Aligning Data Debt with AI-Integrated Software Project Lifecycle Processes: A Standard-Based Mapping Approach

Nilay Yorgancılar Akgül ¹HAVELSAN A.Ş., ²Graduate School of Informatics, Data Informatics Dept., Middle East Technical University Ankara, TURKIYE nyorgancilar@havelsan.com.tr Tuğba Taşkaya Temizel Graduate School of Informatics, Data Informatics Dept., Middle East Technical University Ankara, TURKIYE ttemizel@metu.edu.tr Özden Özcan Top Graduate School of Informatics, Information Systems Dept., Middle East Technical University Ankara, TURKIYE ozdenoz@metu.edu.tr Pelin Dayan Akman Graduate School of Informatics, Information Systems Dept., Middle East Technical University Ankara, TURKIYE pelin.dayan@metu.edu.tr

Abstract—Artificial Intelligence (AI) technologies have become increasingly central to software development, enhancing efficiency with tools such as intelligent code assistants and driving innovations in products like chatbots, recommendation engines, and predictive analytics. Despite these advancements, the inherent complexity of AI-integrated software projects often leads to the accumulation of technical debt (TD), which can compromise the reliability and sustainability of systems in the long term. Managing TD effectively in these projects can be achieved by adapting international standards. Although these standards are not designed for TD management, they can be systematically applied to detect and address TD by aligning with AI system lifecycle processes. The aim of this study is to demonstrate how AI-related TD correlates with various AI lifecycle processes, thereby enabling systematic detection and management of TD in AI-integrated software projects. To achieve this, we studied 73 unique cases of TD, each reflecting either an instance or a root cause of datarelated TD. These cases were subsequently mapped to the processes and activities outlined in the ISO/IEC 5338 AI Systems Lifecycle Processes standard. Subsequently, the accuracy of these mappings was validated bidirectionally by a large language model and two domain experts. Our findings revealed that data-related TD categories are associated with a diverse range of processes such as design definition, quality management and human resource management and tend to accumulate more significantly in certain areas within the AI lifecycle. This study not only serves as a proof of concept for developing a management approach for AI-related TD, but also enhances the body of knowledge on managing TD in AI projects by detailing how TD interacts with and impacts various AI lifecycle processes.

Keywords—Technical Debt, Artificial Intelligence, AI Life Cycle, TD Management, ISO/IEC 5338, ISO/IEC 122207, LLM

I. INTRODUCTION

Artificial intelligence (AI) technologies have recently become integral to many software development (SD) projects, serving two main purposes. First, they can improve software development processes through tools like intelligent code assistants, which streamline coding, debugging, and testing. Second, they form core components of AI-integrated products, such as chatbots, recommendation engines, and predictive analytics tools. The second group is considered within the scope of this study.

As in traditional software development projects, technical debt (TD) also arises in AI-integrated software products. The concept of TD was first introduced in Ward Cunningham's experience report [1], where it is defined as the consequences of short-term decisions made during the SD process, which increasingly hinder software development and maintenance over time. While the concept of TD is quite important in traditional SD projects [2], it requires even greater emphasis in AIintegrated SD projects. This is because AI systems not only encounter the typical TD issues found in SD projects, such as code debt and configuration debt, but also face additional challenges unique to AI [3]. Although there is not yet sufficient research in this area, TD types such as data smells [4, 5, 6], code smells [7] and prompt smells [8] are encountered in AIintegrated SD projects. Wiese et al. [2] state that TD awareness is quite low in AI-integrated SD companies; which tend to manage project issues reactively, intervening only after problems arise. This lack of a preventive approach can lead to costly consequences for companies in the long term. Therefore, in AI-integrated SD projects, TD management (TDM) is critical to ensuring the project quality dimension.

Data Debt, one of the most common types of technical debt in AI-based systems, refers to issues related to the collection, management, processing, and storage of data [9]. When it leads to immediate problems, the issue becomes more apparent, but it can also remain hidden, creating long-term risks for the system. It includes challenges like data dependencies, quality issues such as weak data relevance or data bias, labeling deficiencies, and anomalies in the data [9]. Data plays a critical role in the performance of AI systems, requiring meticulous attention. Gudivada et al. [10] have explored data quality issues in the context of big data and machine learning (ML) projects, addressing topics such as missing/duplicated data, data heterogeneity, semantic data integration, and the bias and variance tradeoff. Foidl et al. [6], in the context of AI-based systems, have investigated "data smells," and categorized them into three main types: Believability Smells, Understandability Smells, and Consistency Smells. Their study examined data smells through the dimensions of causes, consequences, and tools for detection. On the other hand, AI-integrated SD companies often lack sufficient awareness of recently emerging AI standards that can provide valuable guidance throughout the

AI project life cycle. This limited awareness hinders their ability to adopt a structured approach to managing TD effectively. As highlighted in previous studies [11], many practitioners remain unaware of these standards, which could otherwise serve as a foundation for ensuring quality, reducing risks, and improving the sustainability of AI-integrated projects.

Our study proposes a standards-based approach designed to help AI-integrated software development companies adopt a TD perspective. Our approach aims to increase awareness of, recognize, monitor, and prevent the recurrence of problems that arise during the AI-integrated SD projects. It also aims to provide guidance regarding how different types of major AI life cycle processes should be interpreted from a Data Debt point of view. Additionally, it intends to explore the specific practices within these standards that can help mitigate the formation of Data Debt.

In line with these purposes, we focused on the processes outlined in ISO/IEC 5338 - AI Life Cycle Processes [12] and, where ISO/IEC 5338 directs, the related ISO/IEC 12207 Software Life Cycle Processes [13], to identify how neglecting them can lead to the accumulation of Data Debt. We mapped 73 data debt-related cases, along with associated root causes and solutions obtained from interviews with 18 industry practitioners, and were mapped to the processes and activities outlined in the ISO/IEC 5338 and ISO/IEC 12207 standards. In addition to manual mapping of data debt-related cases with standard practices, we examined how large language models (LLMs) can be leveraged to identify processes/practices related to Data Debt, and what prompt templates could be most effective in the verification context.

The structure of the paper continues as follows: Section 2 presents background and related work, providing an overview of existing research on TD and AI-integrated SD. Section 3 describes the research methodology, detailing the approach used to map data debt-related cases to AI lifecycle processes. Section 4 presents the results, followed by the discussion in Section 5. Section 6 provides validity threats and actions taken to mitigate them. Finally, Section 7 presents the conclusions.

