PROTOTYPING AN ETHICAL ISSUE DECISION SUPPORT TOOL FOR USER STORY ANALYSIS

Lappeenranta-Lahti University of Technology LUT

Master of Science (Tech.) in Software

Engineers for Green Deal, Master's thesis

2024

Cindy Aprilia

Examiners: Prof. Jari Porras D.Sc (Tech) (Lappeenranta-Lahti University of Technology)

Prof. Henry Muccini, Ph.D (University of L'Aquila)

Prof. Dr. Patricia Lago (Vrije Universiteit Amsterdam)

SE4GD
Software Engineers for Green Deal

With the support of the Erasmus+ Programme of the European Union

This thesis has been accepted by partner institutions of the consortium (619839-EPP-1-2020-1-FI-EPPKA1-JMD-MOB). Successful defence of this thesis is obligatory for graduation with the following national diplomas:

- Master of Computer Science (University of L'Aquila)
- Master of Science in Technology (LUT University)
- Master of Computer Science (Vrije Universiteit Amsterdam)

This thesis has been conducted in Chalmers University of Technology with close collaboration with Hochschule Furtwangen University.
The supervisors of this these were Professor Birgit Penzenstadler (Chalmers University of Technology) and Professor Stefanie Betz (Hochschule Furtwangen University).
This thesis was graded by Professor Henry Muccini (Università degli Studi dell'Aquila).

ABSTRACT

Lappeenranta-Lahti University of Technology LUT

Your school: LUT School of Engineering Science

Your degree programme: Software Engineers for Green Deal (SE4GD)

Cindy Aprilia

PROTOTYPING AN ETHICAL ISSUE DECISION SUPPORT TOOL FOR USER STORY ANALYSIS

Master's thesis

2024

73 pages, 15 figures, 10 tables and 0 appendices

Supervisors: Prof. Birgit Penzenstadler, Ph.D. (Chalmers University of Technology)
Prof. Dr. Stefanie Betz (Furtwangen University)
Examiners: Prof. Jari Porras D.Sc (Tech) (Lappeenranta-Lahti University of Technology)
Prof. Henry Muccini, Ph.D (University of L'Aquila)
Prof. Dr. Patricia Lago (Vrije Universiteit Amsterdam)

Keywords: Ethics, Ethical AI, Ethical IT, Ethical Decision-Making, Agile Software Project Management Automation, Automation, ECCOLA, GPT-3.5, Ethical Risk Management, Software Development Ethics, Large Language Models (LLM), Ethical Assessment Tools

Automating ethical recommendations in software project management might potentially enhance the quality of software products. However, since machines are not ethical agents, full automation is impossible. Thus, this study explores the feasibility of automating ethical decision-making processes within agile software project management.

To navigate these viewpoints, the research employs a narrative literature review and develops a prototype utilizing the ECCOLA ethical model and the GPT-3.5 language model. This prototype aims to seamlessly integrate ethical considerations into the agile development process by providing ethical suggestions based on user stories. Although the research is in its early stages and faces limitations related to both the inherent nature of ethics and the prototyping methodology, the study demonstrates that while full automation of ethical decision-making is not feasible, automated systems can offer valuable support to human practitioners.

ACKNOWLEDGEMENTS

I would like to extend my heartfelt appreciation to the exceptional individuals who have provided invaluable support and assistance throughout the process of writing this master degree's thesis.

First and foremost, I express my deepest gratitude to God for His boundless blessings and unwavering guidance. His grace and mercy have served as a constant source of inspiration and strength, enabling me to successfully complete this project.

Secondly, I am also immensely grateful to my esteemed supervisors, Prof. Dr. Birgit Penzenstadler and Prof. Dr. Stefanie Betz, for their unwavering guidance and unwavering encouragement throughout this thesis.

Third, I would like to extend my gratitude to Prof. Jari Porras, Mrs. Susanna Koponen, Prof. Henry Muccini, Prof. Patricia Lago, and all other members of the SE4GD committee whom I cannot list individually, for providing me with the opportunity to be part of the SE4GD program. This experience has been transformative, allowing me to explore, learn, and grow. Their support and guidance have been invaluable in shaping my academic and professional journey.

Furthermore, I extend my sincerest thanks to Dr. Ville Vakkuri, Dr. Kai-Kristen Kemell and their teams for taking time out of their busy schedule to give guidance for this thesis. Your guidance and willingness to share your datasets are pivotal in shaping my research and refining my prototype. Without your unwavering support, this project would not have come to fruition.

My deepest appreciation also is delivered to Prof. Hoa Khanh Dam and his teams, Dr. Shola, Ahnaf, Quazi, Hieu, Mhreteabe, Mr. Manu Setälä and SOLITA. Your participation and willingness to share your data have been a key factor to making this thesis research possible.

I would also like to express my appreciation to Nsidibe Bassey and Mercy Bamiduro for taking time to proof-read this thesis.

I am profoundly grateful to my family, whose unwavering belief in me has been a constant source of motivation. Their continuous support and encouragement have inspired me to strive on this journey.

Last but certainly not least, I would like to express my sincere appreciation to my colleagues from the SE4GD cohort 2. Their insightful discussions, motivation, and valuable feedback have played an integral role in enhancing the overall quality of this project.

In conclusion, it is with immense gratitude that I am thanking the invaluable contributions of all those mentioned above. Their unwavering support, guidance, and collaborative spirit have been the driving force behind the successful completion of this project.

DECLARATION OF AI USAGE

Ilmoitus tekoälyn käytöstä

As required by the LUT academic regulations, I hereby declare the usage of Artificial Intelligence (AI) tools and systems in the development of this thesis.

Table 1 outlines the specific AI tools used and their respective purposes:

Table 1: List of AI tools used and their purposes

Tools	Description
Perplexity(Perplexity 2024), Gemini (Google 2024a), Bing (Microsoft 2024)	These tools were used to collect and aggregate literature for the narrative literature review section of this thesis. Since the topic of this paper is still novel, traditional literature searches through indexed websites did not yield fruitful results, as the keywords returned results primarily from titles and abstracts. Therefore, Perplexity, Gemini, and Bing were employed to extend the search, as these AI tools can retrieve content where the keywords appear within the body of the text, rather than just in titles and abstracts. The results were then carefully read, manually selected, analyzed, and incorporated into the literature review (Chapter 2).
ChatGPT-4 (OpenAI 2024a)	For supporting writing phase, ChatGPT-4 was utilized to general sentences' review and proofreading:correct grammar, punctuation, and rephrase certain sections of the text to ensure a more formal and academic tone of the sentences. To further enhance the writing clarity, beside supervisors and examiners, this thesis has been manually proof-read by two external proof-readers (Ms. Nsidibe Bassey and Ms. Mercy Bamiduro).
GPT3.5 API (OpenAI 2024b)	LLM (GPT3.5 API) are employed as transformation logic of the prototype developed as the part of this thesis (Section.4.5).
ChatGPT-4 (OpenAI 2024a)	Assistance was provided by ChatGPT-4 to debug several errors when developing the prototype as part of the thesis project.

I understand that this thesis will be subject to an AI detection check using Turnitin. Therefore, I declare that all AI tools used in the development of this thesis are in line with the university's standards and guidelines (LUT University 2023 and Vrije Universiteit Amsterdam n.d.). The information provided above accurately reflects the extent and nature of AI usage in this work.

ABBREVIATIONS

Abbreviations

AI Artificial Intelligence

ANN Artificial Neural Network

API Application Programming Interface

CPU Central Processing Unit

CSV Comma-Separated Values File(s)

ECCOLA A method for implementing ethically aligned AI systems

FR Functional Requirement

GDPR General Data Protection Regulation

JIRA a proprietary bug and product tracking product developed by Atlassian for (agile) software project management

LLM Large Language Model

NFR Non-Functional Requirement

RAM Random Access Memory

TIPP Timing, Intent, Proportionality, and Perception

TRIP Transparency, Respect, Integrity, and Patient Focus

TRIP/TIPP TRIP TIPP model

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Ethical and social considerations in software development have become increasingly important (Saltz et al. 2019). Ethical decisions bear significant implications for end-users, organizations, and the environment (Gogoll et al. 2021), because ethical lapses can lead to privacy breaches, biases, and a lack of transparency, undermining public trust, legal and financial repercussions, and many other negative feedback (K. C. Laudon and J. P. Laudon 2020).

Abdulhalim et al. 2018 found that **employing ethical tools in software development leads to superior quality outcomes**. Despite its critical importance, the ethical challenges in agile software project management are still not adequately addressed (McNamara et al. 2018) (Section. 2.2).

Since the survey by McNamara et al., which revealed a **lack of awareness regarding ethical issues in IT**, an extensive literature has been devoted to tackling this dilemma. Using the snowballing technique, by March 2024, more than 100 papers have been cited that attempt to address the gap identified by McNamara et al. Yet, this scholarly discourse has predominantly concentrated on applying ethical standard to AI-related projects, and limited exploration into other crucial areas such as data science, software engineering education, curricular development, and ethics in project management. Moreover, between those papers, we did not find any paper about automating ethical suggestion in software project management. This narrow focus marks a considerable gap, especially considering the distinct demand for effective ways to bring straightforward decision-making tools into software projects (Raith et al. 2017).

The primary objective of this research is **to assess the extent to which ethical decision-making processes in agile software development can be automated**. We focused on agile approach between other available methodologies in software project management because agile methodology, characterized by iterative development, collaboration, and flexibility, is widely researched in software project management (Dybå et al. 2014).

However, agile methodologies, while promoting flexibility and rapid delivery, often need extra reminder on risk of the product, due to their dynamic and fast-moving processes. This oversight includes ethical risk, which can lead to significant negative consequences, such as privacy violations, biased outcomes, and a lack of accountability (K. C. Laudon and J. P. Laudon 2020).

Mitchell et al. suggest **that automation can support ethical decision-making** by providing consistent recommendations, potentially enhancing the ethical quality of software projects.

In contrast, Zuber et al. 2022 claims that ethical design and decision-making processes are not possible to be applied to autonomous systems since **machines are not a moral agent**. **This different perspectives creates a research gap that this study aims to address by evaluating the potential and limitations of automating ethical decision-making in agile software development.**

The research aims to bridge the gap between theoretical and practical debates on the possibility of automating ethical suggestions in software project management. It contributes to the development of tools that support ethically responsible software development, thereby integrating ethical considerations into the agile development process.

Automating ethical suggestions might produce superior quality outcomes, especially since, although ethics is important, there is still a lack of awareness in this area. However, there is a contrasting perspective on the possibility of such automation because a machine is not a moral agent. To find a middle ground, the following central research question will be addressed: **To what extent can ethical decision-making processes in software development be automated?**
The main research question are broken down further into several sub-research questions:

Sub RQ1: In which ways can ethical suggestion automation in agile software project management be carried out?

Sub RQ2: What kind of input is needed for the prototype to have sufficient information to make suggestions?

Sub RQ3: What transformation algorithm is employed to enable the prototype to generate ethical suggestions?

Sub RQ4: What kinds of output formats are useful for the prototype to effectively deliver ethical suggestions?

Sub RQ5: What kind of actions should be taken to address the privacy concerns of clients' private data?

Sub RQ6: How can the prototype be evaluated?

To answer the research questions, we employed a **mixed-methods research** design to bridge the gap between debates on the possibility of automating ethical suggestions in software project management. This design allows for an exploration of both the theoretical impacts of automated ethical assessment in software project development through a **narrative literature review** and the creation of an applicable **prototype** to provide practical support for developers and project managers. The prototype was developed by leveraging the ECCOLA ethical model and the GPT-3.5 language model.

The developed **prototype leverages the ECCOLA ethical model and the GPT-3.5 language model** to provide automated ethical suggestions based on user stories within agile projects. The prototype functions by analyzing user stories, classifying them according to ethical dimensions derived from the ECCOLA model, and generating tailored ethical recommendations for each of the ethical dimension. These suggestions aim to guide developers and project managers in addressing potential ethical issues proactively.

Ethic is inherently mix between empathy and logic. Therefore, the primary challenge addressed in this research is the dynamic nature of ethic, where different time and place might have different view toward "good ethic". To address this inherent challenges of automating ethical decision-making, this research adopts several strategies, as shown in Subsection.4.4.3. First, the prototype incorporates a feedback loop, allowing users to refine and improve the system continuously based on real-world experiences and evolving ethical standards. Second, it emphasizes transparency by opening the flow of decision-making process, enabling users to understand and trust the recommendations provided. Third, privacy concerns are mitigated by processing data locally wherever possible and adhering to strict data protection policies when using external services.

This research investigates the feasibility and impact of automating ethical decision-making in agile software development through a novel prototype leveraging the ECCOLA ethical model and the GPT-3.5 language model. By exploring this intersection, this study contributes on enhancing ethical awareness for decision-making in agile projects, a topic of increasing importance in today's technology-driven world.

The subsequent chapters will detail the methodology (Chapter 3), literature reviews of previous studies (Chapter 2), prototype development (Chapter 4), case studies (Section 4.8), and findings (Section 5), providing a comprehensive examination of the potential for automating ethical decision-making in agile software projects. Additionally, as this project is still a novel idea on this research field, there are limitations that need to be addressed in the future work, which are shown in Chapter **??**.

