Engineering Future Critical CPSs with Trustworthy GenAI Across the Lifecycle

Alessio Bucaioni
Mälardalen University
Sweden
alessio.bucaioni@mdu.se

Antonio Cicchetti
Mälardalen University
Sweden
antonio.cicchetti@mdu.se

Gordana Dodig Crnkovic
Mälardalen University
Sweden
gordana.dodig-crnkovic@mdu.se

Romina Spalazzese
Malmö University
Sweden
romina.spalazzese@mau.se

Emma Söderberg
Lund University
Sweden
emma.soderberg@lu.se

Dániel Varró
Linköping University
Sweden
daniel.varro@liu.se

Abstract

One of the most transformative developments today is the integration of generative artificial intelligence into the development of critical software-intensive cyber-physical systems. From autonomous vehicles to industrial robotics, these systems are entering a new era shaped by artificial intelligence-driven development and automation. In this paper, we consider software engineering, artificial intelligence, artificial intelligence ethics, and social aspects, to explore how such technologies can be harnessed safely, transparently, and with human values at the center. Our contributions include a *vision* for software engineering, guiding the engineering of future trustworthy safety-critical cyber-physical systems under the influence of generative artificial intelligence. We critically analyze how three established certification principles can be leveraged to cope with the societal and technical tensions introduced by generative artificial intelligence adoption, and propose a research and practice agenda to ensure that future cyber-physical systems development and operations cycle remain trustworthy, both from a system (hardware and software) and from a societal perspective.

Keywords

Software Engineering, Generative Artificial Intelligence, Trustworthy

1 Introduction

Safety-critical cyber-physical systems (CPSs), such as self-driving vehicles, industrial robots, and smart factories, play an increasingly

central role in modern society. Their failure may lead to severe financial loss, environmental damage, or even human casualties. While automation has long improved productivity in CPS engineering (e.g., through code generation and model-based systems engineering), assurance standards (e.g., DO-178C [33] for avionics, ISO 26262 [21] for automotive) mandate rigorous evidence that such systems are safe and secure for their intended purpose. Certification processes thus enshrine established principles of accountability and societal trust (Figure 1). Three principles are particularly foundational:

- *Human responsibility*, every engineering decision must be traceable to a human responsible, across the system lifecycle.
- *Independent eyes*, validation and verification across the cycle to be made by experts not involved in the design.
- *Human-first*, requires safety-critical CPSs that collaborate with humans to consider not only critical system qualities but also broader ethical, legal, and societal concerns.
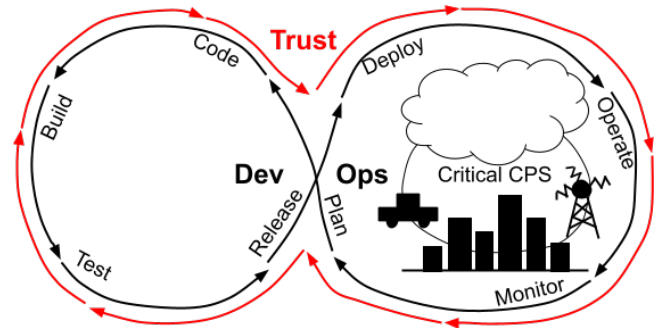

Figure 1: Overview of the Lifecycle When Developing Critical CPS With Trustworthy DevOps.

The rapid emergence of generative artificial intelligence (GenAI), in particular large language models (LLMs), directly challenges these principles. Early studies indicate that LLMs can support requirements engineering, architectural design, domain modeling, testing, simulation, and documentation [6, 32]. For industry, this suggests major productivity gains, especially in labor-intensive tasks such as testing and documentation. At the same time, visions of fully autonomous LLM agents [16] capable of developing complete software systems raise fundamental questions: *How can one*

guarantee to build critical CPSs with trustworthy AI techniques? How can established certification practices evolve to address AI-driven development? How do we ensure that human engineers, regulators, and affected communities remain in meaningful control?