II. BACKGROUND AND RELATED WORK

A. Background

AI-integrated SD necessitates following lifecycle models and standards. In the literature, there are relatively few established data science life cycle frameworks. TDSP (Team Data Science Project) comprises five interconnected and iterative main stages. "Business Understanding", "Data Acquisition and Understanding", "Modeling", "Deployment", and "Customer Acceptance". Another well-known framework, CRISP-DM, which is no longer maintained, includes six stages similar to TDSP life cycle. In a recent study [14], Microsoft proposed a workflow that integrates software engineering principles into the development process of AI applications, which includes nine stages, some of which are data-focused while others are model-focused.

Organizations like ISO, IEEE, and the International Telecommunication Union (ITU) have released numerous

standards for AI/ML processes, with efforts continuing to expand. A June 2024 report by the National Institute of Standards and Technology (NIST) [15] cataloged these standards, noting 63 from ISO, seven from IEEE, ten from CEN¹, and 26 from ITU². ISO leads in scope, covering areas such as AI reliability, data lifecycle frameworks, bias mitigation, and AI-assisted decision-making.

In the scope of this study, it was observed that a more detailed breakdown was needed compared to TDSP or CRISP-DM life cycle models to ensure better TD management. These models primarily focus on general data science workflows and lack sufficient granularity to address TD challenges. The recently introduced ISO/IEC 5338 - AI Life Cycle Processes standard was identified as the most suitable model to meet this requirement which defines various processes and activities for AI system life cycle, building on ISO/IEC 15288 and ISO/IEC 12207 standards, with modifications specific to AI processes. ISO/IEC 5338 standard outlines a total of 33 processes, specifying the Purpose, Outputs, Activities, Tasks, and AIspecific considerations for each. The ISO/IEC 5338 standard organizes AI system life cycle processes into four main categories: (1) Agreement processes, (2) Organizational projectenabling processes, (3) Technical management processes, and (4) Technical processes. The technical processes include activities such as design definition, AI data engineering, and implementation, which are directly related to the development and deployment of AI systems. The standard particularly discusses data-related activities under the "6.4.8. AI Data Engineering" key technical process that outlines data-related activities and tasks such as data acquisition, labeling, data quality issues, data preprocessing. It provides recommendations and guidance for executing these activities.

On the other hand, the primary standard in the software engineering domain is ISO/IEC 12207 Software Lifecycle Processes. This standard defines 30 processes for managing software lifecycle activities, including the development, operation, maintenance, and disposal of software systems [13]. While the primary focus of our research is mapping Data Debt types to ISO/IEC 5338 processes, ISO/IEC 5338 directs readers to ISO/IEC 12207 for processes that remain unchanged for AI systems, instead of redundantly defining them. Therefore, we have also reviewed and mapped the processes in the ISO/IEC 12207 standard.

B. Related Work

Although many studies in the literature focus on the TD concept at the coding level in software projects, relatively few explore TD from alternative perspectives [16]. These include integrating TD management into project management [2], developing taxonomies for TD prioritization [17], defining debt stories with a holistic approach that accounts for social dimension [18], utilizing AI in TD manage activities [19], and creating prototype web applications to visualize and interact with various TD-related data [20].

In contrast, research on TD in AI-integrated SD projects is limited. As Amershi et al. [14] pointed out, AI-integrated SD projects possess certain characteristics that distinguish them

² https://www.itu.int/

from traditional software development projects. From the perspective of TD identification, Holt et al. [3], by highlighting these differences, have explored several ML-specific TD. Bogner et al. [9] specified TD and antipatterns in AI systems, and defined four new debt types —model, ethics, data, and configuration— alongside established TD types. Moreschini et al., [21] identified 15 potential TD types and their justifications, which align with the interview data used in this paper. Addressing the dimension of TDM; Wang et al. [22] proposed a TDM model for industrial ML based on interviews with 15 practitioners. The model begins with raising awareness of TD, followed by identifying its type and assessing the effort needed to address it. TD is then prioritized, actions are taken to resolve it, and the process concludes with a retrospective to evaluate outcomes.

Although these studies significantly advance our understanding of TD, a comprehensive perspective linking TD with specific project lifecycle processes is still lacking in the AI domain. There is a noticeable absence in the literature of studies that systematically map the processes and practices within the AI lifecycle processes in relation to TD, clearly defining the specific types of relationships involved. Establishing this relationship through a structured mapping of relevant processes and practices could be essential in enhancing the identification, management, and even prevention of TD.

III. METHODOLOGY

A. Research Design

In line with the purpose of the study, we aim to answer the following research questions:

1) Which processes outlined in the ISO/IEC 5338 standard, when overlooked, contribute to the root causes of Data Debt in AI-integrated SD projects?

2) What is the distribution of processes that contribute to the root causes of Data Debt across Data Debt sub-TD categories in AI-integrated SD projects? *3)* What ISO/IEC 5338 and ISO/IEC 12207 activities help prevent Data Debt, and how are these activities distributed across data-debt cases in projects?

In this study, we adopted a qualitative research approach to analyze and interpret the relationships between Data Debt and specific processes and activities within AI development lifecycle. Through detailed data debt-related case analysis, we systematically linked each case to relevant processes and practices in ISO/IEC 5338 and ISO/IEC 12207. By examining these connections, we analyzed the outcomes to identify insights regarding the formation of TD in AI-integrated software projects. The workflow of our methodology is given in Figure I, illustrated using the business process management notation (BPMN).

B. Data Preparation

1) The Source Data

This research used a sample drawn from the dataset created by our previous study [23] in which 309 cases were collected through semi-structured interviews with 18 AI domain experts involved in AI-integrated software development projects. These cases were analyzed and categorized using a coding process, followed by an evaluation conducted by academic experts. The study includes 15 companies from various countries, such as Türkiye, Germany and the U.S., representing a range of sizes from small businesses (1-50 employees) to large enterprises (10,000+ employees), where their years of operation range from 1973 to 2019. This diversity in size, geography and longevity ensured a broad yet mature industry perspective on ML-related challenges. The business domains of the companies varied significantly such as online fraud detection. education, content recommendation, image recognition and time series forecasting. This previous research aimed to investigate the TD issues, root causes, solutions and band-aid problems in ML projects considering the CRISP-DM lifecycle stages. As a result of this research, 13 main and 53 sub-categories of technical debt were defined through an itera-



Fig. 1. Workflow of the Study

tive approach based on 145 examples from 21 different AI-ML projects obtained through interviews. The categories and their relationships were verified systematically.

The following were identified as the main TD categories using our interview dataset: People Debt, Project Management Debt, Team Debt, Requirement Debt, Resource Management Debt, Data Debt, Deployment Debt, Code Debt, Dependency Debt, Design Debt, Model Debt, Privacy and Compliance Debt, System Maintenance and Security Debt. In this current study, we focused only on the TD cases related to the Data Debt sub-TD category.

