Quantifying and qualifying the human-in-the-loop task in the Digital Product Passport production process

Anonymous ACL submission

Abstract

The *Digital Product Passport 4.0* (DPP 4.0) initiative drives the transition from linear to circular economies. From 2027, it will enable seamless data exchange between manufacturers, suppliers and end-of-life actors using Asset Administration Shell-compliant *digital twins* (DTs). However, creating and validating these DTs is tedious, error-prone, and potentially overwhelming even for experts due to the complexity of formal descriptions and data. We turn this disadvantage into an advantage by leveraging the hierarchical DT structures as an *interlingua* to communicate with a *Large Language Model* (LLM). Through automatic tree navigation, prompts are generated for substructures, which an LLM is asked to correct/revise. Clear cases run automatically, while a *human-in-the-loop* (HiTL) is involved for necessary decisions through specific dialogues. We quantify and qualify the HiTL-tasks and, most importantly, show how to automatically obtain the appropriate context for the *user-centered decision-making dialogues*. A prototype implementing key aspects is currently under evaluation.

1 Introduction

For the transition from linear to sustainable circular economy it is essential to provide detailed, standardized information throughout a product's lifecycle to connect producers, material suppliers, and end-of-life actors, i.e., recyclers (see, e.g., ZVEI, 2023, Plociennik et al., 2024, or Kühn et al. 2025). The *Digital Product Passport 4.0* (DPP 4.0[1]) is an initiative to document key information throughout a product's lifecycle. It employs *digital twins* (DTs) accessible via a globally unique ID link (cf. IEC 61406) to ensure seamless data exchange across the value chain. The DPP 4.0 initiative recommends modelling the DTs in the *Industry 4.0* standard, the *Asset Administration Shell*[2] (AAS), to facilitate efficient information exchange. However, successfully creating the AAS is a challenging and labor-intensive process (Shi et al., 2025) particularly for non-expert users in *small and medium sized enterprises* (SMEs; cf. the CIRPASS study, 2024; Mohandes et al. (2024) and Neligan et al. (2023) point out significant barriers to AAS adoption and DPP realization, respectively).

Prompted by the increasing use of AI to create AAS, this paper seeks to answer the research question: How can an AI-based AAS review, correction, and extension process be designed to be user-friendly and effective for non-experts irrespective of its creation method (manual/rule-/AI-based)?

As machine-learning systems advance, *human-in-the-loop* (HiTL) methods are increasingly essential (see, e.g., Wu et al. (2022) for a survey on training accurate prediction models with minimum cost by integrating human expertise). In computational linguistics, human-in-the-loop approaches have been widely explored in *Topic Modeling* (HLTM) where user feedback refines models through operations, such as adding words to topics, merging topics, or removing documents (cf. Kumar et al. (2019) for an evaluation). Generally speaking, human-in-the-loop systems are key for *Industry 5.0*, which is often characterized as *human-centric* in the literature (cf. Xun et al., 2021).

[1] The DPP 4.0 (https://dpp40.eu/) is part of the EU's *Ecodesign for Sustainable Products Regulation* (ESPR) which entered into force on July 18th of 2024 as part of the *European Green Deal* for the *European Climate Pact* and becomes mandatory 2027 for the first product types, such as batteries, textiles, electronics and plastics.

[2] The Industrial Digital Twin Association (IDTA, https://industrialdigitaltwin.org/dpp4-0) develops the Asset Administration Shell as the IEC 63278 standard for digital twins in Industry 4.0 (I4.0, www.plattform-i40.de) use cases.

In this paper, we explore HiTL interventions in the context of AI-based AAS creation for DPPs (see Section 2 on the state of the art). Inspired by ongoing research in *explainable AI* (XAI), we adopt key design principles for *interactive explanation user interfaces* (cf. Chromik & Butz, 2021): "Human-XAI interaction happens in stages or turns, where the correct mapping between UI functions and the user's intentions and feedback by the UI has to be ensured to bridge the *gulf of execution* (Norman et al., 1986).". The key idea is that users weigh up the effort and benefits of an interaction. Thus, HiTL actions must balance guidance and efficiency. Following the principle "Sensitivity to the Mind and Context" by Chromik and Butz (2021), automatically generated queries should neither overtax the user's current state of knowledge with unnecessary complexity nor request for context-free, isolated information (e.g., asking for a missing value without relevant context). Aligning with the "Flexibility through Multiple Ways to Explain" principle in XAI-UI, the system should respect user preferences in task execution. For AI-based AAS validation, we extend these principles by avoiding boring questions (e.g., repetitive confirmation checks). However, correctness and completeness of the AAS must become assessable for the human.

