Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 277 (2026) 111–120

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Industry of the Future and Smart Manufacturing
(former International Conference on Industry 4.0 and Smart Manufacturing)

A conceptual framework for Zero-Error AI agents using Digital AI Passport approach

Foivos Psarommatis[a,b,e]*, George Siaterlis[a], Dimitris Apostollou[c], Alexandros Bousdekis [c], John Soldatos[d], Chrysostomos Stylios[f]

[a] *Zerofect GmbH, Albisblick 49, Allenwinden, 6319, Switzerland*
[b] *University of Oslo, SIRIUS labs, Gaustadalleen 23B, 0373, Oslo, Norway*
[c] *Information Management Unit, Institute of Communication and Computer Systems (ICCS), School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), 10682, Athens, Greece*
[d] *Netcompany-Intrasoft*
[e] *Dept of Informatics and Telecommunications, University of Ioannina*
[f] *Industrial Systems Institute, Athena Research Center, Patras, Greece*

Abstract

Artificial Intelligence (AI) systems are increasingly deployed in high-stakes domains where even minor errors can have severe consequences. Existing lifecycle documentation frameworks—such as Model Cards, Datasheets for Datasets, and AI FactSheets—provide transparency but lack continuous performance monitoring, lifecycle-wide traceability, and mechanisms to preserve full operational context. This paper proposes the Digital AI Passport (DAIP)—a conceptual framework adapted from the Digital Product Passport (DPP) model in manufacturing—to establish a structured, auditable identity for AI models across their lifecycle. At its core is the Zero-Error AI Agent, a concept transferred from the Zero Defect Manufacturing (ZDM) domain, which operates through a two-level triggering mechanism: monitoring and predicting changes in working conditions, followed by detecting or predicting performance loss to enable corrective or preventive actions. This approach combines proactive and reactive strategies to achieve near zero-error operation. Although conceptual, the DAIP provides a foundation for future implementations to validate its feasibility, scalability, and advantages over existing AI governance tools.

Keywords: Zero Error AI; Digital Passport, agents, framework, AI working conditions

* Corresponding author. Tel.: +41786666309
E-mail address: fp@zerofect.com

1. Introduction

Artificial Intelligence (AI) has transformed industries ranging from healthcare and finance to autonomous systems and manufacturing. While AI adoption accelerates, models remain vulnerable to errors due to biases, limited adaptability, and opaque decision-making. In high-risk domains, such errors can have severe consequences. Addressing these risks requires a lifecycle approach to AI governance. Inspired by the EU's Digital Product Passport (DPP)—originally designed for product traceability and circular economy goals—we extend these principles to AI assets via the Digital AI Passport (DAIP). DAIP enables structured identity, continuous performance tracking, and transparent auditability of AI agents, aligning reliability improvements with sustainability objectives. While the DPP was originally conceived within the EU's Ecodesign for Sustainable Products Regulation (ESPR) to enable a circular economy for physical products, the scope of this paper is not circular economy management. Instead, the DPP serves here as a conceptual and structural foundation for building a DAIP, enabling full lifecycle traceability, data integrity, and stakeholder-specific access control for AI models. This repurposing preserves the core principles of transparency, standardization, and accountability inherent in the DPP, but applies them to the domain of AI safety, governance, and reliability. By reducing unnecessary retraining and optimizing inference processes, DAIP can mitigate AI's energy footprint while improving trustworthiness Current AI validation techniques focus on improving accuracy and robustness, yet they fail to ensure a zero-error standard, particularly in high-risk scenarios where even minor mistakes can have catastrophic consequences. This highlights the urgent need for a structured, fail-proof mechanism that continuously validates, tracks, and optimizes AI agents to minimize errors [2].

In this work, the DPP concept is repurposed from its original application in physical product traceability to serve as a centralized, structured repository of lifecycle information for AI models. This approach ensures that every AI model is associated with a persistent digital identity, allowing full traceability of its development history, updates, performance metrics, and operational context. By maintaining data integrity and preserving the context of all recorded information, the DPP-inspired Digital AI Passport (DAIP) prevents the loss of critical metadata that often occurs when models are updated or transferred between systems. Furthermore, this structure facilitates controlled accessibility for different stakeholders—developers, auditors, and regulators—while enabling verifiable, immutable logging through hybrid cloud technologies (such as TSA (Timestamp service) certifications etc.). The adoption of the DPP framework in this context therefore provides a proven, standards-aligned foundation for accountability, interoperability, and lifecycle governance in AI systems, addressing gaps left by traditional validation and monitoring approaches.

Despite advances in techniques such as Explainable AI (XAI) and adversarial training, which improve interpretability and resilience, there is still no comprehensive framework capable of enforcing real-time error elimination while preserving AI adaptability and decision transparency throughout the model's lifecycle. Current validation strategies typically end after deployment and lack mechanisms to provide AI agents with a systematic, verifiable identity that ensures accountability across their operational history. This gap is especially critical in high-stakes environments, where the absence of continuous, auditable error management can lead to unacceptable risks.