2 LITERATURE REVIEW

In this chapter, we focus on gathering relevant studies of automating ethic assessment in software project management. By employing narrative literature review, we analysed the papers for acquiring insights about the state of arts of ethic assessment in software project management.

Since the topic of this paper is still novel, traditional literature searches through indexed websites did not yield fruitful results, as the keywords returned results primarily from titles and abstracts. Therefore, Perplexity, Bard, and Bing were employed to extend the search, as these AI tools can retrieve content where the keywords appear within the body of the text, rather than just in titles and abstracts.

We employed Google Scholars(Google 2024b), Perplexity(Perplexity 2024), Gemini(Google 2024a), and Bing(Microsoft 2024) to collect and aggregate literature for the narrative literature review of this thesis. The keywords given to the tools are listed on Table 2. After received the possible list of relevant studies, we accessed the full papers through LUT University 2024, then carefully read, manually selected, analyzed, and incorporated into the literature review. The selection of the studies were based on the criteria listed on Table 3.

Table 2: Search Keywords for Literature Collection

Search Keywords for Literature Collection
Studies focus on ethical IT Framework
Studies focus on ethical risks in software project management
Studies focus on challenge on automating ethic assessment
Studies focus on automatic impact/risk assessment on project management
Studies focus on automatic ethical assessment on project management
Studies focus on automatic analytic for IT project management
Studies focus on evaluation of automation in the context that not fully-rational, empathy-depended area, and ethics

Table 3: Literature Selection Criteria

Inclusion	**Exclusion**
Relevant chapters to the topic are available	Studies not available
Studies in English	Studies not in English
Reputable sources	Non peer-reviewed sources
	Only abstract or poster paper available

2.1 Background of the Study

This section provides an **overview of previous studies** that highlight **the background and the current situation (state-of-arts) regarding automating ethical assessment** in software project management. It defines ethics and discusses the **importance** of integrating ethics into software development and the existing practices. This section also focuses on the area of software project management that we focused on, which is Agile Software Project Management. Additionally, it presents the current views of IT practitioners regarding the benefits, feasibility, and challenges of automating ethical assessments in their work.

These studies provide a foundation for understanding the current landscape and identifying gaps that this research aims to address.

2.1.1 Software Project Management (Agile)

Software project management is a collection of activities that producing valuable outcome, which at least has planning, executing, and monitoring stages. Software project management has a critical impact toward the end product. Therefore, extensive researches in agile software development have been done in the past decade, are happening now, and for sure will be done in the future as well. (Dybå et al. 2014).

Agile software development is one of approach to manage software development projects. It relaxes the up-front plans and relies on informal collaboration, coordination, and learning (ibid.).

The result of systematic literature review done by Hoda et al. 2023 identified three important areas of agile project management, where prescriptive analytics would be an effective support:

1. Backlog item identification: Employing prescriptive analytics for automatically extracting and processing input into new backlog items. The inputs might come from different data sources such as requirement specifications, customers' feature requests, reported bugs reported, technical debts, retrospective meetings notes, members' discussions, end users' reviews, and even experiences from previous projects. Beyond creating user stories, the automation is expected to be able to recommend inter-dependencies between the items.

2. Backlog item refinement: Employing prescriptive analytics for decomposing user stories into granular size.

3. Risk mitigation: Employing prescriptive analytics for recommending future opportunities or reminding future risks, and even showing the implication of each decision option.

2.1.2 Ethics is a Concern in Software Development

Ethical concerns are considered as part of the quality requirements of the software product as industry should comply to restrictive laws and regulations on their respective business areas. Actions against the law and regulations might put the company under the risks of incurring judicial or administrative sanctions, significant financial losses or damage to reputation or self-regulation. On the other hand, the management of compliance risks helps the growth of the companies (Settembre-Blundo et al. 2021).

Therefore, ethical values should be explicitly and formally accounted when developing software system to have an identifiable and justifiable effect on the design (Lurie and Mark 2016), same as other risks. The finding of literature review done by Abdulhalim et al. 2018 validated that employing ethical tools in software development leads to superior quality outcomes.

The key ethical considerations to be complienced on software development are including privacy, explainability and inclusivity (fairness and non-discrimination). Privacy means providing people with control and decision over their data. Explainability articulates that system should allowing oversight, which include translating operations into intelligible outputs and providing information about where, when, and how they are used. Finally, fairness and non-discrimination promoting inclusivity and avoiding biases or prejudices. These considerations are intrinsic principles for ethical behaviour and ethically responsible action, such as respect for individuals' self-determination, justice, or trust, reinforcing awareness about the ethical dimension within their projects (Ibáñez and Olmeda 2022).

2.1.3 Ethics, Ethical Dilemma, and Ethical Judgement

Ethical issues always persistent in contemporary society. Actions or policies that affect other people always have an ethical dimension. Ethics is a branch of philosophy focused on understanding the principles that shape, guide, and limit our social interactions, as well as with the more abstract consideration of moral evaluation itself. It also examines the the possibility of general principles, that may help us negotiate ethical dilemmas (Matthews 2020).

Ethical dilemma involves making a choice between several ethical values options, which when one option is chosen, it will compromise or violate some other ethical principles or values. Ethical dilemma is not always involving solution that resulting in pleasant situation (Holt 2020).

A response to a ethical dilemma is not always a matter of “right versus wrong,” as both courses of action or decision could both seem ethically correct or incorrect. In some cases, decision have to be taken for choosing the better or best way to respond when faced with two or more similarly “right“ or ”bad” courses of action or decisions to select from (Holt 2020).

To reach ethical agreement, people with different view might do ethical debates. In ethical debate, it is hard to accept or decline the other point-of view, therefore to be reasonable reasoning from both sides should be considered (Radford University 2020). Therefore, together, they will analyze the situation further and make an ethical judgment.

Ethical judgment involves making a judgment, claim, or statement about whether an action or motives is ethically right or wrong. To reach ethical agreement in judgement, people with different view might do ethical debates. In ethical debate, it is hard to accept or decline the other point-of view, therefore to be reasonable reasoning from both sides should be considered (ibid.). Because, ethical judgment recognises the multiple, fluid, and contradictory nature of ethical life and judgment does not solely proceed by rational deliberation (Gibson et al. 2022). Ethics is not what is legal, and not ruled by religion or political views, everyone point of view should be considered (Radford University 2020). Effective ethical judgment necessitates a harmonious integration of empathic and rational deliberation.

2.1.4 IT Practitioners’ Views of Ethical Responsibility

Currently, there is still vary opinion regarding the role of software engineers in ethical responsibility. Mitchell et al. 2022 interviewed 10 software professionals at different career stages, who, at the time of the interview, were working in various organizations, including startups, SMEs, and large enterprises, to explore the state of art regarding incorporating ethics into software engineering. Some participants believed they had a degree of ethical responsibility, but found the current code of ethics did not provide enough aid in technical decision-making. Some participants even did not believe their works had ethical implications.

Regardless of the participants’ personal awareness toward ethical computing, all of the participants feel that their ethical responsibilities were unclear, or that their role did not entail ethical responsibilities at all. However, all the participants were holding concerns about ethics in computing as an industry (ibid.). Furthermore, Mitchell et al. also were proposing for enhancing software development and project management tools to assist the incorporation of ethical considerations during development, such as formalised protocol for documentation, ethical considerations identification, and automatic identification of unsafe data handling (ibid.).

Project management has a large socio-technical element with many uncertainties arising from variability in human aspects (Dam et al. 2019). Therefore, having assistance of making informed decision will be helpful toward the executioners (Mitchell et al. 2022). However, there is currently a shortage of tools supporting the design, implementation, and auditing by both technical and non-technical stakeholders. Additionally, there is a lack of consensus, standards, and tools, indicating that much progress is still needed to realize ethical computing (Barletta et al. 2023).

2.1.5 Automating Ethical Risk Mitigation on Software Project Management

The rise of Artificial intelligence (AI) has the potential for transforming the practice of software project management, as AI might addresses the challenges in agile project management by processing massive amounts of data generated from software projects to harvest useful insights and train to perform complex tasks such as estimating effort, task refinement, resource management, and sprint planning (Dam et al. 2019).

AI can assist developers by automating tasks and enabling the risk prediction, providing actionable recommendations, and even making decisions. Therefore, AI technologies are adapted and integrated to support various areas of agile project management, from automating basic administration tasks to delivering analytics-driven risk predictions and estimation, facilitating project planning, and making actionable recommendations (ibid.).

AI plays a significant role in providing risk assessment in agile project management by leveraging predictive analytics to forecast future risks. Furthermore, AI's capability in prescriptive analytics uses the results from predictive analytics to recommend the best course of actions for agile teams in specific situations, thereby aiding in risk mitigation. This includes showing the implications of each decision option to take advantage of future opportunities or mitigate future risks (ibid.).

However, fully automated ethical decision-making is not achievable, as ethical consideration is a skill learned through practice, rather than simply following a checklist (Gogoll et al. 2021). Another reason why fully automated ethical decision-making is not possible is because ethical concepts are also application-dependent. Besides the lack of knowledge and available tools for automating ethical decision-making, the types of concerns vary because both the process and the issues may differ for different applications Ibáñez and Olmeda 2022. Therefore, this state of art prompts us to the question of how possible is automating ethical assessment.

2.2 Related Works

This section reviews a variety of **previous approach** that have been done for assisting practitioners for resolving ethical concern on software development projects. While the **automation of ethical assessments in software project management is a relatively new field**, this review includes relevant studies from **adjacent areas**. It covers qualitative ethical risk assessments currently used in software project management, automated risk assessments that do not focus on ethics, and attempts to automate ethical risk assessments in other fields.

These studies gives information regarding the approach that have been taken by previous researchers for clearing the gap.

2.2.1 Qualitative Ethical Risk Assessment of Software Project Management

Embedding ethical considerations in software development processes has been explored frequently before. However, the research usually focuses on specific sectors chosen by each researcher. Therefore, sensitive areas like healthcare and profitable areas like industry have been explored more than other sectors. For examples, the work of Amugongo et al. 2023 on operationalizing AI ethics through the agile software development lifecycle was providing a case study on applying ethical principles only in e-health applications.

Further enriching the landscape of ethical integration in software development are the initiatives like SENTINEL (Trantidou et al. 2022) and DEFeND (Piras et al. 2019), which focus on cybersecurity and data protection, providing actionable frameworks for ensuring privacy and security in digital solutions. These projects exemplify the practical application of ethical principles in areas critical to user trust and regulatory compliance.

Another issue found by systematic literature review done by Raab is most of the previous researches often fall short in providing clear guidance on how these principles can be implemented in real-world scenarios (Raab 2020).

On the other hand, the work of Brey et al. 2022 are more detailed. The framework provides a holistic approach to understanding the broad spectrum of ethical considerations associated with emerging technologies and designed to be adaptable across different technological domains. As the methodology came with proper guidance, it might help to systematically identify, analyze, and address ethical risks throughout the lifecycle of a project. The inclusion of this methodology in discussions on ethical risk assessment frameworks enriches the dialogue by providing a structured, adaptable, and holistic toolset for navigating the complex ethical landscape of emerging technologies. However, this methodology is a workflow methodology, and therefore, it needs a person in charge to study and apply this methodology in project management.

Same approach were taken by Zahedi et al. 2023 discussion on risk management frameworks within Agile software development as well. Their methodologies emphasizes the integration of risk management activities into Agile workflows. This approach aligns with the collaborative and iterative nature of Agile methodologies, offering a pathway for continuous ethical and risk assessment throughout the project lifecycle. Additionally, their focus are in the wide spectrum of risks, and not specially on ethical risks.

Another workflow framework explored by Wagner 2020 emphasizes the significance of one of the ethical offsets: accountability. It stresses the importance of not treating accountability as an afterthought but as a foundational element in the design and execution of technology research projects. By embedding accountability mechanisms from the outset, technology projects can proactively address ethical, legal, and societal concerns, potentially mitigating risks and unintended consequences.

Beside frameworks, several experts published guidelines to cover this issue as well. Artificial Intelligence Set Up by the European Commission 2019 released "Ethics Guidelines for Trustworthy AI". Although not directly focusing on the Agile methodology, it contributes to establishing a workflow framework for developing trustworthy AI systems. This work underscores the importance of ethical frameworks that are both comprehensive and capable of being operationalized within the development processes.

Another innovative approach is gamification, as exemplified by Vakkuri, Kemell, Jantunen, Halme, et al. 2021. ibid. explored ECCOLA—a model for implementing ethically aligned AI systems, which stresses the importance of principles such as transparency, equality, accountability, and safety. The most prominent feature of this research is that, even though it is not yet automated, the framework comes with elaborate documentation on how to assess ethics easily.

2.2.2 Ethical Risk Assessment in Agile

The work of Kemell et al. 2022 on utilizing user stories brought an innovative approach to integrating ethical considerations directly into the Agile development process. By making model through building ethically proper user stories, ibid. highlighted the potential for user stories beyond the tools for capturing requirements, but as vehicles for ethical reflection and assessment.

To refine the understanding of ethical risk assessment within Agile methodologies, insights from the book titled ``Scenarios as "Grounded Explorations". Designing Tools for Discussing the Desirability of Emerging Technologies'' offer a valuable perspective. This approach emphasizes the creation of scenarios as a method for conducting ethical assessments, where detailed narratives about the use and implications of technologies are developed to explore ethical dimensions in a grounded, and tangible manner (Lucivero and Lucivero 2016). These scenarios, akin to expanded user stories with an ethical focus, provide a qualitative method for assessing the desirability and ethical implications of technologies during the Agile development cycle.

