

Mind the Debt: Characterisation, Detection, and Evolution of Technical Debt in AI-enabled Systems

Abstract - Ph.D. Thesis

University of Salerno

Ph.D. Program in Computer Science – XXXVIII Cycle

Academic Years 2022–2025

Author: Gilberto Recupito

Supervisors: Prof. Fabio Palomba, Prof. Dario Di Nucci

Ph.D. Program Director: Prof. Andrea De Lucia

The widespread adoption of Artificial Intelligence (AI) in software systems introduces new layers of complexity to software quality assurance. Rapid technology evolution leads developers to make rushed design decisions, resulting in long-term maintainability and reliability issues. When these issues emerge in AI components, they manifest as AI-specific Technical Debt (AITD)—a phenomenon that may severely compromise software quality and lead to project failure. Although AITD has recently received increasing attention, prior studies have predominantly focused on categorizing debt types, providing limited insights into how to detect and manage them effectively.

The thesis project aimed to (i) characterize the properties of AITD issues, (ii) propose and validate novel automated detection approaches, and (iii) analyze the evolutionary accumulation of AITD in AI-enabled systems over time. The work centers on three categories of AITD—architectural debt, Machine-Learning-specific code smells (ML-CSs), data smells, and extending the investigation to self-admitted technical debt (SATD).

The findings reveal distinct patterns for each debt type. Architectural issues are difficult for developers to perceive because they require system-level visibility. In contrast, ML-CSs and data smells are frequently overlooked, suggesting their high prevalence in AI-enabled systems. For these two categories, the thesis proposes two publicly available and empirically evaluated tools to support AITD detection: CLAIRE and CodeSmile.

As a key contribution, the thesis provides a comprehensive body of knowledge and actionable insights for understanding, detecting, and managing AITD.



Mind the Debt: Characterisation, Detection, and Evolution of Technical Debt in AI-enabled Systems

Abstract - Tesi di Dottorato

Università degli studi di Salerno

Corso di Dottorato in Computer Science and Information Technologies – XXXVIII Ciclo

Anni Accademici 2022–2025

Candidato: Gilberto Recupito

Supervisori: Prof. Fabio Palomba, Prof. Dario Di Nucci

Coordinatore del Corso di Dottorato: Prof. Andrea De Lucia

L'adozione diffusa dell'intelligenza artificiale (AI) nei sistemi software introduce nuovi livelli di complessità nella garanzia della qualità del software. La rapida evoluzione tecnologica porta gli sviluppatori a prendere decisioni affrettate in materia di progettazione, con conseguenti problemi di manutenibilità e affidabilità a lungo termine. Quando questi problemi emergono nei componenti AI, si manifestano come debito tecnico specifico dell'AI (AITD), un fenomeno che può compromettere gravemente la qualità del software e portare al fallimento del progetto. Sebbene l'AITD abbia recentemente ricevuto una crescente attenzione, gli studi precedenti si sono concentrati prevalentemente sulla classificazione dei tipi di debito, fornendo informazioni limitate su come individuarli e gestirli in modo efficace.

Il progetto di tesi mirava a (i) caratterizzare le proprietà dei problemi AITD, (ii) proporre e convalidare nuovi approcci di rilevamento automatico e (iii) analizzare l'accumulo evolutivo dell'AITD nei sistemi basati sull'IA nel corso del tempo. Il lavoro si concentra su tre categorie di AITD: debito architettonico, code smell specifici del machine learning (ML-CS) e data smell, estendendo l'indagine al debito tecnico auto-ammesso (SATD).

I risultati rivelano modelli distinti per ogni tipo di debito. I problemi architetturali sono difficili da percepire per gli sviluppatori perché richiedono una visibilità a livello di sistema. Al contrario, gli ML-CS e i data smell sono spesso trascurati, il che suggerisce la loro elevata prevalenza nei sistemi basati sull'intelligenza artificiale. Per queste due categorie, la tesi propone due strumenti disponibili al pubblico e valutati empiricamente per supportare il rilevamento dell'AITD: CLAIRE e CodeSmile.

Come contributo chiave, la tesi fornisce un corpus completo di conoscenze e approfondimenti pratici per comprendere, rilevare e gestire l'AITD.