While existing AI lifecycle documentation frameworks—such as Model Cards, Datasheets for Datasets, and AI FactSheets—provide valuable transparency [3], they lack continuous performance monitoring, lifecycle traceability, and secure data integrity mechanisms. The proposed DAIP extends these ideas by integrating TSA certification for immutable record-keeping, federated learning for adaptive performance improvement, and zero-trust security principles. This combination allows for both real-time accountability and full lifecycle context preservation, advancing the state-of-the-art beyond theoretical transparency tools toward an operational governance model inspired by DPP.

To address this gap, this paper proposes a conceptual framework that adapts principles from the manufacturing domain—specifically Zero Defect Manufacturing (ZDM)—and transfers them to the AI domain. In manufacturing, ZDM focuses on continuous monitoring, error prevention, self correction, and lifecycle optimization of physical products. Here, we extend these principles to AI models by treating each model as a "product" and its operational data as the equivalent of manufacturing process metrics. The proposed DAIP framework introduces a structured, auditable identity for AI agents, enabling real-time validation, lifecycle monitoring, and error minimization. By integrating information traceability techniques, federated learning for adaptive performance improvements, and zero-trust security for risk mitigation, DAIP creates an environment where AI agents can operate within a near-zero-error paradigm. This approach aims to make AI systems not only more transparent and trustworthy but also aligned with rigorous lifecycle governance practices that have proven successful in the manufacturing sector.

2. Literature Review

The adoption of Artificial Intelligence (AI) tools has surged across all sectors in recent years, driven by advancements in data processing capabilities and computing power [1]. AI solutions—particularly deep learning techniques that leverage artificial neural networks—are increasingly being developed and applied in various industries and sectors, from consumer-facing markets to advanced manufacturing [2]. Advanced AI techniques include Transfer Learning, Reinforcement Learning, and Deep Learning Neural Networks, while other, more traditional techniques such as ensemble learning (e.g., random forest, gradient boosting), clustering (e.g., k-means, tree-based, DBSCAN), decision tree learning, dimensionality reduction (e.g., PCA, t-SNE), regression analysis (e.g., linear, logistic, lasso), and linear classifiers (e.g., Fisher's linear discriminant, SVM) continue to play a crucial role [1]. AI is considered critical to achieving sustainability goals and expected to be utilized broadly across various industries and sectors, with automation and intelligent systems reducing the need for human intervention [2]. For instance, Vinuesa et al. [3] highlight AI's potential to advance 134 of the total 169 targets set forth in the United Nations' Sustainable Development Goals (SDGs), across all three groups of Society, Environment and Economy. However, problems arise regarding the quality of the data and its ownership. Nishant et al. [4] observe that AI is primarily utilized for problem solving and technical solutions such as supply chain intelligent logistics, broader smart manufacturing sector for resource efficiency [5], energy forecasting [6] or even in efficient yield production for the agriculture sector [7]or the healthcare sector for disease diagnosis and personalized medicine with an overall increased accuracy, minimizing human error [8]. However, Palomares et al. [9] argue that a hurdle in implementing such AI solutions is in utilizing big data to offset computational costs, aligning with our zero-error approach. This efficiency is exemplified in Zero Defect Manufacturing (ZDM) methodology, where Psarommatis et al.[10], [11] integrate machine learning and Digital Twins (DT) for real-time anomaly detection, showcasing that current AI implementations minimize errors under Industry 4.0 frameworks championed by initiatives like the European Commission's digitalization strategy [12].

However, despite these advancements, AI systems remain susceptible to errors, often stemming from opaque decision-making processes that obscure the origins and management of inaccuracies. This opacity represents a critical scientific gap: the inability to fully explain and control error generation in AI models, particularly in high-risk applications where precision is paramount. Existing methodologies, including XAI and adversarial training, enhance interpretability and robustness but fail to establish a zero-error benchmark or provide a comprehensive mechanism for continuous error tracking and mitigation throughout an AI system's lifecycle, something that should be considered in error optimization and the Fair AI principle [13] Rudin [14] further argues that relying on black-box models in critical domains often sacrifices accuracy for complexity, highlighting the need for inherently interpretable systems to reduce errors effectively. Holzinger et al. [15] emphasize that in domains like medicine, current XAI approaches fall short of delivering the precision required for error-free systems. This necessitates innovative approaches, such as AI agents and the integration of DPPs, to ensure transparency, adaptability, and error elimination.