2.2.3 Automated Ethical Assessment

The exploration of automating ethical assessments in technology and software development is pivotal for advancing ethical considerations in these domains (Mitchell et al. 2022). There are papers that delved into different aspects of this challenge, ranging from theoretical studies to practical applications and frameworks designed to embed ethical considerations into automated systems. This subsection synthesizes insights from the provided documents to construct a comprehensive view of automating ethical assessments.

"Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations" by Mökander et al. 2021 provides a theoretical exploration of auditing automated decision-making systems from an ethical standpoint. It discourse on how ethical principles can be embedded and audited in automated systems, highlighting the challenges and limitations inherent in this process. Although theoretical, the paper underlines the need to have robust methodologies for carrying out ethical audits in automated systems, suggesting a move towards more accountable and transparent technologies.

Although paper by Thompson et al. 2021 does not correlate directly to project management or provide detailed step-by-step processes, but the paper introduced an novel approach in automating ethical assessment focused on swine body scanners. It presents an innovative approach with the potential to digitally automate ethical considerations in a specific case study.

Meanwhile, the "Intelligent Decision Support System for Building Project Management Based on Artificial Intelligence" by Wang 2023 outlines an AI-based system for managing building projects. This paper is written in a technical report style and focuses on the use of historical data in order to optimize decision-making in project management. Although it does not cover ethical assessment explicitly, the work of ibid. gives a hint regarding the applicability of AI and data analytics within the broader setting of fitting ethical considerations among complex decision-making. It suggests that incorporating ethical assessments into decision support systems could enhance their utility and alignment with ethical standards.

2.2.4 Automating Risk Assessment on Software Project Management

The framework proposed by Hung et al. 2018 provides a new approach to automate risk assessment in construction management. This novel framework employs artificial neural networks (ANN) to predict and manage risks in construction projects, concentrating on non-ethical risks like financial and operational ones. Although the main focus of this framework is on those found in construction projects alone, its underpinnings are very useful for automating risk assessment in other areas within software project management. Thus, this is an initial step towards wider adoption of automation of risk assessment which could be undertaken through diverse sectors of software project management activities. It shows how AI can augment decision making processes and improve risk strategies.

2.2.5 Automating Ethical Assessment on Software Project Management

Automating ethical assessments in software project management has been an advancements in technologically supported ethical deliberations.

Hingle et al. 2023 automated engineering ethics assessment codebooks generation using Natural Language Processing(NLP)-based methods. Their research compared a human and NLP collaboration method for codebook generation. By applying these methods to analyze students' perceptions and understanding of ethical concepts, the study demonstrated the potential of NLP to support and enhance the analysis of material in ethical education. Furthermore, the study highlights the feasibility and benefits of using advanced computational techniques to facilitate ethical assessments in educational settings.

Meanwhile, in the biopharmaceutical sector, Poplazarova et al. 2020 introduced a values-based decision-making model. The model emphasizes the application of company values through TRIP&TIPP. TRIP consists of Transparency, Respect, Integrity, and Patient focus. Meanwhile, TIPP stands for Timing, Intent, Proportionality, and Perception. The model proposed by Poplazarova et al. facilitates automated ethical decision-making in R&D contexts by incorporating a structured five-step process that leverages organizational values (TRIP) and contextual factors (TIPP). Although solely focused on biopharmaceuticals, the model shows the potential of making automated decisions based on core values and situational contexts.

2.3 Relevant Works Directly Connected to Our Research

In this section, we examine the key works and foundational research that have directly influenced the design and development of our prototype. The discussion includes the ethical models that form the theoretical basis of our system, the transformation tools that enable practical implementation, and the datasets utilized to personalize the information for each managed software project.

These studies provided essential knowledge and methodologies integrated into the components of our prototype (Figure 1). By synthesizing these components, our research builds on existing work to advance the automation of ethical assessments in software project management.

2.3.1 ECCOLA

As the baseline model to be automated, ECCOLA was chosen as this model has been tested and adjusted over the years (Subsection 4.6.1). ECCOLA is a model designed to implement ethically aligned AI systems. The model is intended to be modular and can be integrated into various existing software engineering (SE) methods. It is built upon the principles of AI ethics and aims to make these principles actionable and practical for developers (Vakkuri, Kemell, Jantunen, and Abrahamsson 2020, Vakkuri, Kemell, Jantunen, Halme, et al. 2021, Kemell et al. 2022, and Halme et al. 2024).

The ECCOLA cards are divided into eight themes that collected from various ethical guidelines of AI ethics: Analyze, Transparency, Safety & Security, Fairness, Data, Agency & Oversight, Wellbeing, and Accountability. Each theme consisting of 1 to 6 cards, with more atomic aspect of that theme. Therefore, ECCOLA consists of a deck of 21 cards, each addressing different themes and specific aspects of AI ethics. To help the user understand more regarding the ethical theme, each card contains three sections: motivation (why the topic is important), actions (what to do about it), and practical examples (making the issues tangible) (Vakkuri, Kemell, Jantunen, Halme, et al. 2021, and Halme et al. 2024).

The cards are sorted into three piles based on project stages: planning, development (sprint-by-sprint), and evaluation. At the start of each sprint (Subsection 2.3.2), select relevant cards, discuss ethical issues according to the cards, and make notes on planned actions. During the sprint, refer to the cards to ensure ethical considerations are addressed, documenting actions and discussions. At the end of the sprint, review and evaluate the notes, adjust the card selection for the next sprint, and document the entire process to maintain transparency and accountability. In summary, that ECCOLA is done in circular manner and regular retrospective meetings should be held to discuss successes, challenges, and improvements, ensuring continuous integration of ethical practices throughout the project (Vakkuri, Kemell, Jantunen, Halme, et al. 2021, and Halme et al. 2024).

2.3.2 User stories as Dataset

In order for the prototype to understand the nature of the project, user stories are used as the primary dataset to be compared against the ECCOLA cards (Kemell et al. 2022, and Halme et al. 2024). The user stories allows the prototype to personalise the ethical recommendation to each of software development project.

User stories encapsulate the functional and non-functional requirements in agile software development. These stories are typically followed a simple format that describes who the user is, what they want, and why they want it (Max Rehkopf n.d.). In agile software development, user stories usually are broken down into smaller user stories, and refined by development team in iterative manner. The sequence when the user stories evaluated and refined is called sprint.

Previous researchers have utilized user stories in various ways to integrate ethical considerations directly into the Agile development process, including the team that build ECCOLA as well (Kemell et al. 2022). They highlighted the potential of user stories for ethical reflection, on top the generic purpose of capturing requirements.

In our research, we employ user stories that are collected, cleaned and structured to facilitate comparison with ECCOLA cards. These stories are either retrieved automatically from project management tools like JIRA using API credentials or manually inputted into a CSV template specifically designed for this purpose (Section 4.5)

By using user stories as the dataset, the prototype ensures that ethical considerations are embedded into the software development lifecycle.

2.3.3 Large-Language Model (LLM)

A large language model (LLM) is a computational model notable for its ability to achieve general-purpose language generation and other natural language processing tasks such as classification. Based on language models, LLMs acquire these abilities by learning statistical relationships from vast amounts of text during a computationally intensive self-supervised and semi-supervised training process (OpenAI n.d.(a)).

Previous researchers have explored various implementations and capabilities of LLMs in different contexts. The works most relevant to ours are by Hingle et al. and Hung et al., which employ Artificial Neural Networks (ANN) and Natural Language Processing (NLP), respectively. Both studies show positive outcomes when applying automation in risk assessment. However, their focuses differ from ours. Hingle et al. focused on automating the generation of engineering ethics assessment codebooks, while Hung et al. concentrated on automating risk assessment in construction management.

In the context of this research, the chosen LLM is GPT-3.5 (Subsection 4.6.3). It is employed to automate ethical suggestions within agile software project management. The LLM model is fine-tuned with ECCOLA questions, and then integrated into the prototype with the relevant project data to provide ethical recommendations based on the ECCOLA model.

The system architecture includes various layers to handle data retrieval, preprocessing, model training, and output generation (Subsection 4.5). The Model Layer is focused on integrating the GPT-3.5 API, fine-tuning the model, and evaluating its performance.

Additionally, we put focus on protecting user's privacy as well. While, GPT-3.5 is run in third party server, we studied their data privacy policies (OpenAI n.d.(b), OpenAI 2023, OpenAI n.d.(c), and OpenAI n.d.(d)).

3 METHODOLOGY

In the following section, we outline the objective, research questions, hypotheses, scope and design, and statistical analysis of our study.

The research was carried out over a period of approximately seven months. Further elaboration on the specifics of the study can be found in the subsequent sections and subsections.

3.1 Research Objectives

The backdrop frames built from the Chapter 2 bring our inquiry into the uncharted territories of ethics in IT, focusing on the practices within software engineering and project management. Therefore, the objective will be to undertake a close examination of what has become an issue of contestation: whether ethical decision-making can be automated.

Our central question probes the extent and appropriateness of such automation, given the multifaceted ethical challenges inherent in various domains of software engineering. Through this investigation, we intend to bridge the identified gaps and contribute to the establishment of more comprehensive and practical ethical frameworks within the IT industry.

3.2 Research Questions

The central focus of this research is to investigate the possibility of autonomous ethical checking during software development process. **The research objective is to evaluate whether automating ethical checks is helpful for the software development process**.

Therefore, the primary research question guiding this study is:
RQ : To what extent can ethical decision-making processes in software development be automated?

To answer the main research question, the following sub-research questions have been made:

Sub RQ1: **In which ways can ethical suggestion automation in agile software project management be carried out?** This question is to answer in what form automation can be offered to help users solve ethical problems, without ignoring that ethics is dynamic?

Sub RQ2: **What kind of input is needed for the prototype to have sufficient information to make suggestions?** This question aims to determine the kind of input needed for the prototype to have sufficient information to make suggestions?

Sub RQ3: **What transformation algorithm is employed to enable the prototype to generate ethical suggestions?** This question aims to determine the most suitable transformation algorithm according to our project's timeline and availability of resources.

Sub RQ4: **What kinds of output formats are useful for the prototype to effectively deliver ethical suggestions?** This question aims to determine how the automation can deliver its output to the user, while aligning with the baseline model?

Sub RQ5: **What kind of actions should be taken to address the privacy concerns of clients' private data?** This question aims to determine the strategies that need to be applied to the prototype, considering the sensitivity of project management data.

Sub RQ6: **How can the prototype be evaluated?** What kind of evaluation mechanism is most suitable to cover the dynamic and multi-faceted nature of ethics, based on our project's timeline and availability of resources?

3.3 Research Design

To align with the research question and identified gap in the integration and automation of ethical considerations within Agile project management tools and processes, the design and methodology of the research are influenced by both the theoretical foundation and practical applications of AI in enhancing Agile methodologies, especially in the domain of ethical decision-making.

Therefore, mixed-methods research design could be particularly effective in the consideration on the lack of automated ethical assessment in Agile IT project management, which were informed by the works of Dam et al. 2019, and further insights from Hoda et al. 2023.

This design allows for an exploration of both the theoretical impacts of automated ethical assessment in software project development and applicable prototype to be used on one of the branching of automated risk assessment of software project management (Dam et al. 2019):

1. Theoretical Phase: Employ narrative literature review to explore previous researches, and surveys regarding the possibility of automated ethical assessment on software de-

velopment. This phase can provide deeper insights into the current dynamics, considerations, and challenges encountered previously (Chapter 2).

2. Prototyping Phase: The prototyping phase aims to build tools for automating ethical suggestions using experimental or quasi-experimental designs. The prototype serves as a proof of concept for the theories derived from the literature review. This phase involves creating a prototype to generate automated ethical suggestions in agile projects management. The prototype operates through three stages: Input, Transformation, and Output. Users can upload project information, which the Large Language Model (LLM) processes to classify user stories and provide ethical suggestions based on the ethical baseline model (ECCOLA). The prototype is then evaluated using several approaches, including consistency checks with ethical theories and models, documentation and rationalization of system decisions, and most importantly, the implementation of a feedback loop (Chapter 4 and Section 5.1).

4 PROTOTYPE DEVELOPMENT AND IMPLEMENTATION

In order to have hands-on experience toward the possibility of automating ethical concern in agile project management, a prototype of automation was made. The input consists of two main parts: ethical baseline standard models and the data that carries information regarding the project. These inputs are transformed into a series of suggestions on how the project can be improved to align better with ethical standards.