The main dataset was systematically validated by an independent researcher from the study team, who was not involved in the development of the dataset. This validation employed a 5-point Likert scale, assessing the dataset based on the criteria of understandability, level of detail, internal consistency, and ease of use, as outlined in the research by Sonnenberg and Vom Brocke [24]. It was determined that only 1% of cases scored below four for the Understandability criterion, while 2% scored below four for the Level of Detail criterion. For the Internal Consistency criterion, the alignment between TD types and sub-TD types was assessed and deemed appropriate, as the categorization ensured logical coherence and sufficiently distinguished the boundaries of each sub-TD type within its respective TD type. Regarding the Ease-of-Use criterion, it was observed that the dataset is well-structured, includes detailed content, and provides the necessary information for the intended mapping process. In conclusion, the dataset has been deemed adequate and ready for use in this study.

2) Sampled Data

The dataset then was filtered to include only entries categorized under "**Data Debt**". The study utilized attributes detailed in Table I, where the Abstract Case Description provides a summary of the issue, and the Root Cause field identifies the contributing factors. Multiple root causes or subcategories of Data Debt can be linked to a single case description. For instance, the description "No previously labeled data" includes root causes such as "Customers' video data do not inherently contain labels" (Data Labeling Design and Construction Debt) and "Lack of data collection at the customer side" (Data Acquisition Debt). Table I provides an example of the dataset used in this study and all of these fields are provided from the study in [23].

 TABLE I.
 Example Data Debt-Related Cases from Data Subset

Sub-TD Category	Abstract Case Description	Root Cause of TD			
Data Acquisition Debt	Missing downstream data during time series forecasting	Lack of monitoring of dependent downstream data			
Data Labeling Design and Construction Debt	Redundant labeling processes	The ML team's unawareness of existing pre-trained models that could be used for data handling			
Data Preprocessing Debt	The annotation team cannot handle the increased workload of data labeling	High maintenance requirements for data labeling due to the rapid expansion of the project			

The Sub-TD Category field given in Table I includes the following sub-categories for Data Debt given in Table II. The created subset consists of 73 cases divided into 7 categories. The distribution of the subset according to subcategories and the total number of the cases is given in Table II.

TABLE II. DATA DEBT SUB-TD CATEGORIES AND THEIR FREQUENCIES (F)

Data Debt Sub-TD Category	Definition	F
Data Acquisition Debt	Difficulties in obtaining essential data for ML projects, such as challenges with third-party data and infeasible data collection	21
Data Labeling Design and Construction Debt	Poorly planned and executed data labeling processes lead to ambiguous protocols, inefficient use of resources, and an absence of standardization	18
Data Labeling Outsourcing Debt	Issues arising from outsourced data labeling, resulting in reduced quality and project delays	2
Data Labeling Quality Debt	Quality issues in labeled data, including bias, inconsistencies, and limited labeled instances	6
Data Preprocessing Debt	Difficulties in preparing and transforming data for ML projects, involving complex data handling and labor-intensive preprocessing tasks	7
Data Quality Debt	Challenges with data integrity, consistency, and reliability in ML projects, such as missing values, noisy data, and inconsistencies in data types	9
Dataset Structure and Representation Debt	Challenges in structuring and representing datasets to comprehensively address all relevant use cases	10
	Total	73

C. Mapping Data Debt TD Attributes with Processes

When associating TD with these processes, three critical aspects were jointly considered: the Data Debt sub-TD Category, the Abstract Case Description, and the Root Cause of TD. For each mapped process, the specific Activity and Task in either ISO/IEC 5338 or ISO/IEC 12207 to which it aligns were identified where applicable, alongside the detailed information supporting this alignment. Additionally, the source of this alignment was documented. We identified all the processes related to the Data Debt sub-TD categories, regardless of the extent of their coverage. The process can either fully or partially cover the issue mentioned in the debt description. Table III below shows three examples of how we matched data debt-related cases with the standard's processes.

A sample from our mapping processes can be seen in Table III. For instance, the case description "Missing downstream data during time series forecasting" and its root cause "Dependencies on upstream data sources" were observed in a project where the company depends on an upstream time series data for forecasting purposes. After deployment, the company discovered that data was missing at different dates. The company had assumed the upstream source would operate reliably and, as a result, had not implemented monitoring mechanisms or conducted a risk assessment for this dependency. Hence, we associated this debt with the ISO/IEC 5338 Risk Management (Mgmt.) Process. Dependencies on upstream data sources mean that any disruption or inaccuracy

in these sources can lead to incomplete or delayed downstream data, impacting forecast reliability. This case is connected to Data Acquisition Debt type because, the failure to properly acquire, monitor, or maintain downstream data introduces inefficiencies and risks in the system's ability to function effectively. In the context of risk management, these issues require a continuous monitoring and management approach to ensure data dependencies are addressed, and the accuracy of the forecasting models is maintained.

The data debt-related case "Redundant labeling processes" and its root cause "The ML team's unawareness of existing pretrained models that could be used for data handling" were observed in a project within the computer vision domain focusing on self-checkout loss prevention. Initially, the team employed a basic key point model for annotation tasks, manually labeling skeletal structures such as hand and shoulder positions. However, this approach stemmed from insufficient initial research into alternative solutions. At a later stage, the team conducted a more thorough investigation and identified pre-trained models capable of delivering superior performance without requiring additional annotations. These publicly available models proved to be highly effective, leading the team to abandon the redundant labeling processes in favor of leveraging these advanced pre-trained solutions. We associated this debt with the ISO/IEC 5338 Human Resource Management (Mgmt.) Process, as this process encompasses the activity "Acquire and provide skills" which involves ensuring qualified personnel are available when skill deficits are identified. Both issues outlined in this case indicate gaps in the team's knowledge and skill set. Redundant labeling processes suggest inefficiencies that could be reduced by training personnel in best practices or by implementing automated or optimized labeling techniques. Similarly, the lack of awareness of existing pre-trained models highlights a knowledge gap that could be addressed by upskilling the team on current AI tools and resources. Addressing these deficits ensures the team is equipped with the necessary skills and knowledge to perform tasks effectively, optimizing resource use and improving the overall efficiency of the AI project.

 TABLE III.
 Example Mapping of Data Debt TD Attributes with the Processes

Data Debt Sub-TD Category	Abstract Case Description	Root Cause of TD	Processes	Source
Data Acquisition Debt	Missing downstream data during time series forecasting	Lack of monitoring of dependent downstream data	Risk Mgmt. Process	ISO / IEC 5338
Data Labeling Design and Construction Debt	Redundant labeling processes	The ML team's unawareness of existing pre-trained models that could be used for data handling	Human Resource Mgmt. Process	ISO / IEC 12207

³ https://github.com/MEdata4/TD

Mapping was performed by one of the authors of this study over a period of 15 person-days, following the established mapping guideline³.