AI Techniques and Performance Across Applications

AI's performance in critical applications hinges on balancing accuracy, efficiency, and transparency, yet significant trade-offs persist. Sophisticated models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) deliver high accuracy but often at the cost of interpretability, rendering error causation difficult to trace [15]. Lundberg and Lee [16] propose SHAP (SHapley Additive exPlanations) to clarify feature contributions in predictions, yet this method struggles to address real-time error dynamics in complex environments. Vaswani et al. [17] note that Transformer models, while effective for translation tasks due to their attention-based architecture, incur substantial computational overhead from multi-head attention, limiting scalability in resource-constrained settings. Similarly, Goodfellow et al. [18] highlight how adversarial examples reveal deep learning's vulnerability to small input perturbations, amplifying error risks in high-stakes domains. These gaps—opacity, inefficiency, and susceptibility to failure—underscore the need for frameworks like DAIP to enforce error-free operation across diverse applications.

According to Russell and Norvig [19], AI agents, defined as autonomous entities capable of perceiving their environment, reasoning, and acting accordingly, offer a potential solution to these challenges. Unlike static models, agents exhibit adaptability, enabling ongoing improvement and error correction. Agents employing Reinforcement Learning (RL) refine decision-making through iterative environmental feedback, a capability supported by Sutton and Barto [20] and demonstrated in healthcare where RL agents optimize treatment protocols based on patient responses [21]. The primary value of AI agents lies in their autonomy and capacity for real-time adaptation, facilitating proactive error mitigation. Nevertheless, limitations persist: Lipton [22] highlights that model performance often hinges on the alignment of training data with real-world conditions, while agents' decision-making transparency remains inadequate, potentially yielding unpredictable outcomes in unfamiliar circumstances. This dynamic adaptability supports the DAIP's aim of reducing errors continuously, though enhancing transparency remains a key focus.

The DPP, mandated by the EU's Ecodesign for Sustainable Products Regulation (ESPR), provides a structured framework to enhance data transparency and traceability across a product's lifecycle [23]. DPPs are applicable in all sectors such as textile, battery, construction and food. In manufacturing, Psarommatis and May [24] integrate DPPs with advanced manufacturing techniques, yielding cost efficiencies and sustainability benefits by linking data analytics to lifecycle data, while research conducted for healthcare devices, Stodt et al. [25] suggest that utilizing artificial intelligence with DPPs could further advance DPP capabilities through predictive analytics and real-time monitoring. Beyond this domain, DPPs hold potential for tracking AI model performance in applications such as healthcare or finance, offering a verifiable identity that records updates, errors, and decisions. Ompusunggu et al. [26]incorporate causal AI within DPPs to compute Key Performance Indicators (KPIs) across supply chains, highlighting agents' capacity for dynamic performance monitoring, while Samatas et al. [27] demonstrate AI-IoT pairings for predictive maintenance in smart manufacturing. However, integrating AI with DPPs presents challenges, namely: processing voluminous datasets and their heterogeneity require advanced techniques, while opacity persists, prompting Alexander et al. [28] to advocate XAI for transparent decision-making. This integration informs the DAIP's design, treating AI models as traceable assets with lifecycle accountability.

Literature underscores a persistent limitation: AI's black-box nature obscures error causation, and current frameworks lack a holistic mechanism to enforce zero-error performance across domains [15]. XAI improves interpretability [29], and adversarial training bolsters resilience [20], yet neither ensures comprehensive error elimination. AI agents offer adaptability but falter in unfamiliar circumstances due to transparency and robustness deficits [24]. DPPs provide a robust data infrastructure, though their synergy with AI remains underexplored beyond manufacturing contexts [28]. Studies note that XAI approaches, including inherent transparency and post-hoc explanation methods, aim to clarify model behavior, yet measurable interpretability gaps persist [15],[30]. The proposed Digital AI Passport (DAIP) framework addresses these shortcomings by integrating AI agents with DPPs, leveraging a hybrid cloud DPP platform for traceability and at the same time assuring scalability, federated learning for adaptability, and zero-trust security for risk mitigation, establishing a transparent, fail-proof paradigm for critical applications.

3. Methodology

The methodology is organized into four components: (1) Zero-Error AI Agent architecture; (2) Definition of triggering factors and working conditions; (3) Integration with DAIP for lifecycle monitoring; and (4) Security, scalability, and sustainability considerations. The Zero-Error AI Agent operates continuously, detecting and predicting performance degradation using techniques such as concept drift detection (e.g., ADWIN, DDM) and meta-learning. In the context of the DAIP framework, working conditions are defined as any operational changes—occurring in either the physical system or the digital environment—that have the potential to affect an AI model's accuracy, reliability, or stability. In the physical domain, this may include variations in sensor performance, equipment wear, environmental factors such as temperature or vibration, and changes in production processes. In the digital domain, working conditions encompass factors such as data distribution shifts, variations in input data quality, adversarial inputs, evolving user behavior patterns, and software or infrastructure updates. The Zero-Error AI Agent continuously monitors these conditions and uses predictive analytics, including concept drift detection and anomaly detection algorithms, to anticipate potential performance losses before they occur. All detected changes, corresponding model adaptations, and system interventions are recorded immutably within the DAIP to ensure full auditability and lifecycle traceability.The Zero Error AI Agent is a component designed to ensure the continuous, reliable performance of AI systems, even under changing operational conditions. The Figure 1 illustrates the operational workflow of the Zero Error AI Agent, detailing its processes for detecting, predicting, and mitigating potential performance losses.