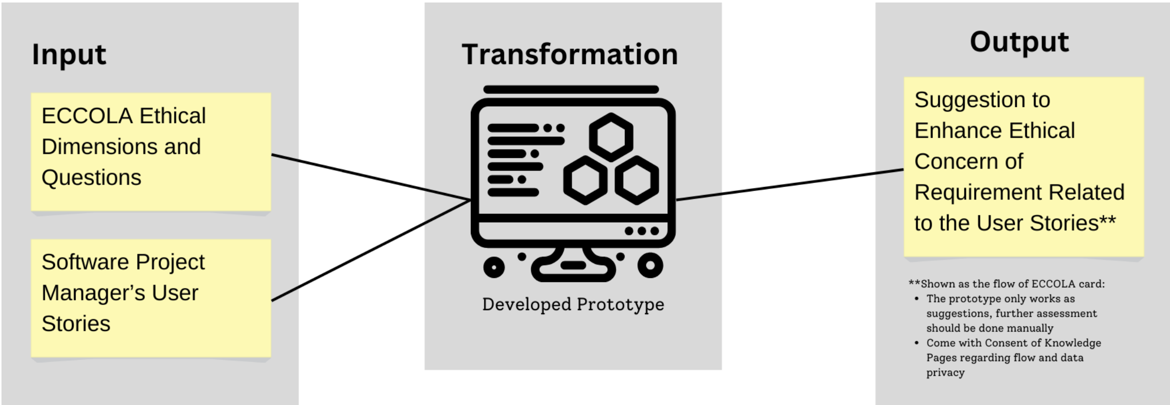

Figure 1: Flow of the system

4.1 How the Automated Tools Provide Suggestions

The prototype consists of three stages: Input, Transformation, and Output (Figure 1). The input stage involves delivering the baseline model and project information to the prototype. The transformation stage is where the LLM processes the input. The output stage delivers the transformation result to the user.

In the input stage, users upload information in the form of user stories in a CSV file. This information can be sourced from various places, such as manually listing them or exporting from agile tools like JIRA. Additionally, users are expected to upload a list of ECCOLA cards. The ECCOLA card list is included with the replicable package. By uploading the ECCOLA cards in the CSV, users have the freedom to include only the cards relevant to their project.

In the transformation stage, the LLM processes the user stories and classifies each one according to the type determined by the ECCOLA card list. To achieve this, the LLM was specifically trained with the list of ECCOLA questions and user stories related to ethical situations.

Finally, the prototype will deliver suggestions following the ECCOLA flow, where the client receives a list of questions one-by-one in sequence. These questions will be annotated with information on whether there are previous user stories relevant to them, how relevant they are, and guidance on how to expand the user stories further.

4.2 Prototype Requirements

This section outlines a detailed set of requirements expected from the prototype. This chapter will cover both the functional requirements, which ensure the prototype can effectively meet user needs, and the non-functional requirements, which are essential for the overall performance and usability of the prototype.

4.2.1 Functional Requirements

Table 4 details the functional requirements expected of the prototype. The functionalities are not presented in the sequence they will be encountered by users, rather than by their level of importance.

Table 4: Functional Requirements

Requirement ID	**Description**
	FR1.Authentications
FR1.1	The prototype must incorporate User Authentication and Authorization. Clients should be able to to input their own JIRA and OpenAI API credentials into the source.
	FR2.Data Retrieval
FR2.1	The prototype must facilitate users in retrieving the user stories template.
FR2.2	The prototype should enable users to access the ECCOLA questions list.
FR2.3	The prototype should support the retrieval and downloading of JIRA user stories and related data through API integration in CSV format, particularly for users who utilize JIRA for recording their user stories.
	FR3.Ethical Assessment
FR3.1	The prototype must classify user stories based on predefined ethical codes (ECCOLA) using the fine-tuned model.
	FR4.Data Processing
FR4.1	All operations should be executed on the client side to ensure data security and privacy.
	FR5.Model Training

FR5.1	The prototype should allow users to upload ECCOLA questions, definitions, and training samples.
FR5.2	The prototype should enable clients to adjust and re-upload data if the results are not up to their standard, thereby creating a feedback loop.
FR5.3	The prototype must then fine-tune an AI model using the provided training data.
FR5.4	The prototype logs the training process and provides real-time updates to the user, ensuring transparency and tracking of the training activities.
FR6.Result Storage and Export	
FR6.1	The prototype should store the classification results in a structured format.
FR6.2	The prototype must also allow users to export these classification results to a CSV file.
FR6.3	The exported CSV file along with ECCOLA questions (FR5.1) should be usable by users as a feedback loop for further fine-tuning, enhancing the iterative improvement of the model.
FR7.Title	
FR7.1	User Interface must offer a graphical user interface for uploading files, starting processes, and viewing results.
FR7.2	User Interface should be lightweight and simple to minimize device resource consumption.
FR3.Acknowledgment and Consent (Subsection 4.4.2)	
FR8.1	The prototype must display an acknowledgment section that outlines the steps taken by the code.
FR8.2	The acknowledgement section should explain how and where the data are processed.
FR8.3	It is imperative that users acknowledge their understanding of the process and consent to the use of their data.

4.2.2 Non-Functional Requirements

Table 5 details the non-functional requirements of the prototype.

Table 5: Non-Functional Requirements

Requirement ID	**Description**
NFR1.Scalability	

NFR1.1	The prototype's architecture should support the addition of new features and modules with minimal changes.
NFR2.Compliance	
NFR2.1	The prototype must include an acknowledgment notice that outlines the system flow, ensuring users are informed about how their data will be processed.
NFR2.2	The privacy policies of third-party components must be thoroughly reviewed.
NFR3.Performance	
NFR4.1	The prototype should be able to run smoothly on the available device (Section 4.3).
NFR4.2	The prototype must also process and classify user stories within a reasonable time frame, specifically within 5 seconds per user story in the dataset.
NFR4.3	The prototype must provide real-time feedback during data processing and model training.
NFR5.Usability and Documentation	
NFR5.1	The prototype should provide comprehensive documentation to assist users in navigating and utilizing the system effectively.
NFR6.Maintainability	
NFR6.1	The prototype's codebase follows standard coding practices and is well-documented, facilitating easy updates and modifications.
NFR7.Interoperability	
NFR7.1	The prototype should integrate seamlessly with existing JIRA infrastructure and allow for manual uploads in CSV format for other project management tools
NFR7.2	The prototype must support file formats for data import and export in CSV.
NFR8.Reliability	
NFR8.1	The code should handle errors gracefully, providing meaningful error messages to the user.

4.3 Specifications of Available Device

In Table 6, we provide a detailed overview of the hardware specifications of the device used for the enhancement and testing of the prototype. Understanding the technical capabilities of this device is crucial for replicating the study.

Table 6: Device Specification

OS Name	Microsoft Windows 11 Home Single Language
Version	10.0.22631 Build 22631
OS Manufacturer	Microsoft Corporation
BIOS Version/Date	American Megatrends International, LLC. X415EP.308, 27/04/2022
SMBIOS Version	3.3
BIOS Mode	UEFI
System Manufacturer	ASUSTeK COMPUTER INC.
System Model	VivoBook_ASUSLaptop X415EP_A416EP
BaseBoard Manufacturer	ASUSTeK COMPUTER INC.
BaseBoard Product	X415EP
System Type	x64-based PC
System SKU	
Processor	11th Gen Intel(R) Core(TM) i7-1165G7
Processor Clock Speed	2.80GHz, 2803 Mhz
CPU Physical Cores	4 Core(s)
Logical Processor(s)	8
RAM	11.7 GB

By presenting this information, we aim to provide a clear understanding of the technical environment in which the prototype was tested. For further research in this area, we would suggest to employ higher device specification.

4.4 Ethical considerations and complexity in software projects

The following section outlines several ethical considerations that are crucial for the success of this project. These concerns and challenges must be noted and addressed to ensure the project's integrity and ethical compliance.

4.4.1 Data privacy, and security

Implementing AI in Agile project management involves processing potentially sensitive project data. Addressing data privacy and security concerns is crucial (Raab 2020). Therefore, as the enhancement might have the critical data of users, the whole process ideally should be processed locally or in the server owned privately by users. However, due to the availability of capable machine, in this prototype we employed a third-party LLM, whose data protection policies have been thoroughly reviewed (OpenAI n.d.(b), OpenAI n.d.(d), OpenAI 2023, and OpenAI n.d.(c)).

4.4.2 Transparency and accountability of automated tools

Ensuring that automated ethical assessment tools are transparent in their decision-making processes and accountable for their recommendations is essential to gain trust from Agile teams (Nicodeme 2020).

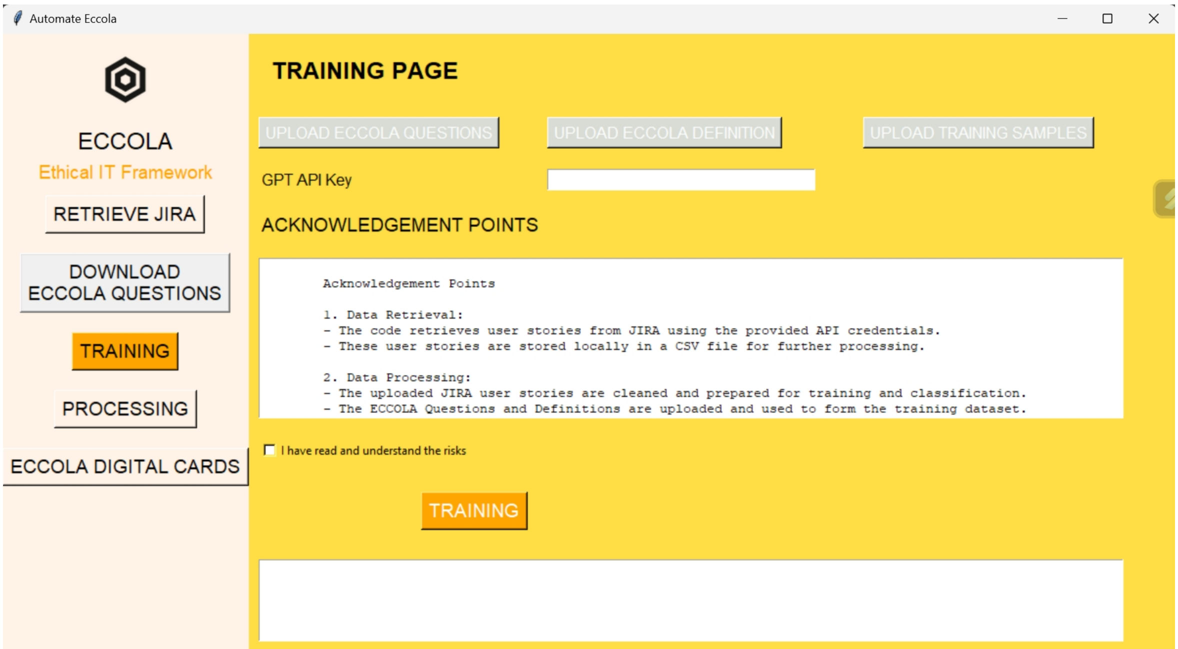

Figure 2: Acknowledgement Notes on Flow and Data Processing for the Training Feature

To achieve this purpose, an acknowledgement page was made for educating software practitioners regarding the process flow and processed data of the system. The acknowledgement section in Figure 2 is intended for people to know how their data is used for fine-tuning the LLM. And, the acknowledgement section in Figure 3 are placed for user to know how the system able to generate the suggestions.

All the process flow and processed data is opened to adhere on public Engagement & Awareness, data privacy, and transparency purpose.

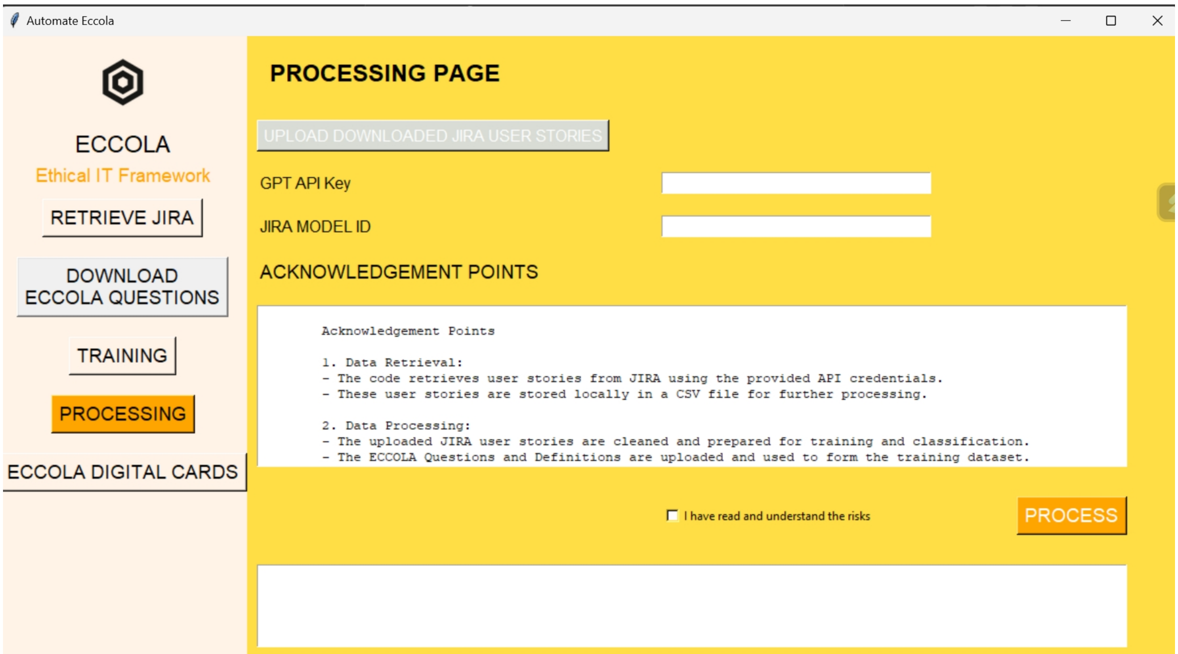

Figure 3: Acknowledgement Notes on Flow and Data Processing for the Processing Feature

4.4.3 Evaluation technique

Ethical decision-making in software projects can be highly complex and context-dependent. Ensuring that automated tools can navigate these complexities without oversimplification is a significant challenge (Gogoll et al. 2021).

