EVC1000	AI-supported holistic control strategy and EV demonstrator	L	M	M	L	M	M	H
TUBE	Identification of biomarkers for early detection of brain disease related to air pollution in CCAM activities	H	H	M	H	H	M	H
MODALES	AI-powered understanding of the nature of driving behaviour concerning vehicle emissions	H	H	M	H	H	M	H
DIAS	Detection of tampering using AI-powered On-Board Diagnostics and Monitoring	M	M	M	M	M	H	H
L3PILOT	Piloting Automated Driving on European Roads	M	H	H	M	H	H	H
ENSEMBLE	AI-powered platooning and situational awareness	L	M	M	M	H	H	H
HEAD START	AI-supported testing and validation procedures of CAD	L	L	L	L	M	H	M

Note that the projects listed in Table 1 are selected from the EU's Cordis platform that presents the repository of the EU-funded projects[1] and many of these are still ongoing and new advancements in AI-driven solutions can be reported. According to the analysis results, AI has been widely adopted in nearly all analysed projects. AI-powered solutions have been addressed mainly for optimising vehicle performance, monitoring vehicle status, sustainable urban mobility and transportation planning, charging and energy management, emission analysis, driver behaviour analysis, AD and CAVs, multimodal mobility, road infrastructure monitoring, safety-critical modelling, etc. It has also been identified that the majority of the projects did not address the explainability and trustworthiness as they did not position the 3WAI in the main scope. However, there is a significant effort in covering the privacy, security and safety factors and topics like fairness, robustness and reliability are gaining importance.

VI. Conclusions and Future Directions

This paper presents a strategic overview of the strategy transition towards more TAI and XAI-oriented smart solutions, namely 3WAI, in line with the CCAM vision of Europe. This vision applies to integrated and multimodal transportation strategies in many countries where AVs are getting widespread for both people and goods transportation. The visionary approach tackles the key research directions like trust, explainability and interpretability, bias-awareness, situational awareness and V&V in the CCAM domain. The paper also presents a dynamic n-sprint V-cycle methodology presenting an effective strategy to improve the 3WAI solutions in new generation AI-based systems. A review of the EU-funded projects in the recent H2020 programme shows that the concept of explainable and trustworthy AI has not been fully covered in current projects. However, it is noted that there is an increasing consciousness of developing more explainable interfaces and trustworthy solutions that may improve the acceptability of AI in the CCAM context. In further studies, reflections of the presented strategic vision are planned to be presented in CCAM applications which are supposed to rely on the proposed V-cycle methodology. The strategic vision is expected to be extended with new advancements in CAS and smart cyber-physical systems. Moreover, a conceptual semantic framework is planned to be developed which is supposed to be built on an ontological basis. The ontology-based approach will be used to describe, model and analyse the V-cycle operations and the interconnections between cycle iterations, i.e. conducting V-cycle functions.

Acknowledgement

This work relies on the ongoing EUREKA ITEA3 project "Trustworthy and Smart Communities of Cyber-Physical Systems" (TioCPS, nr. 18008) and Horizon 2020 project "Data-based analysis for SAFETY and security protection FOR detection, prevention, mitigation and response in trans-modal metro and RAILway networkS" (Safety4Rails, nr. 883532).

[1] https://cordis.europa.eu/

REFERENCES

[1] https://www.ccam.eu/

[2] European Commission, Directorate-General for Mobility and Transport, Next steps towards 'Vision Zero': EU road safety policy framework 2021-2030, Publications Office, 2020.

[3] http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes

[4] https://transport.ec.europa.eu/news/european-commission-welcomes-launch-global-plan-un-decade-action-road-safety-2021-2030-2021-10_en

[5] https://h2020-trustonomy.eu/commission-presents-its-plan-for-green-smart-and-affordable-mobility-automated-mobility-a-key-factor/

[6] https://spectrum.ieee.org/driverless-cars-inspire-both-fear-and-hope

[7] Deloitte, Investing in trustworthy AI, 2021 Japanese Cabinet Office, https://www8.cao.go.jp/cstp/english/society5_0/index.html

[8] https://ieee-cas.org/sites/default/files/cass-sp-2020-2024.pdf

[9] Gunning, D.; Aha, D. DARPA's explainable artificial intelligence (XAI) program. AI Mag. 2019, 40, 44–58 .10.1609/aimag.v40i2.2850

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[11] Ha, Taehyun, et al. "Effects of explanation types and perceived risk on trust in autonomous vehicles." Transp. res. part F, 73 (2020): 271-280

[12] Daniel J. Fremont, Alberto L. Sangiovanni-Vincentelli, Sanjit A. Seshia: Safety in Autonomous Driving: Can Tools Offer Guarantees? DAC 2021: 1311-1314.