Trustworthiness is a very complex and sensitive concern along the whole system lifecycle and including both objective and subjective perspectives. This is exacerbated even more when considering LLMs and GenAI coming into play in future critical CPSs. Efforts have been done to provide an overview and guidelines about trustworthiness and Artificial Intelligence (AI), e.g., [1, 24, 28]. Informally, trustworthiness is about how well a system has been designed and built to do what it is supposed to do. In the scope of this paper, we refer to trustworthiness as a quality aspect of a development life cycle driven by GenAI, while keeping the traditional quality attributes of critical CPSs intact. In the described context, the above questions highlight a growing socio-technical tension. On the one hand, industry is motivated by efficiency and cost reduction. On the other, society depends on, e.g., the transparency, accountability, and fairness of critical systems that affect lives and livelihoods. These concerns echo critiques raised by leading Software Engineering (SE) pioneers, including Parnas[1], Broy and Selic [5], who caution against over-reliance on automation without robust assurance frameworks.

This paper contributes a vision for the future of SE in safety-critical CPSs under the influence of GenAI. We articulate the societal and technical tensions that GenAI adoption introduces, critically examine how established certification principles may be upheld or reinterpreted, and propose a research and practice agenda to ensure that future CPS development remains accountable, human-centered, and societally trustworthy. By doing so, we aim to stimulate discussion within the SE community on how CPS engineering can adapt to the promises and risks of GenAI, while reinforcing software's role as a catalyst for societal trust and responsibility.

2 Background and State of the Art

The engineering of AI-enabled safety-critical CPSs sits at the intersection of AI, SE, and system safety. Below, we highlight selected works that were not part of a systematic review but were manually chosen for their relevance. A number of studies have been conducted in the area of CPSs and LLMs [8, 10, 11, 25, 27, 39]. Investigations include aspects like, e.g., architectures, testing, integration frameworks, human behavior modeling, and AI resilience in CPS where LLMs play roles, directly applicable to designing trustworthy CPS with humans in the loop across the lifecycle. Other efforts focused on providing a detailed, critical overview of the intersection between humans and CPSs from socio-technical and practical industry perspectives, highlighting gaps and avenues for future research and practice [7]. To frame our vision, this section reviews the current research frontier on AI-enabled safety-critical CPSs, examines the emerging role of GenAI in SE practice, identifies the central research gap that motivates our work, and highlights why SE plays a central role.

2.1 Research Frontier on AI-enabled Safety-Critical CPSs

AI-enabled safety-critical CPSs integrate AI-based components with traditional hardware and software to increase autonomy and adaptivity in operation. The international research community increasingly recognizes the need for robust, safe, and trustworthy solutions in this domain. Examples of flagship initiatives include TAILOR (Trustworthy AI Integrating Learning, Optimization and Reasoning) [17], an EU Network of Excellence aiming to develop joint infrastructure such as joint PhD programmes, industry transfer labs, and testbeds. In the US, DARPA's Assurance of AI-Based Systems (AAIS) program aims to assure the safety and correctness of AI-enabled CPSs. Several initiatives on AI-enabled safety-critical CPSs (e.g., [3, 15, 17, 31]) systematically explore how to improve quality and autonomy. Assurance techniques have focused on attributes such as explainability, safety, fairness, liability, and sustainability, often leveraging formal methods, runtime verification, or architectural reasoning. However, these efforts remain largely decoupled from social, ethical, and human-centered assurance mechanisms [18, 36], despite their central importance in safety-critical domains. Furthermore, existing approaches typically assume that system development remains exclusively human-controlled, with AI tools providing limited decision support rather than autonomous action.
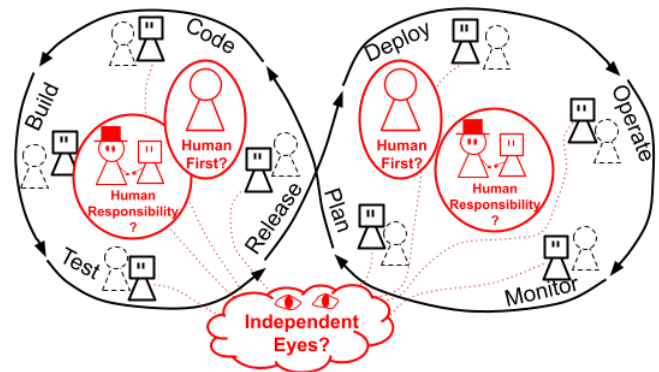
Figure 2: Devops Flow With Bots And People At Different Parts Of The Cycle, Illustrating The Principles 'Human Responsibility', 'Independent Eyes', And 'Human First' And Risks Connected To Trust.