This ethical complexity prevented the prototype to be validated quantitatively. Therefore, stakeholder engagement and feedback process feature are added to raise the credibility of the prototype. This strategy consist of 3 stages:

1. Consistency Checks with Ethical Theories and Frameworks Alignment: Ensure the system developed according to the flow of the ethical baseline model(ECCOLA)

2. Documentation and Rationalization: Ensure that the system provides detailed documentation and rationalization for each decision. This can help in auditing the system's decisions and ensuring they align with ECCOLA principles.

3. Feedback Loop: Use structured feedback to identify and address any logical inconsistencies or gaps in the decision-making process. Collect the system's decisions, let user to give feedback regarding the system's decision, and use it to refine the system (Richardson 2011).

4.5 System Architecture

This section provides a high-level system architecture of the prototype. This architecture will include different components involved in the process, from data retrieval to model training and ethical classification. Figure 4 provides an overview of the components and their interactions within the transformation.

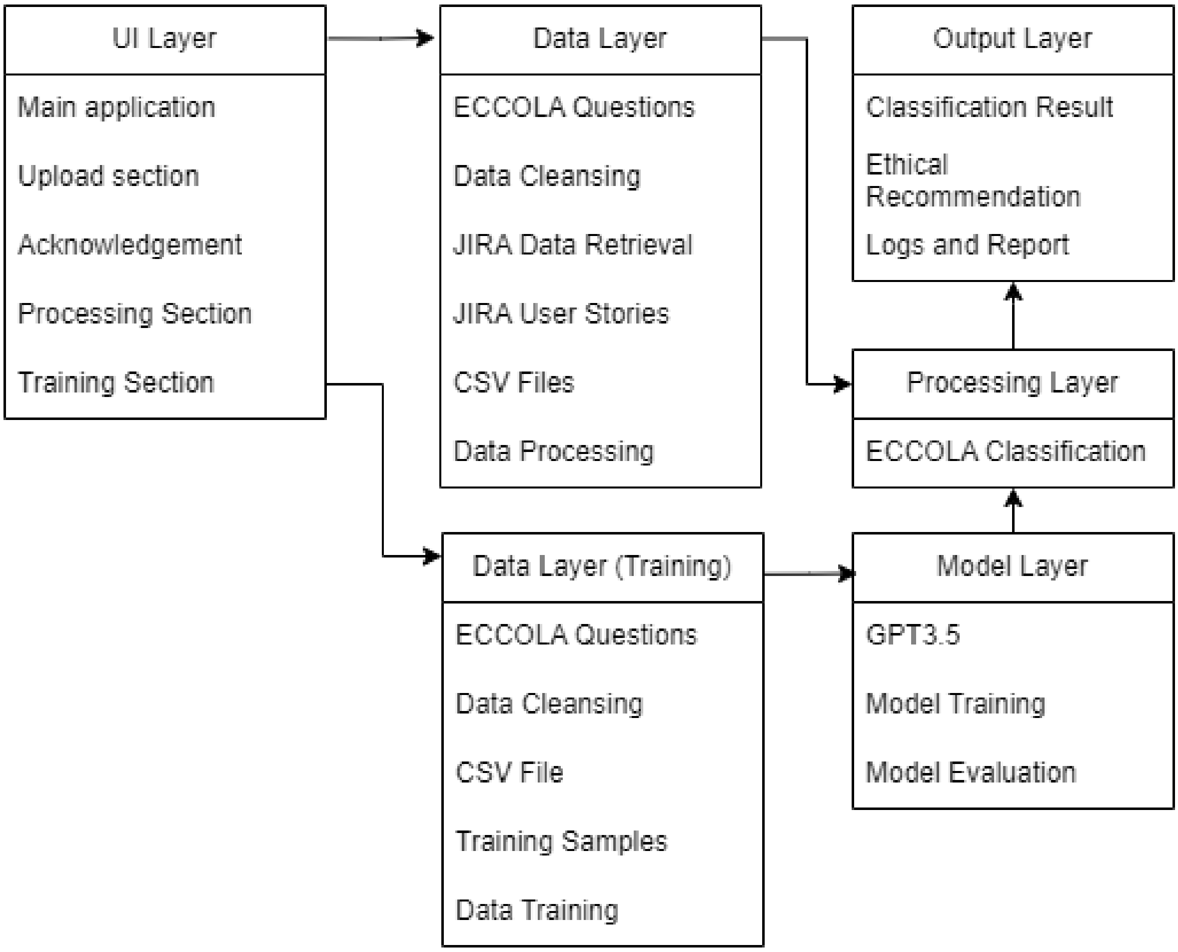

Figure 4: System Architecture

The User Interface (UI) Layer is designed to offer a user-friendly platform for interacting with the system. Its primary goals include providing an interface that allows users to upload data, initiate processes, and view acknowledgment points and logs. The functions of this layer encompass providing a graphical user interface (GUI) where users can upload ECCOLA questions, JIRA user stories, and training samples. It also enables users to start processing and training tasks, displays acknowledgment points, and outputs logs and final classification results. The key components of this layer are the main application, whose front-end was built in Python using Tkinter. The front-end has sections for uploading various datasets (ECCOLA questions, JIRA user stories, and training samples), as well as sections dedicated to processing, training, and receiving suggestions. In any section where user stories will be passed to the LLM, an acknowledgment notice will be displayed.

The Data Layer focuses on managing data storage and retrieval, ensuring data is cleaned and structured appropriately for further processing. Its goals include handling data storage using local CSV files and preparing data through preprocessing for the processing layer. The functions involve managing data storage and performing preprocessing tasks to clean and prepare the data. Components of this layer include local CSV files (for ECCOLA questions, JIRA user stories, and training samples) and data preprocessing mechanisms.

In the Processing Layer, the system retrieves data from JIRA using an API, cleans and prepares this data for model training and classification, and applies the ECCOLA model to classify data and suggest ethical considerations. The goals are centered on data preparation and cleaning, along with ethically classify user stories based on the ECCOLA model. Functions of this layer include handling JIRA data retrieval or manual data retrieval, data cleaning, and processing, as well as using the ECCOLA model for classification and ethical recommendations. The key components are the JIRA Data Retrieval Module, Data Cleaning and Preparation Module, and ECCOLA Classification Module.

The Model Layer integrates the GPT-3.5 API for model training, evaluation, and validating. Its goals include integrating with the LLM model (GPT-3.5 API), fine-tuning LLM model, and evaluating it. Functions in this layer are aimed at fine-tuning the model using ECCOLA questions and training data, evaluating the model's performance, and using the trained model for classifying new user stories and generating ethical recommendations. Components consist of the GPT-3.5 API Integration, LLM-Model Training Module, and LLM-Model Evaluation mechanisms.

Finally, the Output Layer is responsible for outputting classification results and ethical recommendations, logging actions, and providing detailed reports. The primary goal is to offer easily-followed ethical suggestions for IT practitioners. The ethical suggestion is outputted from ethical recommendations derived from ECCOLA. For future enhancements, we suggest adding customized ethical recommendations generated by the LLM-model. This layer also maintains logs and provides detailed reports for users. Components of this layer include sections for Classification Results, Ethical Recommendations, and Logs and Reports.

4.6 Design Decision

In the process of developing the prototype, we encountered numerous pivotal moments that required careful and strategic decision-making. These decisions were essential not only to address immediate technical challenges but also to align the project with its overall objectives.

We have tried or learned from previous studies several approaches for each of the challenges. So, in this chapter, we documented both the successful strategies and the ones that failed, along with the reasoning. This comprehensive record aims to provide valuable insights, enabling future researchers to avoid redundant efforts and build upon our findings effectively.

Table 7, Table 8, and Table 9 are made to document how each option (approach) is assessed and why certain decisions are made. The tables are written by following these standard: The table is divided into sections that address specific concerns, list ranking criteria, and provide options. The first column, titled ”Concern (Identifier: Description),” identifies the ethical concern being addressed. The second column, ”Ranking criteria (Identifier: Name),” lists the criteria used to evaluate the options. The third column, ”Options,” details the possible solutions, each identified by a unique name and description, followed by their evaluation against the ranking criteria. The ”Status” row indicates whether the option is accepted or rejected. The accepted options will be applied into the System Architecture (Section. 4.5) of this prototype. The ”Rationale of decision” provides the reasoning behind each decision based on the evaluation against the criteria.

4.6.1 Ethic Baseline Model

This subsection compares several manual ethical model that potentially be the baseline ethical model of the prototype (Table 7). The ECCOLA Card (Vakkuri, Kemell, Jantunen, Halme, et al. 2021 and Halme et al. 2024) is chosen as the baseline for ethical questions used to build the prototype, due to its peer-reviewed, open-source guidelines specifically optimized for this field. These guidelines have been revised and published multiple times in peer-reviewed papers, ensuring they remain rigorous and credible. They encompass comprehensive ethical considerations that foster transparent and accountable decision-making processes (Table 7).

Table 7: Ethic Baseline Model

Concern (Identifier: Description)		*Con#1: Where do the ethical questions come from?*
Ranking criteria (Identifier: Name)		*Cr#1: Peer-reviewed and proven* *Cr#2: Open source* *Cr#3: Connected to Agile*
Options	***Identifier: Name***	***Con#1-Opt#1: UKRI***
	Description	https://www.ukri.org/publications/ethics-
	Status	*This option is **REJECTED**.*
	Relationship(s)	
	Evaluation	*Cr#1: Unclear whether peer-reviewed and proven* *Cr#2: -* *Cr#3: Not specific to Agile project management*
	Rationale of decision	*The framework is not published in white source*
	Identifier: Name	***Con#1-Opt#2: AMC COC***
	Description	https://www.acm.org/code-of-ethics
	Status	*This option is **REJECTED**.*
	Relationship(s)	
	Evaluation	*Cr#1: -* *Cr#2: -* *Cr#3: Not spesific for agile project management*
	Rationale of decision	*The framework is a bit unclear and did not come with a direct list of questions.*
	Identifier: Name	***Con#1-Opt#3: ECCOLA***
	Description	*Based on ECCOLA Card of Ville Vakkuri et al.*
	Status	*This option is **ACCEPTED**.*
	Relationship(s)	-
	Evaluation	*Cr#1: Peer-reviewed and proven* *Cr#2: Open source* *Cr#3: Optimized for software development management*
	Rationale of decision	*Fulfills all criteria*

4.6.2 Frontend and Backend Placement Configuration for Prototype Deployment

This subsection discusses the selection and justification of the environment where the transformation, front-end and back-end will be placed (Table 8).

The selected option for the end-user placement of the logic/server is a third-party LLM (Language Learning Model) with Python handling both the frontend and backend on the client desktop. This approach is resulting on a lightweight frontend and backend capable of running on a standard client PC while leveraging a third-party server for LLM suggestions.

This option is accepted for the following reasons. First, this option does not require cloud infrastructure, making it cost-effective and manageable. Second, as this option run in same machine aside from the LLM, this option is simplifying integration. Finally, for accessibility, the option offers an interactive user experience with quick processing times due to local execution.

Even though, the setup process requires some Python knowledge to run the user interface, but given that the target users are part of a software project management team with a high level of technical expertise, this criterion is of lower importance.

However, for data security, there is a potential risk of breaching non-disclosure agreements due to data transmission to the LLM server even though we have study the data privacy from the LLM provider (OpenAI n.d.(d) OpenAI 2023 OpenAI n.d.(c) OpenAI n.d.(d)). Despite this, since all components except the LLM operate locally, future enhancements could easily transition to a fully local setup, thereby minimizing technical debt of privacy protection.

Although the option does not fully satisfy all the criterias, the associated risks are deemed manageable. The potential breach of data security can be mitigated in future iterations, and the minor Python knowledge required is considered acceptable given the technical proficiency of the users.