This multi-step process integrates predictive and corrective mechanisms to maintain near-zero error rates, thereby enhancing the robustness and reliability of AI models. As shown in Figure 1, there are four Triggering factors and two main actions to achieve a (near) Zero Error AI model. The related factors and mechanisms are:

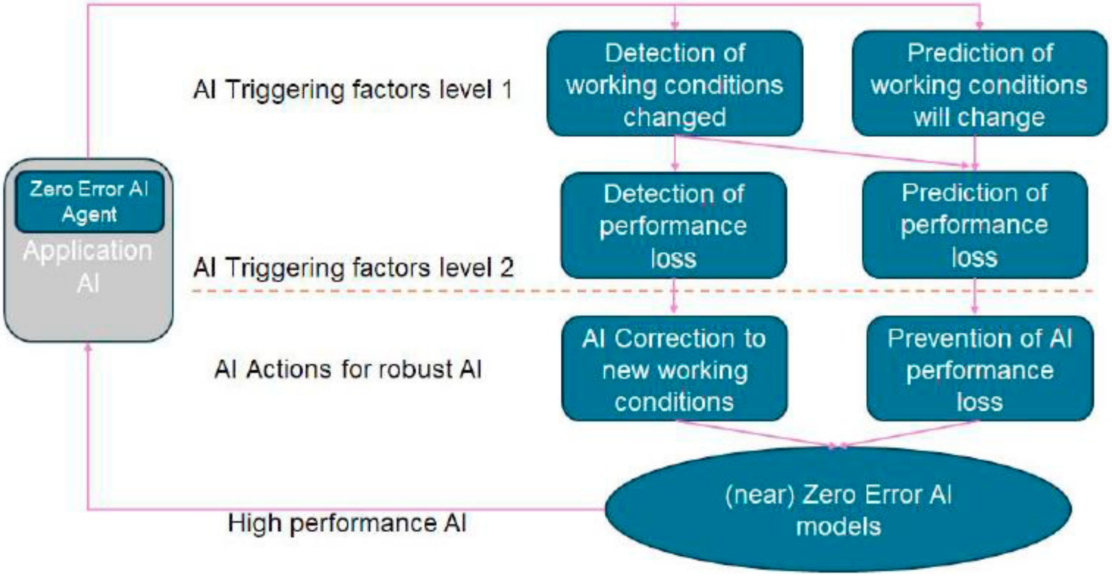

Figure 1 Zero Error AI

- Monitoring and Detection of Changes in Working Conditions
- Prediction of Future Working Condition Changes
- Performance Assessment and Loss Detection
- Corrective and Preventive Actions.

In the proposed framework, the Zero-Error AI Agent operates through a two-level triggering mechanism. At the first level, the agent either detects changes in the working conditions—such as shifts in data distribution, environmental variations, or system configurations—or predicts potential changes before they occur. This information serves as the input for the second level, which focuses on the AI model's performance.

If performance loss is detected, the agent immediately deploys corrective actions to adapt the AI model to the new conditions, while also introducing preventive measures to reduce the likelihood of similar issues in the future. Conversely, if performance loss is predicted, the agent applies proactive measures to prevent the degradation altogether. By combining detection and prediction across both operational conditions and performance outcomes, the framework ensures timely interventions and continuous reliability, enabling near zero-error AI operation even in dynamic environments.

Ultimately, the success of a zero-error AI agent lies in its ability to learn from past mistakes, refine its decision-making processes, and continuously evolve with new data. Monitoring and Detection of Changes in Working Conditions: The Zero Error AI Agent continuously monitors the AI system's operational environment, focusing on detecting any changes in working conditions. These changes may include variations in input data quality, shifts in external environmental factors, or alterations in system load and usage patterns. When the Zero Error, AI Agent identifies that the working conditions have altered, it initiates a sequence of evaluations to assess the impact of these changes on the AI system's performance. This early detection is crucial for pre-emptively addressing any factors that could degrade AI performance.

Prediction of Future Working Condition Changes: Beyond real-time monitoring, the Zero Error AI Agent incorporates predictive algorithms that forecast potential changes in working conditions. This forward-looking capability allows the agent to anticipate challenges before they manifest. By analyzing trends and patterns in the operational data, the agent can predict upcoming changes in the environment that might affect the AI system's effectiveness. This predictive insight enables proactive adjustments, minimizing the likelihood of performance degradation.