D. Verification of the Mapping Output

After mapping the Data Debt TD attributes with the standards' processes and activities, verification was carried out to ensure the accuracy and reliability of the established associations. This involved assessing the consistency of the relationships between the data debt-related cases and the relevant processes, as well as confirming that the reasoning provided aligns with the standards and activities.

The mapping process can fail in two ways: false positives may occur when the rationale for a given mapping is based on weak, false or insufficient explanations, resulting in a mapping that should not have been made. Conversely, false negatives may occur when potential mappings between candidate processes and debt categories are missed. With these threats, we applied a two staged verification process:

1) LLM based Verification

We used LLM-as-a-judge strategy for verification of the mapping process, where LLM effectively acts as a human annotator or expert reviewer/evaluator, providing a mapping between the TD and the ISO document section with its reasoning [25]. This task aligns with the "Question Answering" task in the natural language processing domain, where the TD item serves as the question, and the corresponding section or paragraph in the ISO document constitutes the answer. Here, we presented the context (i.e., ISO sections) with the query (TD item) and tasked the LLM with determining whether any part of the context could be mapped to the given query-an information extraction task. Given the limited dataset size, we opted for a prompt-tuning approach over fine-tuning LLMs. This involved designing and evaluating six distinct prompts using an independent sample dataset, focusing on factors like context, persona, and templates [26]. The prompts were tested on a specific process to assess their effectiveness in establishing relationships with cases and providing reasoning. It is important to note that the effectiveness of prompt templates can vary based on the LLM, its parameters (e.g., temperature), and the dataset. Therefore, prompt engineering requires a human-in-the-loop approach and remains highly context-dependent. We used the ChatGPT tool for bidirectional evaluation of the mapping output. GPT-40 mini model [27] was preferred due to its high request capacity. The temperature value was set to zero to achieve more consistent and deterministic outputs.

"You are analyzing #technical_debt (TD) in #AI_development_projects. Your goal is to evaluate whether each <Data Debt Sub-TD Category>, <Abstract Case Description> and <Root Cause> relates to specific <ISO/IEC Process> and <ISO/IEC Activities, Tasks & Special Notes>. Complete the following sentence for each input group I will give. <Data Debt Sub-TD Category> and <Abstract Case Description> are associated with ISO/IEC 5338 <ISO/IEC Process> because ..." Firstly, to identify false positive associations in the Mapping Table V1.0, we used the above prompt template, which consistently proved to be the most effective across multiple prompt trials. Following this prompt, we provided the TD attribute details for each data debt-related case and the already associated process (i.e. the content given in Table III) line by line. Afterwards, the ChatGPT responses were gathered in the following format (i.e. ChatGPT Output) given in Table IV.

TABLE IV. EXAMPLE VERIFICATION USING CHATGE	PΤ
---	----

Input Parameters							
Data Debt Sub-TD Category:	Dataset Structure and Representation Debt						
Abstract Case Description:	The model has low success in small object detection						
Root Cause:	Lack of diverse training data						
ISO/IEC Process:	6.3.8 Quality Assurance Process						
ISO/IEC Activities, Tasks & Special Notes	6.3.8.3 The processes to be evaluated should include activities for ** tasks for analysing requirements and risks to ensure adequate coverage of the problem domain of interest ** Examples of effects that should be monitored by means of quality assurance include: data offered to the model is of low quality **						
ChatGPT Output							
Dataset Structure and Representation Debt and the model's low success in small object detection are associated with ISO/IEC 5338 Quality Assurance Process because this process includes essential activities for assessing data adequacy and model performance relative							

Quality Assurance Process because this process includes essential activities for assessing data adequacy and model performance relative to the problem domain. The lack of diverse training data limits the model's capability in specific tasks, such as detecting small objects, which highlights a gap in data representativeness. Quality Assurance involves evaluating requirements and analyzing risks to ensure comprehensive coverage, helping to identify data limitations and align the dataset more closely with the domain's needs, ultimately improving model effectiveness.

The relationships were reevaluated by the author conducted the mapping, considering both the reasoning provided by ChatGPT and the author's judgment. For instance, the reasoning given by ChatGPT for the relationship between "Dataset Structure and Representation Debt / Issues related to irrelevant and redundant data collection processes" and "Information Management Process" is that effective information management requires collecting only relevant and necessary data, and mismanagement leads to excessive and irrelevant data. However, it was deemed insufficiently convincing to confirm the relationship. Upon review, this case is not related to the ISO 5338 Information Management Process, because the process focuses on managing existing information rather than the collection or structuring of data. As a result of this initial validation stage, it was determined that two relationships established in the initial mapping (Mapping Table V1.0) were incorrect. Subsequently, these false positive cases were removed, and Mapping Table V2.0 was created.

Secondly, to address false negative associations, we developed five alternative prompts using the Human Resource Management Process as a sample case. The goal was to identify the most effective prompt for linking processes to data debtrelated cases, based on ChatGPT's reasoning. The prompt³ (Accessible from the github page), which was tested with a oneshot learning approach, performed best, achieving 8/9 correct associations and identifying the most extensive set of accurate relationships. The responses were then reviewed, uncovering 58 previously overlooked process relationships, which were added to the mapping table, now updated as Mapping Table V3.0.

2) Manual Verification

The second approach involved two academics to manually inspect the mapping results revised after the LLM verification phase, to evaluate them in terms of completeness and accuracy. A total of 258 cases were randomly assigned for review between an AI expert with over 20 years of experience and a data scientist with five years of experience. Of these, 104 cases were reviewed by the data scientist, who spent 4 man-hours on validation, while the AI expert reviewed the remainder, spending 3 man-hours on validation.

Two domain experts examined three case attributes-"Sub-TD Category", "Abstract Case Description", and "Root Cause of TD"-while investigating the "ISO/IEC 5338/12207 Processes" associated with each data related debt case. During this evaluation, the "Activities and Tasks" that justified the association with the process, along with the supporting "Details" from the relevant sections of the standard and the "Reason for Association" column, were reviewed. After the manual verification, five of the 258 established relationships were removed, and five new relationships were added. As a result, the Final Mapping Table³ (Accessible from the github page) was created, containing 258 established relationships. For each relationship, the table includes the attributes "Sub-TD Category", "Abstract Case Description", "Root Cause of TD", "ISO/IEC 5338/12207 Processes", "Activities and Tasks", "Details", and "Reason for Association".

IV. RESULTS

The results of this study are presented below for each research question:

1) Which processes outlined in the ISO/IEC 5338 standard, when overlooked, contribute to the root causes of Data Debt in AI-integrated SD projects?