Table 8: Frontend and Backend Placement Configuration for Prototype Deployment

Concern (Identifier: Description)		*Con#2: Which configuration of frontend and backend components should be used for the deployment of the prototype?*
Ranking criteria (Identifier: Name)		*Cr#1: Resource availability* *Cr#2: Data security* *Cr#3: Ease of connection with other components* *Cr#4: Ease of setup* *Cr#5: Accessibility*
Options	***Identifier: Name***	***Con#2-Opt#1: Python backend in server | Web frontend***
	Description	*Setting up a server with Python transformation and backend, and a web frontend. Clients can provide the API of their JIRA to the web.*
	Status	*This option is* ***REJECTED****.*
	Relationship(s)	
	Evaluation	*Cr#1: Cloud* *Cr#2: Might not align with NDA* *Cr#3: As the processing is on our side, it is easy to be accessed by other components* *Cr#4: Requires setting up the server again* *Cr#5: Easier for clients to access as they don't need to install any extra software*
	Rationale of decision	*This option does not fulfil the Cr#2, Cr#3, and Cr#4.*

Options	Identifier: Name	***Con#2-Opt#2: Python desktop app on client***
	Description	*Provide the open-source code to be run directly on the client's laptop. This option is feasible as the client consists of software development teams, so they should be able to run a Python script.*
	Status	*This option is **REJECTED**.*
	Relationship(s)	
	Evaluation	*Cr#1:Cloud* *Cr#2:Better Data Security* *Cr#3:Need internet to connect with other components* *Cr#4:No need server. But need to explain to user how to do it* *Cr#5:Harder accessibility, as we expect client to run the apps by themselves*
	Rationale of decision	*Con #1-Opt #2 has better data security. However, the available training data is insufficient to retrain a fast LLM that is light enough to run on a typical client PC.*
Options	Identifier: Name	***Con#2-Opt#3: Transformation, backend, and frontend all compacted in a Python desktop app on our server***
	Description	*Clients will provide us access to their JIRA API or send a CSV file containing their user stories, and we will send back the report.*
	Status	*This option is **REJECTED**.*
	Relationship(s)	
	Evaluation	*Cr#1: Cloud* *Cr#2: Might breach NDA, but the data is processed in another party's LLM* *Cr#3: Easier to connect with other components as it runs in our system* *Cr#4: Easier to set up* *Cr#5: Harder to access and less interactive*
	Rationale of decision	*This option does not fulfill Cr#2, Cr#3, Cr#4, and Cr#5*
Options	Identifier: Name	***Con#2-Opt#4: Transformation with third-party LLM, frontend and backend developed in Python and run on the client desktop***
	Description	*Create a lightweight frontend and backend that can run on a typical client PC and perform LLM suggestions on a third-party server.*
	Status	*This option is **ACCEPTED**.*
	Relationship(s)	
	Evaluation	*Cr#1: No need for cloud* *Cr#2: Might breach NDA due to passing data to LLM. However, since all components besides LLM run locally, it will be easier to move fully local in future enhancements.* *Cr#3: Easier to connect with other components as all besides LLM run on the same machine* *Cr#4: Requires a bit of Python knowledge to run the UI* *Cr#5: Interactive with UX, and the process will be fast as it runs locally*
	Rationale of decision	*This option does not fulfill Cr#2 and Cr#4. However, for Cr#2, it will be the smallest technical debt. For Cr#4, only basic Python knowledge is needed to run the application, and since the client for this prototype is a software project management team with higher technical knowledge, this criterion is not of high importance.*

4.6.3 Transformation algorithm

This subsection compares the transformation algorithms that we have tested and evaluates their suitability for achieving the prototype's goal (Table 9).

The first algorithms we considered were ANN Poplazarova et al. 2020. ANN should be more accurate as it has been validated by a big company, based on the paper ”Ethical decision-making in biopharmaceutical research and development applying values using the TRIP/TIPP model". The model is accessible via API and has been previously proven. But, it is only work for certain context.

Secondly, sentiment analysis, it can be done locally. But, it has no proof of accuracy and has not been previously proven. It does not have pretrained models and the technique is quite old as well.

Next, we tried several Large-Language Model (LLM). The script we used for comparison are all available on the Github(cinapr 2024b).

The first LLM is GPT4ALL. It utilizes several HuggingFace pre-trained models as the base. The models were chosen based on their performance recommendations on the Hugging Face website. The inclusion criteria included models that were in English, lightweight, and capable of running on the device. Therefore, four models were selected:

```
mistral-7b-openorca.gguf2.Q4_0.gguf (424.12 seconds)
Nous-Hermes-2-Mistral-7B-DPO.Q4_0.gguf (544.64 seconds)
orca-mini-3b-gguf2-q4_0.gguf (152.23 seconds)
replit-code-v1_5-3b-newbpe-q4_0.gguf (0.96 seconds)
```

Due to the limited availability of device resources, instead of fine-tuning and comparing them, we first asked a simple question to compare how quickly each model could respond. When asked, *"What is London?"*, the `orca-mini-3b-gguf2-q4_0.gguf` model responded in 152.23 seconds, while `Nous-Hermes-2-Mistral-7B-DPO.Q4_0.gguf` and `mistral-7b-openorca.gguf2.Q4_0.gguf` responded in 544.64 seconds and 424.12 seconds, respectively. Unfortunately, the `replit-code-v1_5-3b-newbpe-q4_0.gguf` model was unloadable due to the wrong number of tensors.

Even though this approach was considered the best for data privacy concerns since it runs locally, it was deemed unsuitable due to the excessive time taken to answer basic questions. The fastest response took more than two minutes, which is impractical for our prototype that needs to loop through numerous user stories.

Next, we compared GPT2. GPT2 is run on a local server. It took 1405.1124 seconds to be trained, and the classification compared to the ECCOLA sample dataset returned a 3.61% similarity. On average, the LLM took 0.195 seconds to classify each user story sample, showing better speed performance than GPT4ALL. Unfortunately, its reasoning is still weak, and future fine-tuning is not feasible as the basic fine-tuning already consumes more than 80% of the CPU and RAM of the available machine (Figure 5).

The next algorithm we compared, after fine-tuning, was BART. The base model for BART is not located on the client machine. It took 2059.9524 seconds to be trained, and the clas-

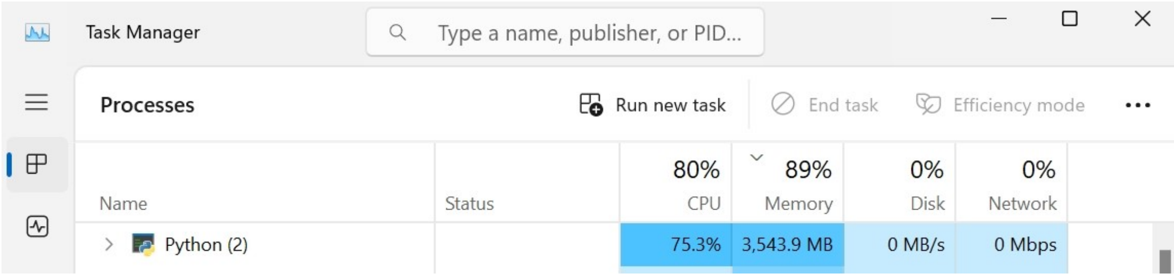

Figure 5: Task Manager Screenshot Showing GPT2 Fine-Tuning Consumption

sification compared to the ECCOLA sample dataset returned a 7.22% similarity. It showed better speed performance than GPT4ALL. Unfortunately, its reasoning remains weak, and future fine-tuning is not possible since the basic fine-tuning consumes more than 90% of the CPU and RAM of the available machine (Figure 6). Moreover, compared to GPT2, BART still falls short as the base model is hosted on an external server.

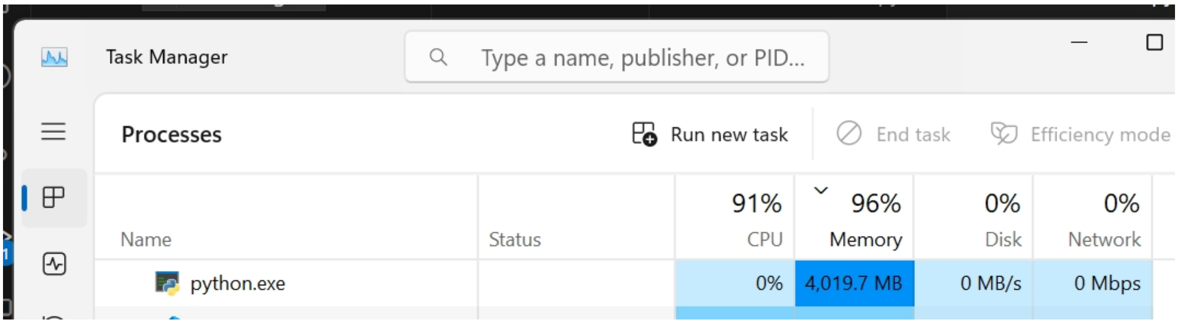

Figure 6: Task manager screenshot that showing BART fine-tuning consumption

The fourth LLM algorithm we compared was GPT-3.5, which runs on a server owned by OpenAI. It took 963.47 seconds to be trained, and the classification compared to the ECCOLA sample dataset returned 26.32%. Although it lacks in terms of privacy, as it runs on an OpenAI server, we have reviewed the privacy policies provided by OpenAI (OpenAI n.d.(b)OpenAI 2023OpenAI n.d.(c)OpenAI n.d.(d)). Since the LLM process does not run on a local machine, this algorithm offers the best performance in terms of speed and minimal CPU and RAM consumption. Therefore, better reasoning can be expected from this algorithm.

From the perspective of machine availability, speed, and project timeline, we chose GPT-3.5, which runs on an OpenAI server. However, we recommend that future researchers with better machine or server availability consider deeper fine-tuning of GPT2 to improve its reasoning skills. GPT2, which runs locally, would be better for user story data privacy.

Table 9: Transformation Algorithm

Concern (Identifier: Description)		*Con#3: Which transformation algorithm should be chosen for generating the suggestion?*
Ranking criteria (Identifier: Name)		*Cr#1: Speed of response* *Cr#2: Data privacy* *Cr#3: Accuracy of classification* *Cr#4: Resource usage* *Cr#5: Previously proven* *Cr#6: Model is openly distributed*
Options	*Identifier: Name*	***Con#3-Opt#1: GPT-4ALL with HuggingFace Pre-trained Models***
	Description	*Using GPT-4ALL with several pre-trained models from HuggingFace based on performance recommendations. It is an open-source alternative to OpenAI GPT.*
	Status	*This option is* ***REJECTED****.*
	Relationship(s)	
	Evaluation	*Cr#1: Response times varied from 152.23 to 544.64 seconds, which is too slow for practical use.* *Cr#2: Ensures data privacy as processing is local.* *Cr#3: Accuracy not thoroughly evaluated due to response time limitations.* *Cr#4: Requires significant resources to run on local machines.* *Cr#5: Not previously proven in similar contexts.* *Cr#6: Models are openly distributed.*
	Rationale of decision	*This option, while excellent for data privacy and openness, is unsuitable due to slow response times and high resource usage, making it impractical for the prototype's needs.*
Options	*Identifier: Name*	***Con#3-Opt#2: GPT2***
	Description	*Testing GPT-2 for performance and classification accuracy.*
	Status	*This option is* ***REJECTED****.*
	Relationship(s)	
	Evaluation	*Cr#1: Faster response time compared to GPT-4ALL models.* *Cr#2: Limited by local machine capabilities, impacting fine-tuning.* *Cr#3: -* *Cr#4: High resource consumption for fine-tuning.* *Cr#5: Proven in various NLP tasks but not specifically for this use case.* *Cr#6: Openly distributed.*
	Rationale of decision	*Despite better speed and openness, the high resource consumption prevents it from being further fine-tuned.*
Options	*Identifier: Name*	***Con#3-Opt#3: BART***
	Description	*Testing BART model for performance and classification accuracy.*
	Status	*This option is* ***REJECTED****.*
	Relationship(s)	
	Evaluation	*Cr#1: Slow training time but better response time.* *Cr#2: Depends on an external server, affecting data privacy.* *Cr#3: Higher accuracy compared to GPT-2.* *Cr#4: High resource usage.* *Cr#5: Proven in various contexts.* *Cr#6: Openly distributed, with APIs that run on HuggingFace server and downloadable models that run locally.*
	Rationale of decision	*Better accuracy and proven effectiveness but still impractical due to server dependence and high resource usage for local fine-tuning.*

Options	Identifier: Name	**Con#3-Opt#4: GPT-3.5 (API)**
	Description	Testing the GPT-3.5 model for performance and classification accuracy. Based on the paper: "A Survey on Large Language Model-Based Autonomous Agents."
	Status	This option is **ACCEPTED**.
	Relationship(s)	
	Evaluation	Cr#1: Fast response time suitable for the prototype. Cr#2: Data privacy concerns due to server processing. Cr#3: Expected better reasoning and accuracy. Cr#4: Low local resource consumption. Cr#5: Proven effective in various applications. Cr#6: Not openly distributed.
	Rationale of decision	Despite data privacy concerns, the fast response time, low resource usage, and proven effectiveness make GPT-3.5 the best option for this prototype.
Options	Identifier: Name	**Con#3-Opt#5: ANN**
	Description	Based on the paper: "Ethical Decision-Making in Biopharmaceutical Research and Development Applying Values Using the TRIP TIPP Model."
	Status	This option is **REJECTED**.
	Relationship(s)	
	Evaluation	Cr#1: Expected to be accurate but no direct speed measurements. Cr#2: Requires external uploading, impacting data privacy. Cr#3: Proven in biopharmaceutical research contexts. Cr#4: Requires API access. Cr#5: Proven effective in specific contexts. Cr#6: Not openly distributed.
	Rationale of decision	While it shows promise in accuracy, it has been proven only in certain contexts.
Options	Identifier: Name	**Con#3-Opt#6: Sentiment Analysis**
	Description	Sentiment analysis using traditional techniques.
	Status	This option is **ACCEPTED**.
	Relationship(s)	
	Evaluation	Cr#1: No proof of speed and accuracy. Cr#2: Ensures data privacy as it can be done locally. Cr#3: Not proven in relevant contexts. Cr#4: Low resource requirements. Cr#5: Not previously proven in relevant contexts. Cr#6: No pre-trained models found.
	Rationale of decision	While data privacy and resource usage are advantages, the lack of proven accuracy and the outdated technique make it less desirable.