To apply these XAI-UX-principles to AI-assisted AAS creation, we use filled AAS instances[3] as an *interlingua* for LLM communication. This abstraction has two main advantages: (1) substructures in the AAS tree can be corrected with high accuracy by an LLM; (2) LLM-generated refinement requests are directly linked to standardized submodels, ensuring automatic enrichment with sufficient surrounding concepts from the overall AAS structure in queries to the human.

For (1), a support system must provide a tree navigation mechanism to select partial structures and coin them into LLM prompts. The LLM's unambiguous replies are logged as fully automatic changes to the structure. For (2), the structure-inspection procedure selects siblings as values and/or one super-structure level and generates precisely formulated intervention requests for the human-in-the-loop.

The paper is organized as follows. First, we outline the state of the art in manual and automatic DT creation focusing on AAS representation for DPPs. Section 3 details on the intervention dialogues to ensure correctness and completeness. In Section 4, we present our prototype's interface illustrating an informed request to the user. In Section 5, we draw conclusions and address future work.

2 State of the art in DT/AAS creation

The precise definition of mandatory elements within a DPP is still to be established (ZVEI, 2023). In AAS, information is modeled in standardized *submodel templates* (SMTs) for a broad range of use cases of varying complexity from *Handover Documentation* to *Plant Asset Management* with constraints and hierarchical dependencies. Currently, 93 SMTs are registered[4]. Relevant SMT for DPPs in the context of circular economy are, for example, *Digital Nameplate*, *Carbon Footprint*, and *Technical Data*.

Commonly, AAS are created using tools like the *Eclipse AASX Package Explorer*[5] or APIs (e.g., *FA³ST*[6]). In turn, standard compliance can be tested with the *IDTA AAS Test Engine*[7]. However, these tools are primarily designed for AAS engineers or software developers and require intricate expert knowledge on the tools and the AAS standard.

Automating the AAS creation process can be based on rule-based methods which are primarily designed to transform a specific data model (e.g., *UML*) into AAS format. For a detailed survey of rule-based approaches, see Xia et al. (2024). However, all approaches are limited by a lack of semantic understanding of data and rigidity (Xia et al., 2024). This can be remedied by NLP-based approaches that provide automatic mapping of raw data to the ECLASS[8] dictionary as a foundation for creating the structured representation that is required in AAS (cf. the survey by Xia et al., 2024). The approaches by Both et al. (2021), Cartus et al.

[3] For official DPP 4.0 product carbon footprint showcases using AAS instances, see ZVEI (2023) or https://pcf.dpp40-2-v2.industrialdigitaltwin.org/pcf2.

[4] For the official list of IDTA Submodel Templates, see https://industrialdigitaltwin.org/en/content-hub/submodels

[5] Eclipse AASX Package Explorer: https://github.com/eclipse-aaspe/package-explorer

[6] FA³ST – Fraunhofer Advanced Asset Administration Tools for Digital Twins by the Fraunhofer IOSB: https://www.iosb.fraunhofer.de/en/projects-and-products/faaast-tools-digital-twins-asset-administration-shell-industrie40.html

[7] The current IDTA Test Engine (https://github.com/admin-shell-io/aas-test-engines) V1.1 supports rule-based testing against the meta model of the AAS specification and the SMT *Digital Nameplate*. Other SMTs are not supported.