Fig. 2. The percentage of the unique data debt related cases across the processes is shown. X axis shows the total count of the cases.

Figure 2 was created using the number of unique data debt related cases based on processes mapped to each AI process. To calculate it the relationships were filtered by processes, and repetitive abstract case descriptions-arising from different root causes or Data Debt sub-TD categories-were eliminated before counting. This approach ensured that each case was associated with each process a maximum of one time, eliminating potential bias in the analysis that could arise from the number of root causes or Data Debt sub-TD categories linked to a given case. As expected, the AI Data Engineering Process holds the largest share at 26.8%, because it directly addresses data acquisition, preprocessing, labeling, and management-core tasks where Data Debt often accumulates. It is followed by the Quality Assurance and Risk Management processes, which are critical for ensuring data quality in terms of accuracy, completeness, usability, accessibility, etc., and for managing risks associated with data operations, such as data unavailability, inconsistent formats, and a shortage of personnel for labeling tasks. However, Data Debt can also arise from a variety of processes beyond technical ones like Design Definition and System Requirements Definition, which are affected by incomplete or evolving specifications. For example, the Acquisition process within agreement processes may lead to Data Debt due to unclear or misaligned contracts.

Similarly, organizational project-enabling processes such as Quality Management or Human Resources Management can contribute to Data Debt when labeling protocols are undefined or skilled personnel for proper data handling are lacking. These processes were found to potentially trigger Data Debt when overlooked or partially implemented.

2) What is the distribution of processes that contribute to the root causes of Data Debt across Data Debt sub-TD categories in AI-integrated SD projects?

This second research question aims to study the distribution of these aforementioned processes across different subcategories of Data Debt. The tree map in Figure 3 provides a visual overview of Data Debt sub-TD categories and their associated processes in ISO/IEC 5338 and ISO/IEC 12207, highlighted by data debt-related TD case counts. Each Data Debt sub-TD category is represented at the top of each colored area, while the processes associated with each category are depicted in different colors as assigned in the legend, allowing for straightforward differentiation. The size of each segment correlates with the number of unique data debt related cases based on processes and Data Debt sub-TD categories, which is indicated in parentheses, highlighting categories and related processes with a higher prevalence. To determine the number of unique data debt related cases based on processes and Data Debt sub-TD categories, the relationships were grouped by Data Debt sub-TD categories, filtered by processes within each group, and repetitive abstract case descriptions-arising from different root causes-were eliminated before counting. This approach ensured that within each Data Debt group, each case was associated with each process a maximum of one time, eliminating the potential bias in the analysis that could arise from the number of root causes linked to a given case. The tree map shows that the AI Data Engineering process represents the largest share across all debt categories, with the Quality Assurance and Risk Management processes being the most influential across nearly all Data Debt sub-TD types. This distribution largely aligns with the findings of RQ1. In the case of Data Acquisition debt, the Acquisition process takes a more prominent role than the other two processes, as expected. Notably, the Human Resource Management (HRM) process is concentrated in the Data Labeling Design and Construction Debt category, which shares a secondary position with Quality Management. One of the TDs related to the HRM process is the lack of skilled personnel for data labeling tasks, often resulting in delays, inconsistencies, or quality issues. While the ISO/IEC 5338 AI Data Engineering process includes the activity of "Conducting data labeling", the related section only specifies that individuals performing the labeling should possess domain expertise in the subject matter being labeled. On the other hand, the HRM process emphasizes the need for skilled personnel throughout the AI lifecycle, highlighting the importance of continuously assessing the appropriateness of human resources and their competencies, as well as implementing necessary training and development initiatives to address potential skill gaps.

Data Acquisition Debt Data Labeling Data Preprocessing Debt Dataset Struct			eling Design and Co structure and Repres	onstruction sentation D	Debt ebt	D D	ata Labeling Outso ata Labeling Quality	urcing [y Debt	Debt		Data Quality Del	ot						
Data Acquisition Debt		Data Labeling Design and Construction Debt			Dataset Structure and Representation Debt		Data Quality Debt			Data Labeling Quality Debt)						
															Al Data	Engineeri	ng(5)	
				Al Data Engineerin	~(11)	Hanaa	IR					AI Data Engineering(6)	Qual Assura (4)	lity ince.				
AI Data Engineerin	g(15)	Acquisition(12)		Ai Data Engineenn	g(11)	wanag	ement(o)	Al Data Engineerin	a(10)	Quali Assurar	ty ce(6)		Desig (Qualit	Qua	ity Ma	ality nage	
						C	onfigurati						n Defini tion (1)	Vana jeme nt(1)	Assura	Project	ent(2) Risk	
					System Re	ea. M	on anageme				n	Risk Management(4)	System Definition	Req. on (1)	Manage ment(1)	Plannin M g(1) n	anage nent(1)	
	Design Definition(5) Definition(4)		(4)	nt(3)	Implementa ement		Data Preprocessing Do) Deb	ot	Data Lab	eling						
	Quality Assurance(6)		Decision Manageme	Quality Management(6)			Know	(5)	tion	(3)	(2)	Al Data	Qua	lity	Acqu	_ Al Dat	a "ov	
		Continuous n Validation(2)	Continuous Validation(2)	Continuous nt(1) Validation(2)			Risk	Mana Me gomo		Acquis	Qualit	Stake holder	Engineering(3)	Assur (2)	n(1)	Engineeri	ng(2)
		(and anon(2)	Implement ation(1)			nt(2)	nt(1)		tion(1) Mana geme nt(1)	Needs &Req. Def(1)		Config	HR	Proje		Acq	
Risk Management (9)	System Reg. Definition	Project on(6) Planning(2)	Stakeholder Needs&Req. Def.(1)	Quality Assurance (4)	Project Planning (3)	Mainten ance(1)	Stakeholde r Needsℜ q. Def(1)	Risk Management (4)	Project Plannir g(1)	t Syster Defin	n Req. tion (1)	Risk Management (3)	Manag ement (1)	Mana geme nt(1)	ct Plann ing(1)	Quality Assuranc e(2)	uisit ion (1)	

Fig. 3. Distribution of Processes Across Data Debt Sub-TD Categories

3) What ISO/IEC 5338 and ISO/IEC 12207 activities help prevent Data Debt, and how are these activities distributed across data-debt cases in projects?

This research question seeks to identify which activities from the ISO/IEC 5338 and ISO/IEC 12207 standards actively contribute to mitigating or preventing Data Debt and to examine the extent to which these activities are associated with the 21 unique projects analyzed. Table V³ (Accessible from the github page) presents examples of the processes, activities or tasks associated with these processes in this study, the Data Debt sub-TD categories linked to these activities/tasks/notes, and the number of unique projects where these relationships are observed. For calculation, the relationships were grouped by process activities/tasks, filtered by Data Debt sub-TD categories within each group, and repetitive company/project ID's were eliminated before counting. In this way, within each process activities/tasks/notes group, each project was associated with each Data Debt sub-TD category a maximum of one time, eliminating the potential bias in the analysis that could arise from repetition. Lines containing activities taken from ISO/IEC 12207 are highlighted in gray. To enhance readability, Table V includes summary notes (marked with "*") instead of full excerpts from the standards.