4.7 Prototype Setup and Documentation

This section provides a visual overview of the prototype through a series of screenshots. These images illustrate the various functionalities and features of the system.

1. Clone the prototype from Github(cinapr 2024a)

   `git clone https://github.com/cinapr/EthicalAutomatedSuggestion.git`

2. Retrieve CSV files containing ECCOLA questions from `DOWNLOAD ECCOLA QUESTIONS` menu (Figure 7).

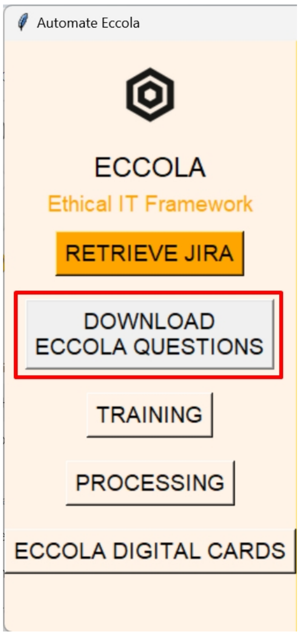

Figure 7: Retrieving ECCOLA Questions

3. Prepare user stories as the personalized information for the software projects to be checked. This can be done automatically from project management tools, such as JIRA (Figure 8) in `RETRIEVE JIRA` menu, or manually by typing them into the template `Template_For_UserStories.csv`, which is included on the clone repository of GitHub (cinapr 2024a) in Step 1. To retrieve the user stories automatically, users need to have JIRA API credentials and input those credentials into the prototype in the relevant input fields.

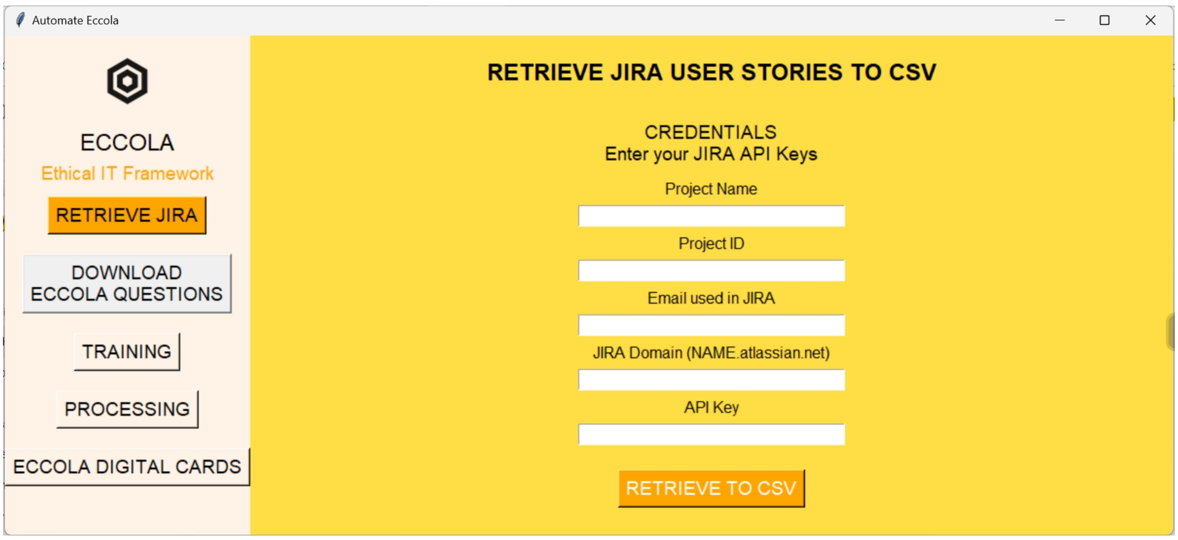

Figure 8: Retrieving JIRA User Stories

4. Before running the suggestions, users should have the model ID of GPT-3.5, which can be acquired in the `TRAINING` menu (Figure 9 and Figure 10). Then, users can upload the ECCOLA Questions from Step 2 to the prototype.

   The file `ECCOLA_Questions.csv` should be uploaded using the `UPLOAD ECCOLA QUESTIONS` button, and `ECCOLA_QuestionsCorrelations.csv` using the `UPLOAD ECCOLA DEFINITION` button, and `TrainingSet.csv` using the `UPLOAD TRAINING SAMPLES` button. Users can adjust the training dataset themselves in `TrainingSet.csv` as well.

   Users can adjust the content of the CSV files. For example, they can delete any questions that are irrelevant to the project being checked (Section **??**). This can also serve as a feedback loop. When users find any irrelevant samples or want to add new samples, they can make these changes directly in the CSV files and re-fine-tune the LLM model.

   It should be noted that users need to acknowledge and give permission before they can proceed with this step, as the data is processed by a third-party LLM (Subsection 4.4.2).

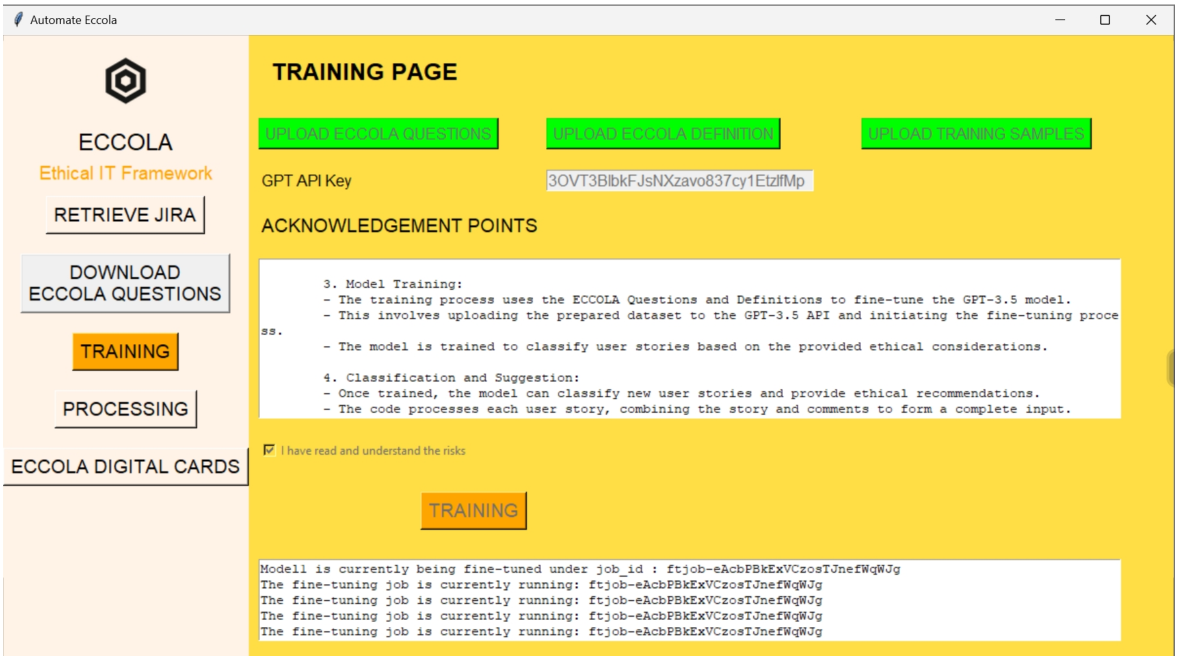

Figure 9: Start to fine-tuning GPT3.5 Model to support ethical question and classification according to ECCOLA

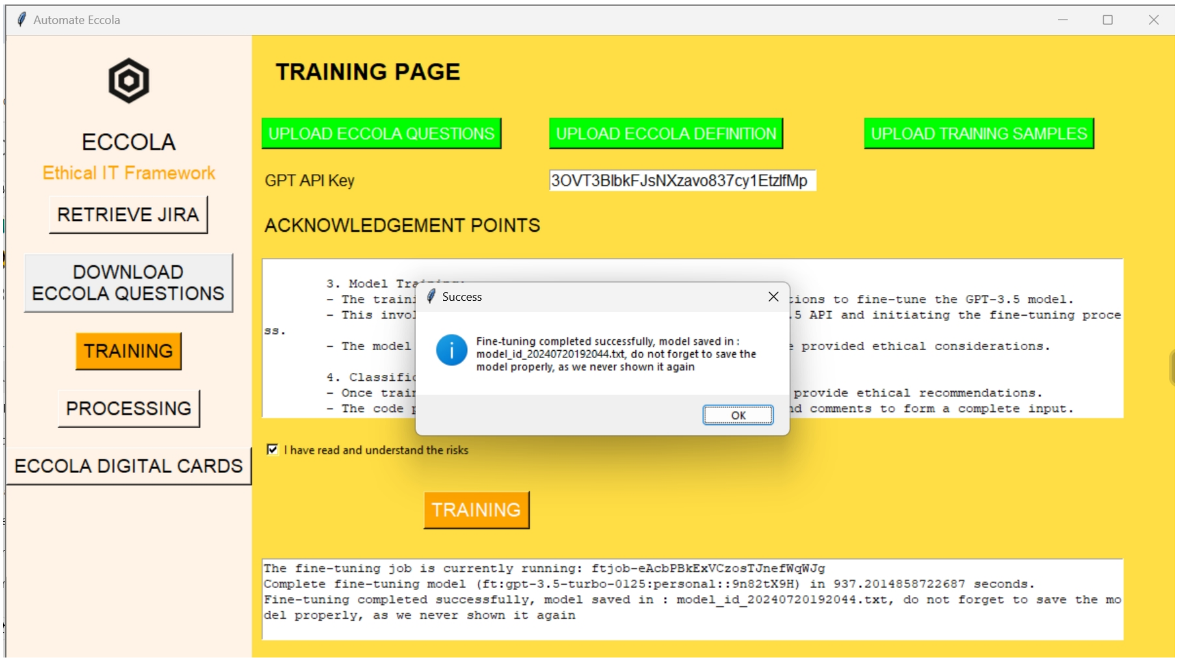

Figure 10: Finish fine-tuning GPT3.5 Model to support ethical question and classification according to ECCOLA

5. Every user story need to be pre-processed before it can receive ethical suggestions. This can be done in the `PROCESSING` menu.
   Users should enter the GPT-3.5 model ID from Step 4 into the `Trained GPT Model ID` input field, and GPT API Credentials into the `GPT API Key` input field. It should be noted that users need to acknowledge and give permission before they can proceed with this step, as the data is processed by a third-party LLM (Subsection 4.4.2) (Figure 11).

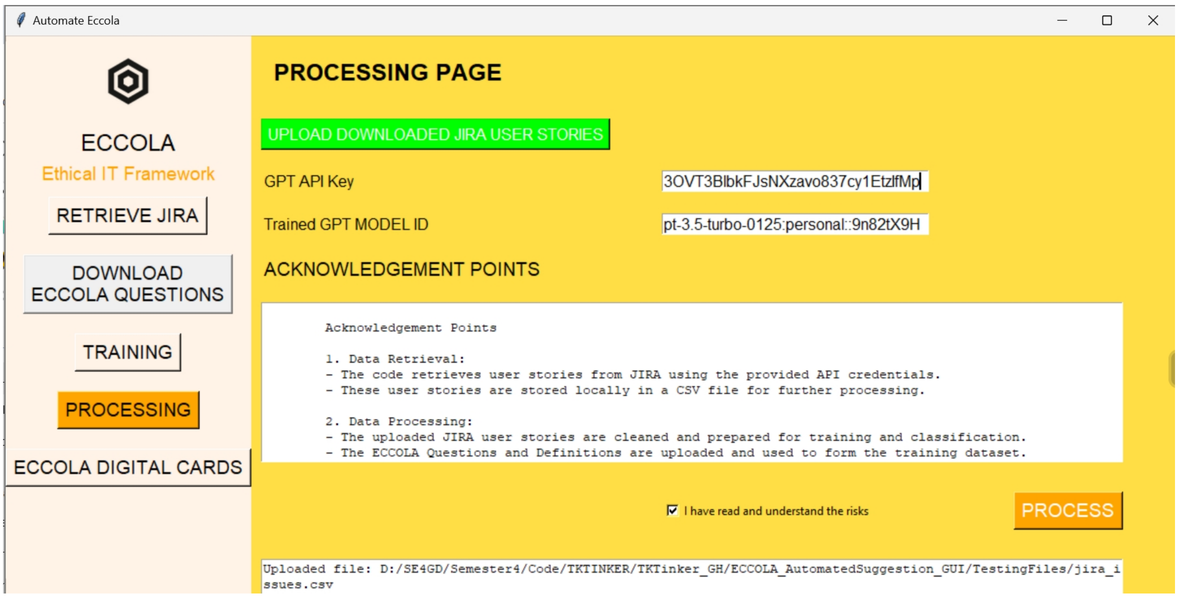

Figure 11: Uploading user story CSV for Pre-processing

While the prototype run the pre-processing, users can see the process from the information textbox (Figure 12).