[8] ECLASS is an ISO/IEC compliant data standard for products & services: https://eclass.eu/

(2022), and Beernann et al. (2023) use trained embedding *BERT*-variants to map technical properties to ECLASS definitions. Xia et al. (2022) employ neural language models to connect manufacturer's data properties with the ECLASS definitions using semantic fingerprints. Shi et al. (2024) use pre-trained LLMs for semantic search for ECLASS entity matching. Shi et al. (2025) evaluate open-source LLMs and *GPT-4* for ECLASS entity matching. A first approach to fully automate AAS creation, is presented by Xia et al. (2024), who evaluate LLM configurations for *Retrieval-Augmented-Generation* (RAG) based on the ECLASS dictionary to transform a snippet of raw textual data entered by the user into a text field into an AAS with the SMT *Technical Data* for further editing or validation in proprietary tools. Heitkamp and Bendorf (2025) use image processing to extract attributes from photos of type shields and, in turn, use LLMs to create AAS with SMT *Digital Nameplate*. The *GPT-4* based *AAS Generator*[9] can create the SMT *Digital Nameplate* or *Technical Data* and map data to ECLASS semantic ids from user input, such as documents, URLs or text.

In respect to HiTL interactions, all state-of-the art approaches in AAS creation rely on the user solely as the initial data provider. The integration of human-AI interactions into the design and creation of DTs remains an open question (cf. Kreuzer et al. (2024) who reviewed AI components in DT design over the years 2018-2022 and found that only a small subset of papers integrated humans as key components of AI-DT systems, highlighting the lack of definitions for roles humans could play in these systems). This aligns with Fernandes et al. (2025) who emphasize the need to consider human actions and intentions in HiTL system design for more effective integration of human expertise. According to Kreuzer et al. (2024), the direction DT-to-physical-system is mainly absent from considerations. Here, HiTL is uncharted territory.

3 Human-in-the-loop in AAS validation

In automating AAS creation, human validation remains crucial (cf. Table 1 for a set of potential error types to be reviewed by the human). Non-experts, in particular, may overly trust in the validity of LLM-outputs (cf. Mittelstadt et al., 2023 and Papenkordt, 2024). Moreover, expertise absent in textual input data has to be integrated by the users. The complexity and volume of structured data in AAS require strategic support for intuitive and efficient user interactions. For instance, the system should offer guidance on the order of error inspection, i.e., prioritising serious errors or missing information, and automize standard conformance and correctness checks to a large extent. Nevertheless, the interface should leave the final navigation strategy to the user (cf. Figure 1 for an overview of the three error categories in a simple color-coding in our prototype for the user to select).

Category	Description
Standard Conformity	(a) Template structure, and (b) data entities adhere to the AAS standard
Completeness	Mandatory elements and information are present
Correctness	Data is factually accurate and valid
Plausibility	Reasonableness and logical consistency of the data
Consistency	Consistent terminology and formatting throughout the AAS
Relevance	Content is relevant to the DPP

Table 1: Prototypical evaluation tool error categories to be resolved in user-friendly dialogues.

As mentioned in Section 2, AAS standard conformity can be checked by the IDTA Test Engine, which supports rule-based file validation. However, it can only validate elements that are explicitly present within the AAS structure, but it cannot detect missing mandatory structures. Moreover, the often long lists of errors have to be inspected one-by-one without concrete correction suggestions[10]. For standard conformity checking, we leverage the structured formality of AAS as an *interlingua* to prompt any LLM[11] for validation tasks—similar to AI-assisted programming. Essentially, we automatically break down the overall structure into appropriate substructures in a bottom-up manner to be "discussed" with an LLM. An appropriate substructure is characterized by just sufficient context determined by the meta-knowledge from the AAS specification (e.g., regulations or EU standards)

[9] https://chatgpt.com/g/g-BcWz9NSxA-aas-generator

[10] See validation-service.admin-shell-io.com for examples of validation output for AAS by the IDTA Test Engine.

[11] For our general discussions here, we probe *ChatGPT 4o Mini*. Fine-tuned LLMs would provide even better answers. In our prototype, we offer finetuned open-source LLMs, such as the EU-Data AI Act compliant *OpenGPTX Teuken 7B* (https://opengpt-x.de/en/models/teuken-7b), to create AAS from unstructured, combined textual data (e.g., sourced from URLs, technical sheets, or data tables).

and corresponding regions of connected ontologies (e.g., ECLASS) underpinned by smaller portions of additional files, such as technical product sheets or even (AI-generated) knowledge graphs. In an extensive series of cases from Table 1, LLMs offer a comprehensive and intelligent validation with high accuracy and hardly any hallucination for these small AAS portions with specific context.