2.2 The Emerging Role of GenAI

Recent advances in LLMs have introduced a qualitatively different paradigm. Beyond decision-support, LLM-based agents promise to increase autonomy in the entire development process. Early work demonstrates that LLMs can derive requirements from safety standards [32], propose high-level architectural designs [37], generate domain models from textual specifications [6], produce test scenarios [35], synthesize code through natural language prompts [29], create or refine documentation [4]. These developments go well beyond augmenting human engineers; they raise the possibility of semi-autonomous or fully autonomous development pipelines, disrupting traditional assumptions about certification and human accountability, illustrated in Figure 2. However, with the introduction

---

of genAI into the DevOps flow, we may lose track of foundational principles of accountability and societal trust.

2.3 Research Gap

Despite advances in GenAI, a central challenge remains unresolved: *How can AI-generated artifacts be aligned with human-controlled, traceable, and certifiable development and operations pipelines in safety-critical domains?*

This challenge is particularly severe in safety-critical CPSs, where failures have direct societal consequences. We build on the premise that *fully autonomous AI agents should not be developed for safety-critical CPSs* [30], and argue instead for rigorous, continuous human oversight.

Our aim is to enable a trustworthy and human-controlled use of GenAI in engineering AI-enabled safety-critical CPSs, explicitly incorporating ethical, legal, and societal concerns. By complementing and extending existing efforts, we aim to inform the definition of next-generation standards and research and practice priorities for LLM-assisted CPS design.

2.4 Why Software Engineering Matters for AI

AI research typically advances solutions to narrowly defined technical problems, but safety-critical CPSs demand more than capability. SE can provide the systemic perspective that connects AI components to stakeholder needs, lifecycle assurance, and certification. For instance, requirements engineering ensures that AI functions address the right problems, while verification and validation assess adequacy in context rather than performance in isolation. SE also offers mechanisms, traceability, standards, and independent review, to embed ethical, legal, and societal constraints into development pipelines. In this sense, SE is not auxiliary to AI but essential for transforming powerful models into trustworthy systems. Without SE's systemic approach, AI-enabled CPSs risk being technically impressive yet socially and ethically unfit for deployment.

3 A Vision for Future TrustDevOps CPSs with GenAI

A central question motivates our work: *How can GenAI be incorporated into the development and operations of safety-critical CPSs while preserving safety, transparency, and human-centered values?*

Ultimately, our argument is that SE is indispensable for transforming AI from a collection of powerful problem-solving tools into trustworthy, certifiable, and societally responsible CPSs. We argue that SE is not merely a supporting discipline in this transformation, but a driving force that must shape how GenAI is responsibly integrated into CPS development. Addressing this question requires a long-term, multidisciplinary perspective that combines foundational technical advances with ethical, legal, and societal considerations.

Building on SE's systemic role in aligning AI with stakeholder needs and societal requirements, our *vision* is for *safe, ethical, and value-driven CPSs accelerated by generative AI technology, with continual human control embedded throughout the lifecycle.* This vision builds on the established certification principles of human responsibility, independent oversight, and human-first design, which we reinterpret in the era of GenAI:

- *Human responsibility*: LLMs should act as collaborators within certifiable workflows, not as autonomous developers without any human control.
- *Independent eyes*: Assurance must become an interactive, human-in-the-loop activity rather than a static, post hoc exercise.
- *Human-first*: Ethical, legal, and societal concerns should be integrated into CPS development and operations from the outset, not treated as downstream constraints.

This vision calls for a new generation of socio-technical methods that make AI-enabled safety-critical CPSs not only technically robust, but also accountable, transparent, and societally trustworthy.