Performance Assessment and Loss Detection: Alongside the detection and prediction of changes in working conditions, the Zero Error AI Agent evaluates the impact on the AI system's performance. The agent assesses whether the changes have led or could lead to a loss in AI performance, such as reduced accuracy, slower processing times, or incorrect outputs. This detection is a critical step for ensuring that the AI system continues to meet its operational benchmarks. Similar to its approach with working conditions, the agent predicts potential performance losses based on upcoming changes. This proactive measure is designed to prevent issues before they affect the system's output.

Corrective and Preventive Actions: Upon detecting or predicting a performance loss, the Zero Error AI Agent initiates corrective actions to maintain the system's effectiveness. When performance loss is detected due to changing conditions, the agent adjusts the AI model or system parameters to adapt to the new environment. This may involve recalibrating algorithms, modifying data inputs, or optimizing processing workflows to restore optimal performance. If a performance loss is predicted, the agent implements preventive measures to avoid the anticipated issues. This could involve pre-emptive adjustments to the AI model or initiating fallback procedures to maintain system reliability. Through the combination of detection, prediction, and corrective actions, the Zero Error AI Agent ensures that the AI system operates with minimal errors. The end goal is to achieve (near) Zero Error AI Models, which maintain high levels of accuracy, reliability, and robustness even in dynamic and potentially adverse conditions. This capability is vital for critical applications where AI performance is non-negotiable, such as healthcare, finance, and industrial automation.

DPP The results of the above agent will be integrated with a software platform the hosts the DPP approach. DPP is used for data collection and information correlation. Especially the suggested integrated solution of this survey is related to a framework that can track and trace information and especially all the available and useful data that is related to the performance and the efficiency of the AI models based on a DPP platform (Figure 3). DPP is considered as a technology that is introduced from the necessity for assets monitoring and the provision of a transparent, reliable and holistic solution that covers the entire lifecycle of each asset, from production to the end of life. The key services of the DPP are the following:

- Unique Digital Identity
- Lifecycle Data Storage
- Interoperability
- Compliance & Sustainability

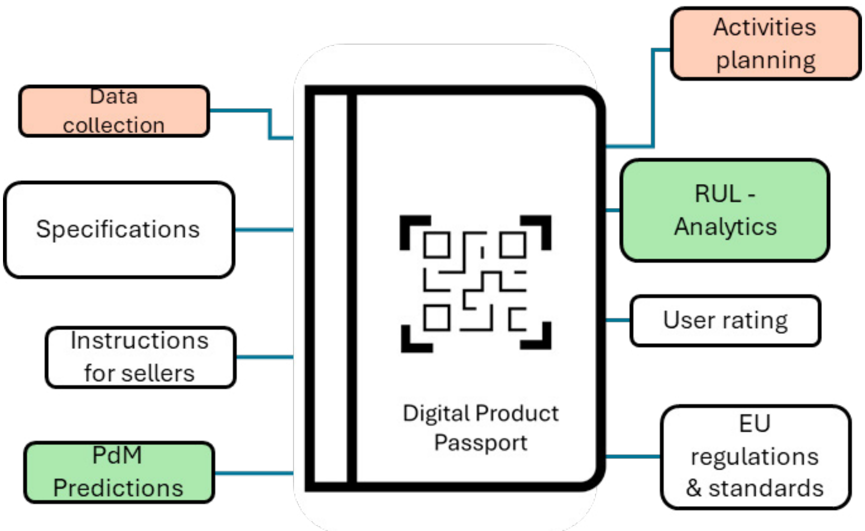

Figure 2 DPP concept

The aim of DPP (Figure 2) technology is tracking and tracing operational technical and any other information regarding assets or activities. Assets can be a product, an activity, a software component, and even an entire production line or factory. In the current case AI models and their results are considered as the products/assets that will be

analyzed by the DPP solution. According to the Eco-design Regulation of the EU, the metrics that are currently defined and included in the Digital AI passport are the following:
Accuracy: The percentage of correct predictions (useful for balanced datasets).
Precision: The proportion of true positive predictions out of all predicted positives (important for cases like fraud detection).
Recall (**Sensitivity**): The proportion of true positive predictions out of all actual positives (crucial for medical diagnosis).
F1-score: The harmonic means of precision and recall (used when there's an imbalance in classes).
ROC-AUC (Receiver Operating Characteristic - Area Under Curve): Measures the model's ability to distinguish between classes.
Inference Time (Latency): The time taken to make a single prediction.
Throughput: The number of predictions the model can make per second.
Memory Usage: The amount of RAM required to store and run the model.
Computer Requirements: The processing power needed, often measured in FLOPs (Floating Point Operations per Second).

The aim of the suggested solution is to provide to the user information graphs and an initial analysis related to the AI model and its performance. The result and the interpreted information will be provided to the Digital Artificial Intelligence Passport (DAIP) by the Zero Error AI agents.