Not only the prompt generation can be automized in our interlingua approach. In general, we classify replies by the LLM into two main categories: (1) Errors that can be automatically corrected, especially syntax violations (e.g., typos in template words), local structure mismatches (e.g., misplaced brackets or empty arrays) or pattern violations (e.g., predefined formatting inconsistencies). These AAS changes are performed automatically with logging for inspections. If not requested otherwise, the user is not bothered. (2) Errors requiring human input—the HiTL cases—are enriched with context to allow informed user decisions.

Here, we distinguish two levels to focus the user's attention: (1) Warnings (yellow) refer to minor cases, like duplicate elements or plausibility checks (e.g., chronology of maintenance date vs. installation date), or confirm unit consistency (e.g., seconds vs. milliseconds). The user sees both elements and a glimpse of the environment with the request to decide which one to retain or remove/revise. For multi-language descriptions, consistency checks ensure that translations convey the same meaning across various EU languages. (2) For critical issues (i.e., violations to standard conformance; red), the system generates context-based suggestions, but expert validation is required before integration (cf. Figure 1). All checks for correctness can include the comparison of AAS data with data in additional documents, e.g., tech sheets.

In the direction DT-to-physical-system, our approach can offer AI-based recommendations for actions in natural language, such as maintenance intervals, derived from real-time operational data captured by the AAS (cf. Cavalieri and Salafia (2020) for predictive maintenance with AAS). These recommendations can be refined or overridden by the human based on operational experience, contributing to system learning.

4 LLM-based AAS-Validation Interface

Figure 1 shows the summary page of our prototype after it has performed all automatic evaluations with an LLM. The user can select a category by clicking. By simple color-coding, we offer all automatic changes in green (simple errors, solved automatically by LLM-interactions), warnings in yellow and red for issues that require mandatory human assistance. The overall number of items is indicated in brackets. On demand, a list is unfolded. Each list is sorted by the error's depth in the overall structure (higher levels first) for searching the review efficiently. In Figure 1, the user focuses on a critical element: The system communicates in a precise, understandable message—varying between a shorter tabular/elaborate text according to the users' preferences—next to the error path that it detected a missing mandatory *idShort* element. Its depth level in the AAS hierarchy can be visualized (cf. the tree icon next to the path). To suggest a value, the system analyzes the context of the missing attribute by considering its definition and the technical data (cf. the list concrete values retrieved from the AAS). The user can accept or elicit the generation of a new suggestion, or enter a new value. In turn, the system directly updates the value in the overall AAS (cf. the blue arrow).

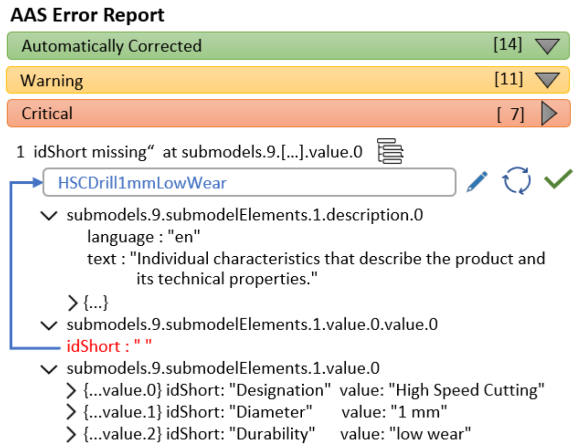

Figure 1: A dialogue with our prototype for AI-aided AAS reviewing using the tabular mode.

The features of our prototype are based on a pre-development survey with 12 AAS experts highlighting a strong desire for intelligent assistance in AAS validation. Currently, we evaluate with non-experts from SMEs.

5 Conclusions

We presented the general concept of our user-centered AI-based AAS validation tool to become part of our AAS creation system. According to the next round of feedback, we rework/expand the dialogues. In general, we aim at a data protection-friendly approach for local machinery by SMEs.