Figure 3: An illustration of the key principles and research objectives we envision for future development of safe, ethical, and value-driven CPSs with generative AI.

4 Research and Practice Agenda

To operationalize our vision, we propose a thematic research agenda that extends current work on AI-enabled safety-critical CPSs by explicitly embedding GenAI responsibly into safety-critical CPSs across the TrustDevOps life cycle, guided by three life cycle principles, illustrated in Figure 3.

Th1 Human control for LLM-assisted design ↦ Principle of human responsibility. Current research often treats LLM output as equivalent to human engineering work, overlooking the tendency of humans to overtrust automatically generated artifacts [29]. We propose that each automated step should be regarded as a potential loss of control unless correctness can be formally justified. Responsibility must remain with human engineers.

– *Collaborative CPS design environments* will allow human engineers and AI agents to co-design systems within traceable, certifiable workflows [6, 32].
– *Socially traceable artifacts* will embed accountability and rationale in all generated outputs, ensuring inspectability by both humans and machines [2].
– *Control surfaces for AI* will provide interfaces and toolchains for engineers to define behavioral boundaries and failure modes of AI components [20, 34].

Th2 **Human control for LLM-assisted assurance ↦ Independent eyes principle.** Assurance today often relies on post hoc explainability or audit measures. Instead, we envision *human-centric assurance frameworks* that integrate LLM agents and ethical values directly into development workflows.
– *Assurance as interaction* will enable dynamic oversight and real-time intervention, moving beyond static documentation [15].
– *Adaptive certification methods* will keep assurance cases current in systems that evolve through AI-generated content [9, 37].
– *Transparent pipelines* will ensure continuous verification and accountability [17, 22].

Th3 **Embedding ethical, legal, and societal concerns ↦ Human-first principle.** Ethical, legal, and societal dimensions concerns must be treated as non-negotiable constraints across the pipeline. Building on interdisciplinary work in ethics, law, and the social sciences [12, 18, 36, 38], we propose methods that embed societal values directly into CPS design processes, ensuring that generative AI applications remain accountable and socially legitimate.
– *Value-sensitive methods*: embed societal concerns into workflows [14].
– *Governance models*: align engineers, ethicists, and regulators [13].
– *Ethical assurance cases*: connect system behavior to normative commitments [19, 30, 36].

Together, these three themes outline a coherent research and practice agenda for LLM-assisted CPS development. They connect technical innovation with ethical, legal, and societal imperatives, ensuring that future safety-critical CPSs remain certifiable, accountable, and human-centered. Importantly, these themes cannot be addressed within SE alone. They require sustained collaboration across different disciplines—including law, ethics, human-computer interaction, and sociology, alongside industrial stakeholders and regulatory bodies. Future investigations should be systemic and transdisciplinary, as also pointed out in [19]. This aligns with Leveson's vision that the analysis of software in isolation, does not guarantee the safety of the whole software-hardware system: a system-level analysis is fundamental [26]. Separation of engineering from human, social and organizational factors is no longer possible: we must take a systemic view.

By articulating these directions, we aim to initiate a community effort toward establishing next-generation standards, methods, and interdisciplinary practices. Advancing this agenda will help ensure that GenAI becomes not a source of risk, but a foundation for trustworthy, socially legitimate, and societally beneficial CPS development.

5 Novelty and Contributions

This paper advances life cycle principles for trustworthy GenAI-enabled CPSs, addressing design, verification, deployment, and certification together. Existing frameworks (e.g., ISO 26262 [21], DO-178C [33], TAILOR [17]) focus mainly on technical assurance or abstract ethics, but do not capture how GenAI reshapes the SecDevOp pipeline. We propose three principles: Principle of human responsibility, the Independent eyes principle (Independent assurance principle), and the Human-first principle, clarifying the distinction between human-centered (how?) and human-first (why) approaches [20, 34]. These principles extend into an actionable agenda: maintaining accountability in design [2, 6, 32], enabling adaptive assurance [9, 22, 23, 37], and embedding ethical and societal trust as non-negotiable constraints [13, 14, 18, 30, 36]. Our contribution is conceptual, methodological, and normative: positioning SE as the discipline that must drive the responsible life cycle assurance of GenAI-enabled CPSs.