Furthermore, in each AI passport information regarding the training data as well as the engagement information as well as literature references especially regarding the SoTa methodologies will be also included. Especially information related to the preprocessing steps, the applied filters as well as the completed dataset will be referred to the AI passport. The suggested solution focuses on the clustering and the digital record of all the information regarding the AI models. The AI passport can also be accessible to users via QR code scanning. The Digital AI Passport platform provides the result via a simple scan to the users. For example, the products in which the AI is applied have a QR code which displays all the information. The QR code has a dynamic behavior since it is able to present different sets of information according to the user's rights and his/her profile. For example, the developer and the manager scans the same QR code and different information are displayed to each profile.

The continuous monitoring of all technical operations is enabled through dedicated communication protocols and modular connectors that interface directly with the AI models and associated systems. The DAIP platform facilitates structured data exchange primarily via JSON-based payloads, ensuring interoperability across heterogeneous environments. Current supported protocols include REST API, WebSocket, and MQTT, providing both synchronous and asynchronous communication capabilities to accommodate real-time monitoring needs. To enable seamless integration, AI solutions must incorporate these communication services within their source code, allowing automated transmission of performance metrics, lifecycle events, and operational context to the DAIP. This ensures that all recorded information remains complete, consistent, and traceable, maintaining data integrity and preserving context throughout the AI model's lifecycle.

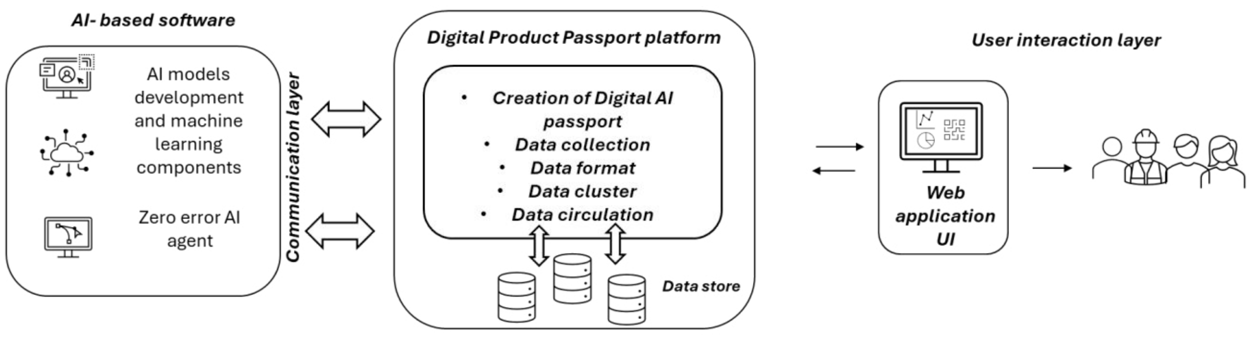

Figure 3 Architecture

4. Implementation

The implementation of the integrated solution is considered as two separate software components which will be correlated to provide an integrated web application for tracking and tracing the AI models and the zero error AI agents. The application follows the client-server architecture including several services regarding communication protocols and data sharing activities. However, before the actual development or configuration of the platform a specific set of activities is performed (Figure 4).Initially, the analysis of each use case is performed, the end users as well as the engineers perform an analysis of the use case and the related steps of the process. The used ISOs and standards are analyzed, and the required information are collected. After the requirements collection the engineers and the end users define the analyzed steps which will be digitalized. The digitalization consists of the modelling of the process including the data that are collected as well as the changes and the updates that will be observed during the engagement of each asset with each step. After the definition of what will be collected, and from which step the connection of the platform with the existing systems that are responsible for the information collection is performed. After the integration and before the connection of the information with the related passports an initial preprocessing and filtering of the information is performed. Especially the connectors of the platform provide some filtering and noise cancellation services in order to increase the quality of the information. After the passports generation and the connection with the data collection systems a continuous recording and update of information starts and each change, update and intervention are recorded in specific digital records based on TSA certification technology.

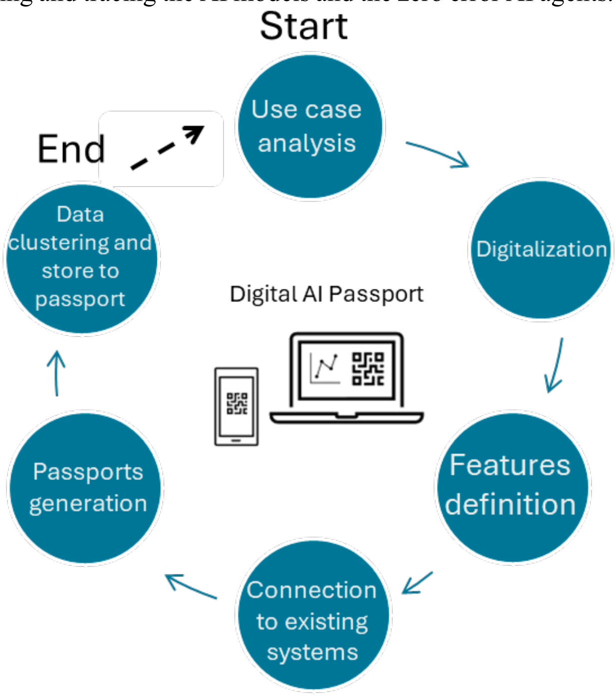

Figure 4 DAIP implementation steps

While DAIP introduces continuous monitoring, its design aims to optimize AI lifecycle operations to reduce unnecessary retraining, minimize carbon footprint, and extend model lifespan. By leveraging targeted updates and efficient inference scheduling, DAIP aligns with the DPP's circular economy principles, balancing reliability with environmental responsibility.