Limitations

As clearly stated, our approach is prototypical, and we acknowledge certain limitations in terms of covered scope and effectiveness:

(1) Our prototypical results were primarily obtained using ChatGPT 4o Mini, which provides valuable insights but may not perform as well as fine-tuned LLMs. To improve our system, we need to use fine-tuned or trained LLMs, specifically relying on open-source models that are compliant with the EU Data AI Act, such as *OpenGPTX Teuken 7B* and/or can run on local/private machines to ensure that critical corporate data remains protected.

(2) The personalized modes to communicate with the human, is currently restricted to the tabular mode and its verbalization by an LLM. Additional modes will be explored in upcoming tests with end users of different target groups (e.g., non AAS specialists from SMEs).

(3) Due to the fact that DTs often contain private product or company data, our tests were limited to non-critical, generalized data and data obtained from public sources, such as company websites, to minimize privacy risks.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used *DeepL Translator* and *DeepL Write* in order to improve language and readability. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Jo Beermann, Rebekka Benfer, Maximilian Both, Jochen Müller and Christian Diedrich. 2023. Comparison of Different Natural Language Processing Models to Achieve Semantic Interoperability of Heterogeneous Asset Administration Shells. *IEEE 21st International Conference on Industrial Informatics (INDIN)*, Lemgo, Germany. https://doi.org/10.1109/INDIN51400.2023.10218154.

Maximilian Both, Jochen Müller and Christian Diedrich. 2021. Automatisierte Abbildung semantisch heterogener I4.0-Verwaltungsschalen durch Methoden des Natural Language Processing. In *Automatisierungstechnik*, 69(11): 940-951. https://doi.org/10.1515/auto-2021-0050.

Alina Cartus, Maximilian Both, Nicolai Maisch, Jochen Müller, and Christian Diedrich. 2022. Interoperability of semantically heterogeneous digital twins through Natural Language Processing methods. *CLIMA 2022 Conference*. https://doi.org/10.34641/clima.2022.143.

Salvatore Cavalieri, and Marco Salafia. 2020. A Model for Predictive Maintenance Based on Asset Administration Shell. *Sensors*, 20: 6028. http://dx.doi.org/10.3390/s20216028

Marcel Chromik and Andreas Butz. 2021. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In *Ardito, C., et al. Human-Computer Interaction – INTERACT 2021*. INTERACT 2021. Lecture Notes in Computer Science(),12933. Springer, Cham. https://doi.org/10.1007/978-3-030-85616-8_36.

CIRPASS. 2024. A study on DPP costs and benefits for SMEs. [Online]. Available at CIRPASS-A-study-on-DPP-costs-and-benefits-for-SMEs-v1.0-1.pdf [Accessed 2 February 2025]

José Marcelo Fernandes, Jorge Eduardo Rivadeneira, André Rodrigues, Fernando Boavida and Jorge Sá Silva. 2025. People 4.0 — A model for Human-in-the-Loop CPS-based systems. *Computer Standards & Interfaces*, 91. https://doi.org/10.1016/j.csi.2024.103895.

Dennis Heitkamp and Gesa Bendorf. 2025. AAS Submodel Generator: Inferring Product Data from Natural Language. *Atp magazine,* 1-2 2025.

IEC. 2020. *IEC 61406: Identification link – A unified approach for product identification based on web links*. Geneva: International Electrotechnical Commission.

Maximilian Kühn, Michael Baumann, Friedrich Volz, and Ljiljana Stojanovic. 2025. Digital Product Passport Design Supporting the Circular Economy Based on the Asset Administration Shell. *Sustainability,* 17(3): 969. https://doi.org/10.3390/su17030969.

Varun Kumar, Alison Smith-Renner, Leah Findlater, Kevin Seppi and Jordan Boyd-Graber. 2019. Why Didn't You Listen to Me? Comparing User Control of Human-in-the-Loop Topic Models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 6323–6330.* Florence, Italy. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1637.

Tim Kreuzer, Panagiotis Papapetrou and Jelena Zdravkovic. 2024. Artificial intelligence in digital twins—A systematic literature review. *Data & Knowledge Engineering,* 151. https://doi.org/10.1016/j.datak.2024.102304.