 TABLE V.
 Example ISO/IEC 5338 & ISO/IEC 12207 Activities for Data Debt Prevention where C Stands for Project Counts

	ISO/IEC 5220 0 12207			
Process	Activity-Task-Special Note	SO/IEC 5338 & 12207 ctivity-Task-Special Note Data Debt Sub-TD Category		
		Data Acquisition Debt	9	
	Acquire or select data	Data Labeling Design and Construction Debt	2	
		Dataset Structure and Representation Debt	4	
		Data Acquisition Debt	2	
		Data Labeling Design and Construction Debt	10	
	Conduct data labeling	Data Labeling Outsourcing Debt	2	
		Data Labeling Quality Debt	2	
)ata Engineering	Analyze and explore data	Data Preprocessing Debt	1	
	for understanding	Data Quality Debt	1	
		Data Acquisition Debt	3	
	Analyze data quality	Data Labeling Quality Debt	1	
[IN]	Anaryze data quanty	Data Quality Debt	6	
		Dataset Structure and Representation Debt	4	
	Document data lineage and data provenance	Dataset Structure and Representation Debt	1	
		Data Acquisition Debt	4	
	Clean, merge and prepare data	Data Labeling Design and Construction Debt	1	
		Data Preprocessing Debt	4	
	Protect sensitive data	Data Quality Debt	1	
	Prepare artefacts for traceability and maintenance	Data Acquisition Debt	1	

Process	ISO/IEC 5338 & 12207 Activity-Task-Special Note	ISO/IEC 5338 & 12207 Activity-Task-Special Note Data Debt Sub-TD Category			
S Validation	Monitor for data drift by applying checks on the model input data	Data Acquisition Debt	1		
Continuous	Monitor requirements that are expected to change over time	Data Acquisition Debt	1		
	Prepare for quality assurance	Data Preprocessing Debt	1		
		Data Labeling Quality Debt	1		
rance	* Monitor AI system evolution and quality	Data Quality Debt	1		
	eronanon and quanty	Dataset Structure and Representation Debt	1		
		Data Acquisition Debt	4		
y Assu		Data Labeling Design and Construction Debt	3		
Qualit		Data Labeling Outsourcing Debt	2		
U	* Evaluate requirement and risk analysis	Data Labeling Quality Debt	2		
		Data Preprocessing Debt	3		
		Data Quality Debt	5		
		Dataset Structure and Representation Debt	4		
ion		Data Acquisition Debt	2		
keq. Definiti	Define system/software requirements	Data Labeling Design and Construction Debt	6		
		Dataset Structure and Representation Debt	1		
tem]	* Address privacy	Data Acquisition Debt			
Sys	systems	Data Quality Debt	1		

The "Conduct data labeling" activity is linked to Data Labeling Design and Construction Debt in 10 projects. Similar numbers are observed for acquisition activities, but overall, the relationships are not heavily concentrated in specific activity or Data Debt sub-TD category pairs. The "Clean, merge, and prepare data" activity, part of the AI Data Engineering process, is associated with five different Data Debt categories, highlighting the important role of data preparation in addressing various data-related issues. The standard emphasizes the importance of robust data preparation practices to ensure achieving the desired outcomes. For example, in the context of Data Quality Debt, improperly cleaned or merged datasets can result in inconsistencies, such as duplicated entries or missing values, which may degrade model performance. The "Define system/software requirement" activity is often overlooked in addressing data-related needs. For example, defining system and software requirements for data labeling—such as the labeling tool, labeling protocol/guidelines, dataset construction plan, and dataset scale—is crucial. When these requirements are inadequately defined, it can lead to Data Labeling Design and Construction Debt. However, merely defining these requirements is not sufficient; they must also be monitored over time to ensure their continued relevance. This is particularly important in the context of Data Acquisition Debt, where

evolving system demands and external factors may render initially defined requirements obsolete. Under the Continuous Validation process, the activity "Monitor requirements that are expected to change over time" plays an essential role. In a situation where a labeling protocol designed for binary classification needs to be updated to accommodate a multi-class problem due to changes in the operational environment, failing to proactively monitor and manage such updates could result in inefficiencies, inaccuracies in data acquisition, and ultimately exacerbate technical debt. The Quality Assurance process is closely associated with various data-debt issues. The standard emphasizes the importance of continuous quality monitoring and assurance to identify performance-impairing issues, such as data drift or concept drift, arising from the evolving nature of AI systems. Under the Continuous Validation process, the activity "Monitor for data drift by applying checks on the model input data" is also critical in addressing these challenges. For instance, in data labeling, new instances (e.g., evolving spam or fraud behavior) or the need for additional anomaly types (labels) may emerge, necessitating adaptive updates to maintain labeling accuracy and relevance. The Quality Assurance process in the standard states that evaluation activities should include requirement and risk assessments to ensure comprehensive coverage of the problem domain. In the case of outsourcing the data labeling process, a proof-of-concept using a small dataset is necessary, followed by ongoing monitoring and a quality assurance plan to ensure high-quality labeled data.

V. DISCUSSION

The integration of AI in SD introduces complex challenges, particularly in managing data-debt issues, making its effective management essential. The ISO/IEC 5338 and ISO/IEC 12207 standards provide a comprehensive list of lifecycle activities and tasks, which can be used to associate data-debt cases obtained from real-world projects. Detailing these activities and tasks from a TD perspective can aid to mitigate long-term risks, improve project outcomes, and ensure the sustainability of AI systems. In this study, we mapped 73 unique cases across seven Data Debt categories obtained from the real-world projects to the processes and activities defined in ISO/IEC 5338 and ISO/IEC 12207. We identified the most critical processes and explored how they were distributed across different types of Data Debt, and defined activities that would help avoid it.

In the literature, Data Debt, including issues such as poor quality, anomalies, and unmanaged dependencies, is categorized into sub-areas such as Management, Anomalies, Quality, Relevance, and Dependencies [9]. In our study, we identified similar issues, but also observed additional sub-categories, including Data Labeling Design and Construction, Data Labeling Outsourcing, and Data Labeling Quality.