5. Conclusion

Nowadays, the high availability of data as well as structured and not information enabled the adoption of AI applications in several sectors. These significant technologies have introduced several new advanced techniques. In conclusion the aim of this study is to present an integrated solution tha focusing on error minimization and monitoring of the performance of the AI driven advanced technologies. The development of a zero-error AI agent requires a comprehensive approach that integrates high-quality data, robust model optimization, and real-time self-correction mechanisms. By leveraging advanced machine learning techniques, uncertainty estimation, and continuous monitoring, the AI system can significantly minimize errors and improve reliability. Although achieving absolute zero error is theoretically unattainable due to inherent uncertainties in data and real-world environments, implementing confidence thresholds, human-in-the-loop validation, and adaptive learning pipelines ensures that the agent makes highly accurate and responsible decisions.

The aim of this study is to present a concept framework which focuses on the AI modules its performance as well as how the continuous training and the AI itself can contribute regarding the minimization of errors and the optimization of the AI performance. The suggested framework provides structured and asset clustered information and a set of information including results from analysis and any other services. The detailed monitoring of transactions and the digital record of any update or intervention along with easy access increases transparency and reliability among the entire value chain. Thus, all the related people from the value chain such as developers, managers, engineers, testers, simple users and generally the communities can adopt the AI activities and not consider it as a black box. The DAIP provides detailed insight into information for the AI models including evaluation results that performed by Zero Error agents. The error minimization as well as the continuous monitoring and the précised insight into the responsible factors will have as a result the optimized design of the product or the service that is analyzed. The use of the suggested integrate solution provides also evaluated filtered and tested sets of datasets for AI models training and in conjunction with the structured clustered and edge processing information increases the quality of the train used data as well as increase the efficiency of the reinforcement learning techniques.

It should be noted that the present work does not assess the direct sustainability footprint of the AI methods themselves, including their computational and energy demands. Instead, the focus is on how the proposed DAIP framework can enable AI systems to contribute to process-level sustainability by improving operational efficiency, reducing waste, and enhancing decision-making accuracy in high-stakes domains. Evaluating and optimizing the energy efficiency of the AI components within the DAIP will be an important avenue for future research.

6. Next Steps

Future research should focus on enhancing the adaptability and explainability of zero-error AI agents. One key direction is the integration of Explainable AI (XAI) techniques, enabling models to provide transparent justifications for their decisions—an essential capability in critical sectors such as medicine and law. Research further emphasizes the need for Responsible AI, incorporating fairness, accountability, and privacy as core principles alongside explainability, offering a roadmap for refining the DAIP to meet ethical and regulatory standards [4]. This includes developing effective methods for measuring explainability, an area that remains underexplored, to ensure DAIP's transparency is both practical and verifiable.

In addition, future work will involve connecting the DAIP platform more tightly with AI agents capable of supporting users and dynamically optimizing the passport's structure, as well as refining the number and nature of features included, in alignment with EU regulations and ISO standards. Secure integration of diverse data sources, as explored in recent studies, will also be pursued to strengthen DAIP's applicability in sectors such as healthcare and finance, while adhering to zero-trust security principles [4].

Crucially, practical implementations and pilot studies will be conducted in future research to validate the proposed conceptual framework. These real-world deployments will test the DAIP's performance, scalability, and sustainability in operational settings, providing empirical evidence to confirm its feasibility and refine its design for large-scale adoption. Although the DAIP framework is presented here as a conceptual model, future work will involve implementing prototypes in high-stakes domains and comparing them directly with existing AI governance tools. These empirical studies will validate its technical feasibility, scalability, and measurable advantages over current approaches, ensuring that the novelty extends beyond conceptual design into demonstrable impact.

Acknowledgements

The presented work was partially supported by the projects CIRCMAN5.0 and SM4RTENANCE, EU H2020 projects under grant agreements No 101178331and 101123490 accordingly. The paper reflects the authors' views and the Commission is not responsible for any use that may be made of the information it contains. Also this research has been partially financed by the European Union : Next Generation EU through the Program Greece 2.0 National Recovery and Resilience Plan , under the call: SUB1.2: Visiting Professors-Visiting Researchers, Project name: Advanced AI methods for optimizing manufacturing processes and achieving Industry 4.0, " Project code: ΥΠ2ΤΑ-0561842.