[13] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature communications, 11(1), 1-10.

[14] https://ai-fairness-360.org/

[15] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), 18.

[16] Shrikumar, A., Greenside, P., & Kundaje, A. (2017, July). Learning important features through propagating activation differences. In International conference on machine learning (pp. 3145-3153). PMLR.

[17] Ustun, B.; Rudin, C. Supersparse linear integer models for optimized medical scoring systems. Mach. Learn. 2016, 102, 349–391.

[18] Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, 2017; pp. 4765–4774.

[19] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " 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 (pp. 1135-1144).

[20] Omeiza, D., Webb, H., Jirotka, M., & Kunze, L. (2021). Explanations in autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems.

[21] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proc. of the IEEE conf. on computer vision & pattern recognition (pp. 2921-2929).

[22] Bojarski, M., Choromanska, A., Choromanski, K., Firner, B., Ackel, L. J., Muller, U., ... & Zieba, K. (2018, May). Visualbackprop: Efficient visualization of cnns for autonomous driving. In 2018 IEEE Int. Conf. on Robotics and Automation (ICRA) (pp. 4701-4708). IEEE.

[23] Karim, M. M., Li, Y., & Qin, R. (2021). Towards explainable artificial intelligence (XAI) for early anticipation of traffic accidents. arXiv preprint arXiv:2108.00273.

[24] Yu, S. Y., Malawade, A. V., Muthirayan, D., Khargonekar, P. P., & Al Faruque, M. A. (2021). Scene-graph augmented data-driven risk assessment of autonomous vehicle decisions. IEEE Transactions on Intelligent Transportation Systems.

[25] Gadd, M., De Martini, D., Marchegiani, L., Newman, P., & Kunze, L. (2020, May). Sense–Assess–eXplain (SAX): Building trust in autonomous vehicles in challenging real-world driving scenarios. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 150-155). IEEE.

[26] Nahata, R., Omeiza, D., Howard, R., & Kunze, L. (2021, September). Assessing and explaining collision risk in dynamic environments for autonomous driving safety. In 2021 IEEE International Intelligent Transportation Systems Conference (ITSC) (pp. 223-230). IEEE.

[27] Wiehr, F., Hirsch, A., Schmitz, L., Knieriemen, N., Krüger, A., Kovtunova, A., ... & Hoffmann, J. (2021). Why Do I Have to Take Over Control? Evaluating Safe Handovers with Advance Notice and Explanations in HAD. Proc. Int. Conf. on Mul. Inter. (pp. 308-317).

[28] Mitchell, R., Frank, E., & Holmes, G. (2020). GPUTreeShap: Massively Parallel Exact Calculation of SHAP Scores for Tree Ensembles. arXiv preprint arXiv:2010.13972.

[29] Zhang, Q., Yang, X. J., & Robert, L. P. (2021). What and When to Explain? A Survey of the Impact of Explanation on Attitudes Toward Adopting Automated Vehicles. IEEE Access, 9, 159533-159540.

[30] Endsley, M. R. (2020). Situation awareness in driving. Handbook of human factors for automated, connected and intelligent vehicles. London: Taylor and Francis.

[31] McAree, O., Aitken, J.M., Veres, S.M.: Towards artificial situation awareness by autonomous vehicles. IFAC-Pap. 50(1) (2017) 7038–43

[32] Li, G., Yang, Y., Zhang, T., Qu, X., Cao, D., Cheng, B., & Li, K. (2021). Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios. Trans. Res. P. C, 122, 102820.

[33] Dreossi, T., Fremont, D. J., Ghosh, S., Kim, E., Ravanbakhsh, H., Vazquez-Chanlatte, M., & Seshia, S. A. (2019, July). Verifai: A toolkit for the formal design and analysis of artificial intelligence-based systems. In International Conference on Computer Aided Verification (pp. 432-442). Springer, Cham.

[34] Riedmaier, S., Schneider, D., Watzenig, D., Diermeyer, F., & Schick, B. (2021). Model validation and scenario selection for virtual-based homologation of automated vehicles. Applied Sciences, 11(1), 35.

[35] Barbosa, R., Basagiannis, S., Giantamidis, G., Becker, H., Ferrari, E., Jahic, J., ... & Sangchoolie, B. (2020, August). The VALU3S ECSEL project: verification and validation of automated systems safety and security. 23rd Euromicro Conf. on Dig. Sys. Des. pp. 352-359. IEEE.

[36] Jingtao, Z. H. A. N. G. "The trend and suggestions of application of digital twin in intelligent transportation." ICT& Policy 46.3 (2020): 24.

[37] Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee, The Committee of the Regions "On the road to automated mobility: An EU strategy for mobility of the future", COM/2018/283 final.