While the ISO/IEC 5338 AI lifecycle standard is relatively new, many companies continue to rely on life cycle models like TDSP or CRISP-DM for their projects. However, ISO/IEC 5338 comprehensively addresses important processes and activities to prevent/resolve data-related TD. We specified that Quality Assurance, Quality Management, HR Management, and Risk Management are the pioneer processes to prioritize when addressing the challenges of data debt in AI systems. This standard can be effectively applied to AI-integrated SD projects from a Data Debt perspective, offering guidance on interpreting and detailing various AI lifecycle processes through this lens. Furthermore, our study highlights specific practices within the standard that can help mitigate the formation of Data Debt. To build upon this foundation, it is essential to recognize that technical processes should not be considered solely within their own domains but rather in conjunction with broader process categories. For instance, even if processes such as Project Planning, Quality Management, and Quality Assurance do not explicitly mention or reference data, they should be considered in connection with the AI Data Engineering process. An example could be the issue of "Ambiguity in the labeling protocol" due to the lack of a standardized definition in labeling, which is associated with Quality Management, yet it should be addressed in a way that includes AI Data Engineering activities.

A significant challenge in AI-based systems lies in their dynamic and ever-evolving nature, which necessitates a distinct approach to quality assurance compared to traditional methods [28]. For example, Felderer and Ramler [28] stated that integrating data from diverse sources with varying characteristics requires continuous system monitoring. In this context, we also found the Continuous Validation process outlined in the ISO/IEC 5338 standard to be significant for mitigating data-related TD. Additionally, effective Risk Management plays a key role in anticipating and planning for potential data debt issues, such as the unavailability of dependent data sources or disruptions in data streaming processes.

The development of TDM models for industrial ML applications has been relatively underexplored in the literature. Wang et al. [22] proposed an extended data sheet to address the specific nature of industrial datasets, which is suggested to be a useful tool for managing TD by reducing biases and improving transparency. However, this approach does not consider the problem from a "process" perspective, which is crucial in AI-integrated software development. While our study highlights the importance of process-oriented practices in managing TD, it does not measure the extent to which these practices impact TD management. Nonetheless, it provides a robust foundation for future research to evaluate the effectiveness of these practices and offers a structured framework for organizations to integrate process-oriented strategies into their TD management efforts.

VI. MITIGATION OF VALIDITY THREATS

In this research, we employed a qualitative mapping technique to establish the relationship between data debt types and AI life cycle processes. As a result, potential validity threats have arisen. Below, we outline these threats and describe the preventive measures we implemented to address the concerns related to construct validity, internal validity, external validity, and reliability.

Construct Validity: In this study, we assumed a theoretical relationship between Data Debt types and AI life cycle processes. Ensuring construct validity depends on clearly defining and accurately representing the dataset attributes, as these are essential for capturing the concept of Data Debt and its relationship to AI life cycle processes. To address this, we developed a TD exploration structure and clearly defined the meaning of each attribute. Additionally, we conducted multiple

review sessions to ensure that the interview data was accurately mapped to the TD exploration structure. To mitigate the risk of survivorship bias in this study, we imposed a criterion that all discussed projects must have at least reached the deployment phase. This criterion ensured that our findings reflect the challenges and consequences of technical debt in real-world, operational settings rather than early-stage or unfinished projects. Besides, although many companies were sourced from our personal network, potentially leading to similar perspectives on people and cultural processes, we mitigated this by including a wide range of business domains such as Security, Finance, Retail, Automotive, Educational Technology, and Public Safety, showcasing a diverse set of ML applications, from image recognition to content recommendation. This strategy provided a wide spectrum of perspectives from various business contexts.

Internal Validity: Achieving it involves addressing systemic or procedural errors that could weaken confidence in the study's outcomes. One threat to internal validity lies in the possibility of alternative mapping choices between Data Debt types and AI life cycle processes. To address this, the mapping of industry Data Debt cases to AI life cycle processes was conducted by a single researcher following an initial agreement on the mapping methodology. This agreement established clear guidelines for how the mapping should be performed. Subsequently, two additional researchers reviewed and evaluated the process to ensure consistency and coherence. It is important to note that the mapping was also significantly influenced by the definitions provided in the ISO/IEC 5338 and ISO/IEC 12207 standards.

External Validity: The data for this research was gathered from a specific group of AI-integrated software development projects, based on insights from 18 practitioners involved in 22 ML projects. Although the study encompasses a diverse range of domains, project sizes, and experience levels among the participants, the findings may not fully extend to all types of AI applications, particularly generative AI systems.

Reliability: In this study, the process of mapping data debtrelated cases to ISO/IEC processes involves a degree of subjectivity. However, the LLM-based validation approach proposed here adds an important dimension to reproducibility. By using the provided prompts, the results can be replicated, enhancing the reliability of the study's findings.

VII. CONCLUSION

This study explores Technical Debt Management (TDM) in AI projects by aligning Technical Debt (TD) with various lifecycle processes. It introduces a standards-based approach, particularly leveraging ISO/IEC 5338 and ISO/IEC 12207, to help AI-integrated software development companies improve TD identification, monitoring, and prevention. The study maps Data Debt cases to these standards based on subcategory, root cause, and specific details, validated by an LLM and experts. Unlike TDSP or CRISP-DM, ISO/IEC 5338 provides a more comprehensive framework for this mapping.

The research addresses three key questions: identifying key processes, analyzing their distribution across different types of Data Debt, and examining real-world mitigation efforts. The findings enhance understanding of Data Debt in AI projects, linking real-world cases to structured processes and offering practical insights for management. While based on a limited dataset, this initial study establishes a foundation for future research and applications.

(1) Cross-process associations of data debt-related cases: It has been observed that data debt-related cases are associated not only with technical processes but also with various processes from the main categories of agreement processes, organizational project-enabling processes, and technical management processes, with a tendency to accumulate more significantly in certain areas.

(2) Interpretation and reading of the standards: Technical processes should not be viewed in isolation, these processes should also be viewed under the scope of other main process categories. For instance, challenges like labeling ambiguities, often linked to Quality Management, should also incorporate AI Data Engineering activities for effective resolution.

(3) Contribution to risk mitigation in AI-integrated projects: This study provides insights that facilitate a targeted approach to risk reduction by uncovering patterns and correlations between specific processes and sub-categories of Data Debt. It emphasizes the significant roles of processes such as AI Data Engineering, Quality Assurance, and Risk Management across all debt categories. This approach enhances the understanding of both the processes at risk and the types of Data Debt they impact, contributing to improved management strategies for AIintegrated projects.

(4) Bidirectional verification using LLM: To verify the mapping process, we employed ChatGPT to evaluate both false positives and false negatives in the results. By analyzing the outputs, we discovered that some of the model's reasoning was not entirely accurate. While the innovative use of bidirectional validation through GPT proved to be highly beneficial in enhancing the overall process, it became evident that the model's responses could not be utilized uncritically. This observation underscores the importance of the "human in the loop" concept, where human expertise is indispensable in validating and refining the AI-generated results. By combining human judgment with AI capabilities, we were able to improve the reliability of the mapping.