References

[1] Chui M, Manyika J, Miremadi M, Henke N, Chung R, Nel P, Malhotra S. Notes from the AI frontier: insights from hundreds of use cases [Internet]. McKinsey & Company; 2018 Apr. Available from: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20AI%20frontier%20Applications%20and%20value%20of%20deep%20learning/Notes-from-the-AI-frontier-Insights-from-hundreds-of-use-cases-Discussion-paper.pdf

[2] Buttazzo G. Rise of artificial general intelligence: risks and opportunities. Front Artif Intell. 2023 Aug 1;6(1). Available from: https://doaj.org/article/f8fe476089b14eca8a4aff3c096aec61

[3] Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2020 Jan 13;11(1):233. Available from: https://doi.org/10.1038/s41467-019-14108-y

[4] Nishant R, Kennedy M, Corbett J. Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. Int J Inf Manage. 2020 Oct;53:102104. Available from: https://doi.org/10.1016/j.ijinfomgt.2020.102104

[5] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: methods and applications. J Manuf Syst. 2018 Jul;48:144-156. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0278612518300037

[6] González Ordiano JÁ, Waczowicz S, Hagenmeyer V, Mikut R. Energy forecasting tools and services. WIREs Data Mining and Knowledge Discovery. 2017 Dec 11;8(2).

[7] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. Sensors. 2018 Aug 14;18(8):2674. Available from: https://doi.org/10.3390/s18082674

[8] Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya A, Almohareb SN, et al. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. BMC Medical Education [Internet]. 2023;23(1). Available from: https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-023-04698-z

[9] Palomares I, Martínez-Cámara E, Montes R, García-Moral P, Chiachio M, Chiachio J, et al. A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. Appl Intell. 2021 Jun 11;51.

[10] Psarommatis F, May G, Dreyfus PA, Kiritsis D. Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. Int J Prod Res. 2020 Apr 19;58(1):1-17.

[11] Psarommatis F. A generic methodology and a digital twin for zero defect manufacturing (ZDM) performance mapping towards design for ZDM. Journal of Manufacturing Systems. 2021 Apr;59:507–21.

[12] European Commission. Industry 4.0: Digitalisation for productivity and growth [Internet]. European Commission; 2015. Available from: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2015)568337

[13] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020 Jun;58(1):82-115. Available from: https://doi.org/10.1016/j.inffus.2019.12.012

[14] Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature Machine Intelligence [Internet]. 2019 May;1(5):206–15. Available from: https://www.nature.com/articles/s42256-019-0048-x

[15] Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923. 2017.

[16] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proc Advances in Neural Information Processing Systems (NeurIPS); 2017. p. 4765-4774.

[17] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proc Advances in Neural Information Processing Systems (NeurIPS); 2017. p. 5998-6008.

[18] Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: Proc International Conference on Learning Representations (ICLR); 2015.

[19] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson; 2021.

[20] Sutton RS, Barto AG. Reinforcement Learning: An Introduction. 2nd ed. MIT Press; 2018.

[21] Yu C, Liu J, Nemati S. Reinforcement learning in healthcare: A survey. ACM Comput Surv. 2019;55(1):1-36.

[22] Lipton ZC. The mythos of model interpretability. Commun ACM. 2018;61(10):36-43.

[23] European Commission. Ecodesign for Sustainable Products Regulation [Internet]. European Commission; 2023. Available from: https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/ecodesign-sustainable-products-regulation_en

[24] Psarommatis F, May G. The role and benefits of Digital Product Passport in the remanufacturing process in the era of Zero Defect Re-manufacturing. Procedia CIRP. 2024 Nov 27;130:695-700. Available from: https://www.sciencedirect.com/science/article/pii/S2212827124013076

[25] Fatemeh Stodt, Ruf P, Reich C. Blockchain-Enabled Digital Product Passports for Enhancing Security and Lifecycle Management in Healthcare Devices. 2024 Dec 4;44–51.

[26] Ompusunggu AP, Tjahjowidodo T, Wicaksono H. Causal AI-powered digital product passports for enabling a circular and sustainable manufacturing ecosystem. In: Proc 12th International Conference on Through-life Engineering Services – TESConf2024; 2024 Jun 6-7; Cranfield, UK.

[27] Samatas GG, Moumgiakmas SS, Papakostas GA. Predictive maintenance - bridging artificial intelligence and IoT. In: Proc IEEE Xplore; 2021. p. 413-419. Available from: https://ieeexplore.ieee.org/abstract/document/9454173

[28] Alexander Z, Chau DH, Saldaña C. An interrogative survey of explainable AI in manufacturing. IEEE Trans Ind Inform. 2024 Jan 1;1-13.

[29] Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ. XAI—Explainable artificial intelligence. Sci Robot. 2019;4(37).

[30] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.