6 Conclusion

In this paper, we explored the open challenge of enabling a trustworthy and human-controlled use of GenAI in engineering AI-enabled safety-critical CPSs explicitly incorporating ethical, legal, and societal concerns.

Our vision is that SE plays a central role to engineer future trustworthy safety-critical CPSs (goal) with the support of GenAI (enabler). More operationally, we provided the Trustworthy DevOps (TrustDevOps) view by leveraging (i) three foundational principles (human responsibility, independent eyes, and human-first) from certification processes, and (ii) the development and operations cycle.

Our vision also points out that needed future investigations must be systemic and transdisciplinary. Involved stakeholders such as engineers, certifiers, policy-makers, and user communities, all bring specific values and concerns that must be taken into organic consideration along the TrustDevOps cycles.

Acknowledgments

References

[1] 2020. *ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence.* Retrieved October 18, 2023 from https://www.iso.org/standard/77608.html
[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi T. Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 3, 13 pages. doi:10.1145/3290605.3300233

[3] Len Bass and Qinghua Lu. 2025. Quality Attributes in the Age of Artificial Intelligence. *SIGSOFT Softw. Eng. Notes* 50, 3 (2025), 43–44. doi:10.1145/3743095.3743103

[4] Henok Birru, Antonio Cicchetti, and Malvina Latifaj. 2025. Supporting Automated Documentation Updates in Continuous Software Development with Large Language Models. In *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering, ENASE 2025, Porto, Portugal, April 4-6, 2025*, Mike Mannion, Tomi Männistö, and Leszek A. Maciaszek (Eds.). SCITEPRESS, 92–106. doi:10.5220/0013286800003928

[5] Manfred Broy, Bran Selić, and John Favaro. 2025. The Effects of Hype in the Software Domain: Causes, Consequences, and Mitigations. *IEEE Softw.* 42, 2 (March 2025), 98–102. doi:10.1109/MS.2024.3511732

[6] Kua Chen, Yujing Yang, Boqi Chen, José Antonio Hernández López, Gunter Mussbacher, and Dániel Varró. 2023. Automated Domain Modeling with Large Language Models: A Comparative Study. In *26th Int. Conf. on Model Driven Engineering Lang. and Syst., MODELS*. IEEE, 162–172.

[7] T. Clemmensen, M. Tourchi Moghaddam, and J. Nørbjerg. 2025. Cyber-physical Systems with Human-in-the-Loop: A Systematic Review of Socio-technical Perspectives. *Journal of Systems and Software* 226 (2025), 112348. doi:10.1016/j.jss.2025.112348

[8] Werner Damm, David Hess, Mark Schweda, Janos Sztipanovits, Klaus Bengler, Bianca Biebl, Martin Fränzle, Willem Hagemann, Moritz Held, Klas Ihme, Severin Kacianka, Alyssa J. Kerscher, Sebastian Lehnhoff, Andreas Luedtke, Alexander Pretschner, Astrid Rakow, Jochem Rieger, Daniel Sonntag, Maike Schwammberger, Benedikt Austel, Anirudh Unni, and Eric Veith. 2024. A Reference Architecture of Human Cyber-Physical Systems – Part I: Fundamental Concepts. *ACM Transactions on Cyber-Physical Systems* 8, 1, Article 2 (Jan. 2024), 32 pages. doi:10.1145/3622879

[9] Ewen Denney and Ganesh Pai. 2013. *Evidence Arguments for Using Formal Methods in Software Certification*. Technical Report. NASA / NTRS. https://ntrs.nasa.gov/citations/20140011544

[10] Weiping Ding, Mohamed Abdel-Basset, Ahmed M. Ali, and Nour Moustafa. 2025. Large language models for cyber resilience: A comprehensive review, challenges, and future perspectives. *Applied Soft Computing* 170 (2025), 112663. doi:10.1016/j.asoc.2024.112663

[11] Maral Doctorarastoo, Katherine Flanigan, Mario Bergés, and Christopher McComb. 2023. Exploring the Potentials and Challenges of Cyber-Physical-Social Infrastructure Systems for Achieving Human-Centered Objectives. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23)*. ACM, 385–389. doi:10.1145/3600100.3626340

[12] G. Dodig-Crnkovic. 2020. The relation between Future State Maximization and von Foerster's Ethical Imperative. *Constructivist Foundations* 16, 1 (2020), 62–64.