A limitation of this study is that some aspects of Data Debt may not fully align with the processes in ISO/IEC 5338 and ISO/IEC 12207. Additionally, the study does not explicitly define the coverage level of Data Debt cases within these standards. Exploring this coverage could provide further insights into how specific processes impact technical debt management, but this aspect remains unexamined.

ACKNOWLEDGMENT

This study was partially supported by funding from the Research Universities Support Program (ADEP) under project number ADEP-704-2024-11481, provided by the Council of Higher Education (YÖK) in Türkiye. This funding facilitated the data collection and analysis phases of the study. Additionally, the mapping of the processes outlined in the ISO/IEC 5338 standard across data-debt cases was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under project number 124E655, funded as part of the TÜBİTAK 1001 program.

References

- W. Cunningham, "The WyCash Portfolio Management System," ACM SIGPLAN OOPS Messenger, 4(2):29–30, 1993.
- [2] M. Wiese, M. Riebisch, J. Schwarze, "Preventing Technical Debt by Technical Debt Aware Project Management," 2021 IEEE/ACM International Conference on Technical Debt (TechDebt), 2021.
- [3] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden Technical Debt in Machine Learning Systems," Advances in Neural Information Processing Systems, vol. 28, 2015.
- [4] A. Shome, L. Cruz, A. Deursen, "Data Smells in Public Datasets," 2022 IEEE/ACM 1st International Conference on AI Engineering - Software Engineering for AI (CAIN), 2022.
- [5] G. Recupito, R. Rapacciuolo, D. Nucci, F. Palomba, "Unmasking Data Secrets: An Empirical Investigation into Data Smells and Their Impact on Data Quality," 2024 IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI (CAIN), 2024.
- [6] H. Foidl, M. Felderer, R. Ramler, "Data Smells: Categories, Causes and Consequences, and Detection of Suspicious Data in AI-based Systems," 2022 IEEE/ACM 1st International Conference on AI Engineering -Software Engineering for AI (CAIN), 2022.
- [7] H. Zhang, L. Cruz, A. Deursen, "Code Smells for Machine Learning Applications," 2022 IEEE/ACM 1st International Conference on AI Engineering - Software Engineering for AI (CAIN), 2022.
- [8] K. Ronanki, B. Cabrero-Daniel, C. Berger, "Prompt Smells: An Omen for Undesirable Generative AI Outputs," 2024 IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI (CAIN), 2024.
- [9] J. Bogner, R. Verdecchia, I. Gerostathopoulos, "Characterizing Technical Debt and Antipatterns in AI-Based Systems: A Systematic Mapping Study," 2021 IEEE/ACM International Conference on Technical Debt (TechDebt), 2021.
- [10] V. Gudivada, A. Apon, and J. Ding, "Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations," vol. 10, 2017.
- [11] G. Recupito, F. Pecorelli, G. Catolino, V. Lenarduzzi, D. Taibi, D. Nucci, F. Palomba, "Technical debt in AI-enabled systems: On the Prevalence, Severity, Impact, and Management Strategies for Code and Architecture," Journal of Systems and Software vol. 216, 2024.
- [12] ISO/IEC, "ISO/IEC 5338: AI Life Cycle Processes," International Organization for Standardization, 2023.
- [13] ISO/IEC, "ISO/IEC 12207: Software Life Cycle Processes," International Organization for Standardization, 2008.
- [14] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann, "Software Engineering for Machine Learning: A Case Study," in 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). IEEE, 2019, pp. 291–300.
- [15] NIST (2024). NIST Trustworthy and Responsible AI NIST AI 100-5, A Plan for Global Engagement on AI Standards. https://doi.org/10.6028/NIST.AI.100-5

- [16] Z. Li, P. Avgeriou, and P. Liang, "A Systematic Mapping Study on Technical Debt and Its Management," Journal of Systems and Software, vol. 101, 2015, pp. 193–220.
- [17] D. Pina, A. Goldman, G. Tonin, "Technical Debt Prioritization: Taxonomy, Methods Results, and Practical Characteristics," 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2021.
- [18] N. Riquet, X. Devroey, B. Vanderose, "Debt Stories: Capturing Social and Technical Debt in the Industry," 2024 ACM/IEEE International Conference on Technical Debt (TechDebt), 2024.
- [19] D. Albuquerque, E. Guimaraes, G. Yonin, P. Rodriguezs, M. Perkusich, H. Almeida, A. Perksich, F. Chagas, "Managing Technical Debt Using Intelligent Techniques - A Systematic Mapping Study," IEEE Transactions on Software Engineering, vol. 49, no. 4, April 2023.
- [20] F. Bi, B. Voger-Heuser, E. Sapera, "Towards an Interdisciplinary Technical Debt Interaction and Visualization Tool," 2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS), 2023.
- [21] S. Moreschini, L. Coba, V. Lenarduzzi, "Towards a Technical Debt for AI-based Recommender System," 2024 ACM/IEEE International Conference on Technical Debt (TechDebt), 2024.
- [22] X. Wang, H. Schuster, R. Borrison, B. Klopper, "Technical Debt Management in Industrial ML - State of Practice and Management Model Proposal," 2023 IEEE 21st International Conference on Industrial Informatics, 2023.
- [23] P. D. Akman, "Analysis of Technical Debt in ML-Based Software Development Projects," M.S. thesis, Informatics Institute, Middle East Technical University, Ankara, Turkey, 2024.
- [24] C. Sonnenberg and J. Vom Brocke, "Evaluations in the science of the artificial-reconsidering the build-evaluate pattern in design science research," in Design Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings 7, Springer Berlin Heidelberg, 2012, pp. 381-397.
- [25] L. Zheng, W. L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, I. Stoica, "Judging llmas-a-judge with mt-bench and chatbot arena. Advances in Neural Information Processing Systems," 36, 46595-46623, 2023.
- [26] K. Ronanki, B. Cabrero-Daniel, J. Horkoff, C. Berger, "Requirements Engineering Using Generative AI: Prompts and Prompting Patterns," in Generative AI for Effective Software Development, Springer 2024, pp. 109–127.
- [27] OpenAI, "ChatGPT (gpt-4o-mini-2024-07-18 model) language model," OpenAI, 2023. Available: https://www.openai.com
- [28] M. Felderer and R. Ramler, "Quality Assurance for AI-Based Systems: Overview and Challenges," in Software Quality: Future Perspectives on Software Engineering Quality. SWQD 2021. Lecture Notes in Business Information Processing, vol. 404, Springer International Publishing, 2021