Brent Mittelstadt, Sandra Wachter and Chris Russell. 2023. To protect science, we must use LLMs as zero-shot translators. *Nature Human Behavior*, 7:1830–1832. https://doi.org/10.1038/s41562-023-01744-0.

Saeed Reza Mohandes, Atul Kumar Singh, Abdulwahed Fazeli, Saeed Banihashemi, Mehrdad Arashpour, Clara Cheung, Obuks Ejohwomu, and Tarek Zayed- 2024. Determining the stationary digital twins implementation barriers for sustainable construction projects, Smart and Sustainable Built Environment, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/SASBE-11-2023-0344

Adriana Neligan, Carmen Schleicher, Barbara Engels, and Thorsten Kroke. 2023. Digitaler Produktpass — Enabler der Circular Economy. Relevanz und Umsetzbarkeit durch Unternehmen. *IW-Report*, 47. Berlin / Köln, Germany.

Donald A. Norman and Stephen W. Draper. 1986. *User Centered System Design: New Perspectives on Human-Computer Interaction.* Lawrence Erlbam Associates, Hillsdale, NJ, USA.

Jörg Papenkordt. 2024. Navigating Transparency: The Influence of On-demand Explanations on Non-expert User Interaction with AI. In *Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI*, HCII 2024, Lecture Notes in Computer Science(), vol 14734. Springer, Cham. https://doi.org/10.1007/978-3-031-60606-9_14.

Christiane Plociennik, Ali Nazeri, Mohammad Hossein Rimaz, Svenja Knetsch, Alice Lopes, Tabea Hagedorn, Julian Baehr, Malte Vogelgesang, Chanchan Li, Wladislaw Benner, Bernd Kellerer, Emanuel Ionescu, Martin Ruskowski, and Anke Weidenkaff . 2024. Connecting Producers and Recyclers: A Digital Product Passport Concept and Implementation Suitable for End-of-Life Management. *Procedia CIRP*, 122(3). *http://dx.doi.org/10.1016/j.procir.2024.02.026.*

Dachuan Shi, Philipp Liedl and Thomas Bauernhansl. 2024. Interoperable information modelling leveraging asset administration shell and large language model for quality control toward zero defect manufacturing. *Journal of Manufacturing Systems*, 77:678-696. https://doi.org/10.1016/j.jmsy.2024.10.011.

Dachuan Shi, Olga Meyer, Michael Oberle and Thomas Bauernhansl. 2025.Dual data mapping with fine-tuned large language models and asset administration shells toward interoperable knowledge representation. *Robotics and Computer-Integrated Manufacturing*, 91. 91. https://doi.org/10.1016/j.rcim.2024.102837.

Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems,* 135:364-381. https://doi.org/10.1016/j.future.2022.05.014.

Yuchen Xia, Zhewen Xiao, Nasser Jazdi and Michael Weyrich. 2024. Generation of Asset Administration Shell With Large Language Model Agents: Toward Semantic Interoperability in Digital Twins in the Context of Industry 4.0. In *IEEE Access*, 12: 84863-84877. http://dx.doi.org/10.1109/ACCESS.2024.3415470.

Yuchen Xia, Nasser Jazdi and Michael Weyrich. 2022. Automated generation of Asset Administration Shell: a transfer learning approach with neural language model and semantic fingerprints. *IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*, Stuttgart, Germany. pages 1-4. doi: https://doi.org/10.1109/ETFA52439.2022.9921637.

Xu, Xun, Yuqian Lu, Birgit Vogel-Heuser, and Lihui Wang. 2021. Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*, 61:530-535. https://doi.org/10.1016/j.jmsy.2021.10.006.

ZVEI. 2023. Implementing Level 2 of the ZVEI-Show Case PCF@Control Cabinet. [Online]. Available at: https://www.zvei.org/fileadmin/user_upload/Themen/Industrie/PCF%40ControlCabinet/White-Paper_Implementing-Level-2-of-the-ZVEI-Show-Case.pdf [Accessed 2 February 2025]