[13] Luciano Floridi and Josh Cowls. 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review* 1, 1 (2019). doi:10.1162/99608f92.8cd550d1

[14] Batya Friedman, Peter H. Kahn Jr., and Alan Borning. 2008. Value Sensitive Design and Information Systems. In *The Handbook of Information and Computer Ethics*, Kenneth Einar Himma and Herman T. Tavani (Eds.). Wiley, Hoboken, NJ, 69–101.

[15] Odd Ivar Haugen, Dag McGeorge, Tore Myhrvold, Christian Agrell, and Andreas Hafver. 2025. Assurance of AI-enabled systems. In *8th Int. Conf. on Advances in Artif. Intell. (ICAAI '24)*. ACM, 7 pages. doi:10.1145/3704137.3704153

[16] Junda He, Christoph Treude, and David Lo. 2025. LLM-Based Multi-Agent Systems for Software Engineering: Literature Review, Vision, and the Road Ahead. *ACM Trans. Softw. Eng. Meth.*, Article 124 (2025), 30 pages. doi:10.1145/3712003

[17] Fredrik Heintz, Michela Milano, and Barry O'Sullivan (Eds.). 2021. *Trustworthy AI – Integrating Learning, Optimization and Reasoning*. LNCS, Vol. 12641. Springer Nature. doi:10.1007/978-3-030-73959-1

[18] Tobias Holstein, Gordana Dodig-Crnkovic, and Patrizio Pelliccione. 2020. Real-world ethics for self-driving cars. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Companion Proceedings (ICSE '20)*. Association for Computing Machinery, 328–329. doi:10.1145/3377812.3390801

[19] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione. 2021. Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In *Machine Law, Ethics, and Morality in the Age of Artificial Intelligence*, Steven John Thompson (Ed.). IGI Global.

[20] Andreas Holzinger, Peter Kieseberg, Edgar Weippl, and Antonella M. Tjoa. 2019. Causability and Explainability of Artificial Intelligence in Medicine. *Frontiers in Artificial Intelligence* 2 (2019), 6. doi:10.3389/frai.2019.00006

[21] International Organization for Standardization. 2018. ISO 26262-1, Road vehicles: Functional safety.

[22] Susmit Jha, John Rushby, and Natarajan Shankar. 2020. Model-Centered Assurance for Autonomous Systems. In *Computer Safety, Reliability, and Security – SAFECOMP 2020*. Lecture Notes in Computer Science, Vol. 12234. Springer, 228–243. doi:10.1007/978-3-030-54549-9_15

[23] Ariel S. Kapusta, David Jin, Peter M. Teague, Robert A. Houston, Jonathan B. Elliott, Grace Y. Park, and Shelby S. Holdren. 2025. A Framework for the Assurance of AI-Enabled Systems. arXiv:2504.16937 [cs.AI] https://arxiv.org/abs/2504.16937

[24] Davinder Kaur, Suleyman Uslu, Kaley J Rittichier, and Arjan Durresi. 2022. Trustworthy artificial intelligence: a review. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–38.

[25] Sanghoon Lee, Jiyeong Chae, Haewon Jeon, Taehyun Kim, Yeong-Gi Hong, Doo-Sik Um, Taewoo Kim, and Kyung-Joon Park. 2025. Cyber-Physical AI: Systematic Research Domain for Integrating AI and Cyber-Physical Systems. *ACM Transactions on Cyber-Physical Systems* 9, 2, Article 19 (April 2025), 33 pages. doi:10.1145/3721437

[26] Nancy Leveson. 2020. Are you sure your software will not kill anyone? *Commun. ACM* 63, 2 (Jan. 2020), 25–28. doi:10.1145/3376127

[27] Xiaoyu Luo, Daping Liu, Fan Dang, and Hanjiang Luo. 2024. Integration of LLMs and the Physical World: Research and Application. In *Proceedings of the ACM Turing Award Celebration Conference - China 2024 (ACM-TURC '24)*. ACM, 1–5. doi:10.1145/3674399.3674402

[28] Madiega Tambiama André. 2024. Artificial intelligence act. https://bitly.cx/SDeIxH

[29] Alan T. McCabe, Moa Björkman, Joel Engström, Peng Kuang, Emma Söderberg, and Luke Church. 2024. Ironies of Programming Automation: Exploring the Experience of Code Synthesis via Large Language Models. In *Companion Proceedings of the 8th International Conference on the Art, Science, and Engineering of Programming* (Lund, Sweden) *(Programming '24)*. Association for Computing Machinery, New York, NY, USA, 12–21. doi:10.1145/3660829.3660835

[30] Margaret Mitchell, Avijit Ghosh, Alexandra Sasha Luccioni, and Giada Pistilli. 2025. Fully Autonomous AI Agents Should Not be Developed. arXiv:2502.02649 [cs.AI]

[31] Roger Nazir, Alessio Bucaioni, and Patrizio Pelliccione. 2024. Architecting ML-enabled systems: Challenges, best practices, and design decisions. *Journal of Systems and Software* 207 (2024), 111860. doi:10.1016/j.jss.2023.111860

[32] Ali Nouri, Beatriz Cabrero-Daniel, Fredrik Torner, Hakan Sivencrona, and Christian Berger. 2024. Engineering Safety Requirements for Autonomous Driving with Large Language Models. In *IEEE 32nd Int. Requirements Engineering Conference (RE)*. IEEE, 218–228. doi:10.1109/RE59067.2024.00029

[33] RTCA. 2012. DO-178C, Software Considerations in Airborne Systems and Equipment Certification.

[34] Ben Shneiderman. 2021. Responsible AI: bridging from ethics to practice. *Commun. ACM* 64, 8 (July 2021), 32–35. doi:10.1145/3445973

[35] Shuncheng Tang, Zhenya Zhang, Jixiang Zhou, Lei Lei, Yuan Zhou, and Yinxing Xue. 2024. LeGEND: A Top-Down Approach to Scenario Generation of Autonomous Driving Systems Assisted by Large Language Models. In *39th IEEE/ACM Int. Conf. on Automated Software Engineering (ASE '24)*. ACM, 12 pages. doi:10.1145/3691620.3695520

[36] Abhilash Thekkilakattil and Gordana Dodig-Crnkovic. 2015. Ethics Aspects of Embedded and Cyber-Physical Systems. In *39th Annual Int. Computers, Software & Applications Conf. (COMPSAC), Symposium on Embedded & Cyber-Physical Environments (ECPE)*. IEEE, 39–44. doi:10.1109/COMPSAC.2015.41

[37] Louis Richard Timperley, Lucy Berthoud, Chris Snider, and Theo Tryfonas. 2025. Assessment of large language models for use in generative design of model based spacecraft system architectures. *Journal of Engineering Design* 36, 4 (2025), 550–570. doi:10.1080/09544828.2025.2453401

[38] Maryam Zahid, Alessio Bucaioni, and Francesco Flammini. 2024. Model-based Trustworthiness Evaluation of Autonomous Cyber-Physical Production Systems: A Systematic Mapping Study. *ACM Comput. Surv.* 56, 6 (2024), 157:1–157:28. doi:10.1145/3640314

[39] Xi Zheng, Aloysius K. Mok, Ruzica Piskac, Yong Jae Lee, Bhaskar Krishnamachari, Dakai Zhu, Oleg Sokolsky, and Insup Lee. 2023. Testing Learning-Enabled Cyber-Physical Systems with Large-Language Models: A Formal Approach. arXiv:2311.07377 [cs] doi:10.48550/arXiv.2311.07377