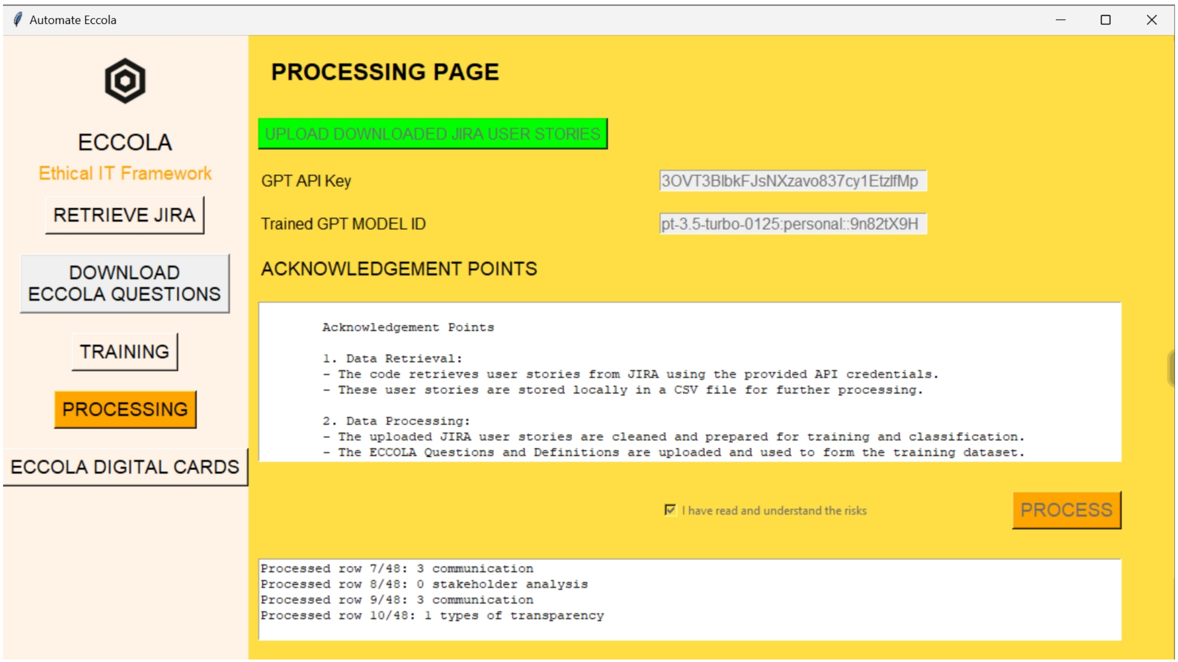

Figure 12: Pre-processing user stories is running

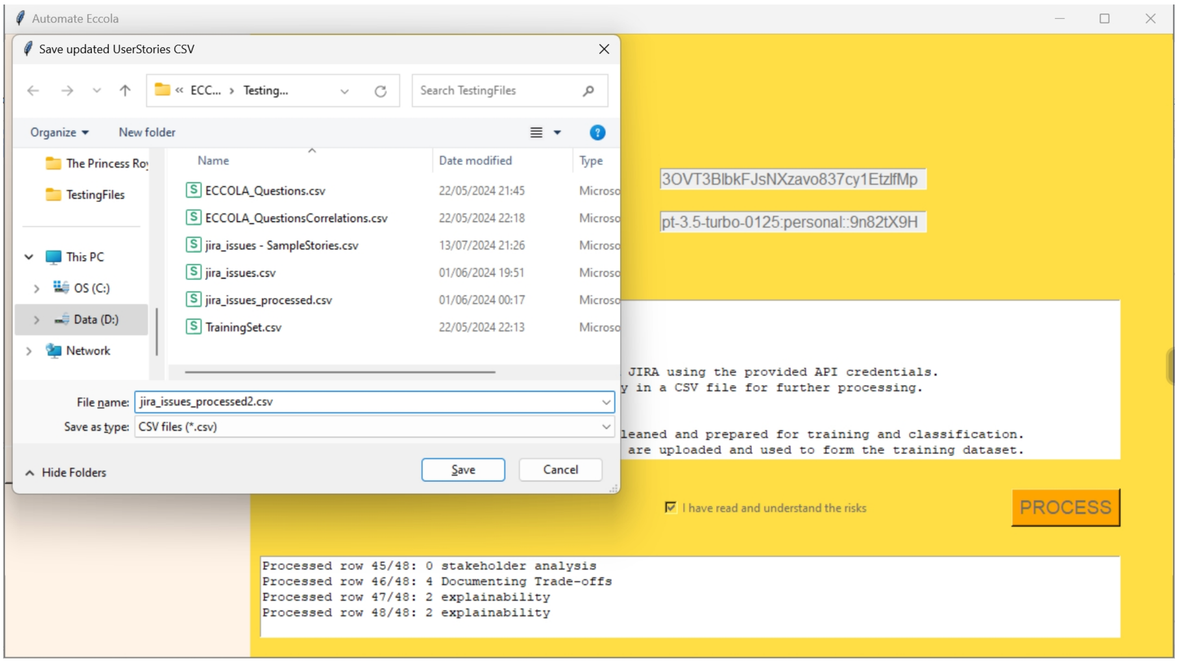

Figure 13: Saving the pre-processed user stories to CSV

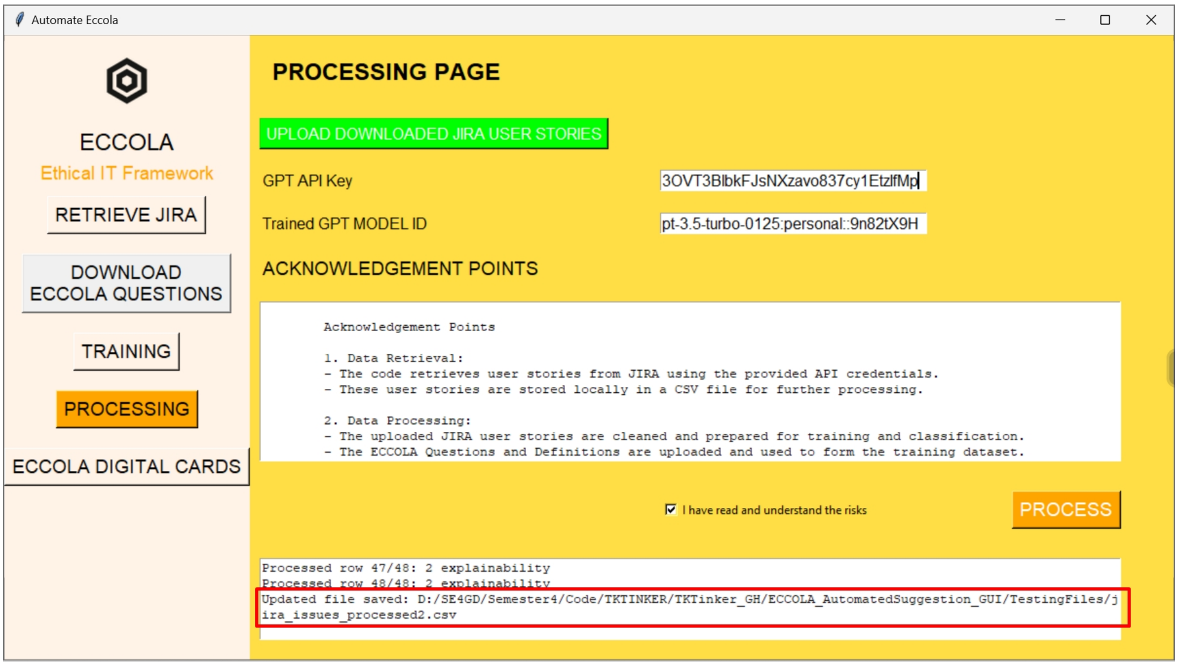

Figure 14: The pre-processed user stories has been saved

After the pre-processing done, user can save the pre-processed user stories into a CSV (Figure 13 and Figure 14).

6. Once the pre-processed CSV is acquired, users can start receiving ethical suggestions for the software project. The user stories are uploaded in the menu `ECCOLA DIGITAL CARDS`. `ECCOLA DIGITAL CARDS` imitates the workflow of ECCOLA (Section **??**), where relevant ethical dimension cards are shown in the textbox. The ethical dimensions are displayed one-by-one, and to move to the next ethical dimension, users can click the `NEXT QUESTION` button. Users can see which ethical concerns are important for each user story. Additionally, suggestions to expand the ethical standards of the user stories are provided as well (Figure 15).

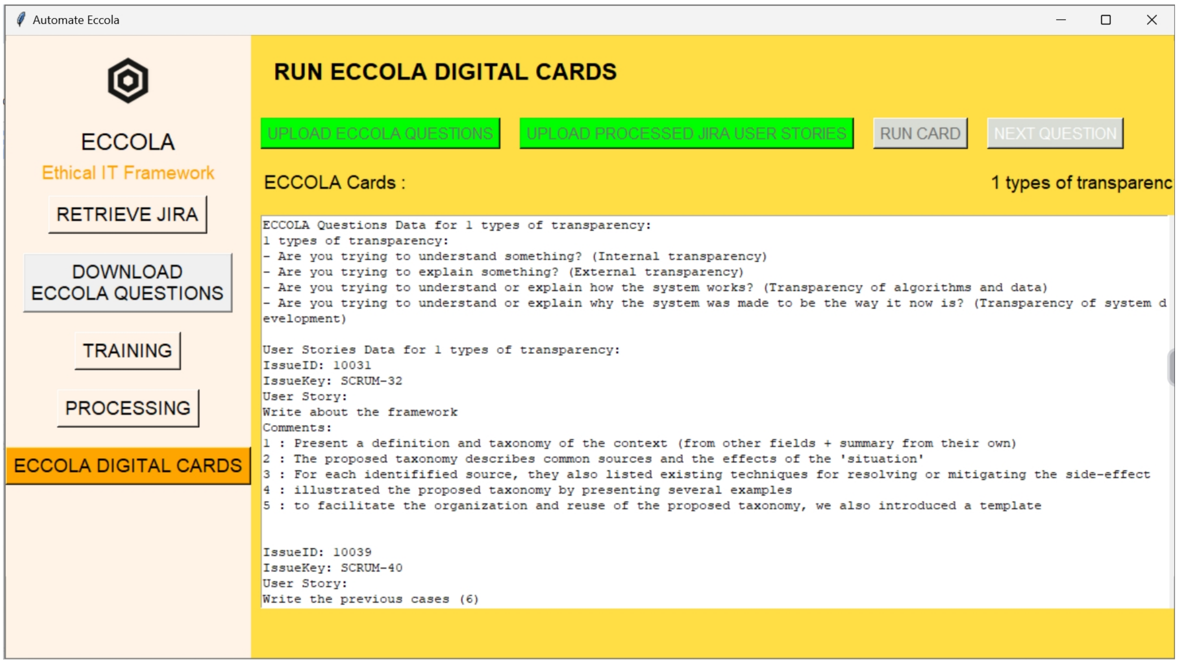

Figure 15: Ethical Suggestions process for software project management

4.8 Case Study for Testing

This section details the process and results of using case studies to test the prototype developed for automating ethical suggestions in software project management.

For the implementation phase, we employed case studies in the form of user stories for training, validating the training, and testing the prototype. The user stories were acquired from several sources, such as Halme et al. 2024 (`https://figshare.com/s/723d290fd27ec98fa139`), JIRA account from student projects, and other valuable contributors that we cannot appreciate one by one.

Initially, we randomly chose 367 case studies from the available data, which consisted of 77 from JIRA and the rest from manual CSV files. Unfortunately, 61 user stories were poorly written, containing only titles without descriptions to be checked, so we ended up with 306 case studies to work on.

For the training and validation phases, we exclusively used data in CSV format. Meanwhile, the user stories employed for testing the prototype were sourced from both CSV files and directly retrieved from JIRA API. We allocated 195 user stories for training, 41 user stories for validating by comparing the classification result with manually annotated ECCOLA types, and 70 user stories for testing the prototype flow in generating suggestions. Table 10 illustrates some of the samples we utilized for testing the prototype.

In SmartHomeAutomation-2 (Table 10), where an energy manager wants to develop an automated energy management system for smart homes so that users can optimize their energy consumption and reduce utility costs, by monitoring energy usage. This feature might raise ethical concerns about privacy and surveillance, as continuous monitoring could infringe on the privacy of individuals within and around the property. There also risk of misusing the data by the third party along with the possibility that the data got intercepted in the network. **From this sample, we discovered that a requirement recorded in user stories might be interconnected with different ethical dimensions(Radford University 2020), but the prototype currently can only give suggestion based on one dimension.**

Another sample is OnlineLearningPlatform-3 (Table 10), where an academic instructor wants to automate grading system to evaluate student assignments. This user story presents concern of bias in the grading algorithm and the possibility of complaints when the students and educators do not understand how grades are assigned. There is privacy issue as well, where the name and personal data of the students might get intercepted on the process. **The prototype, which currently can only give suggestion for one dimension choose to give suggestion for privacy issue, instead of the more pressing human factor.**

Table 10: Case Studies[1]

IssueKey	User_Story	Possible ethical dimension 1	Possible ethical dimension 2	Possible ethical dimension 3	ECCOLA Code from prototype
Health Monitoring Application -1	Continuous Health Data Collection. As a software developer, I want to continuously collect health data from users' wearable devices so that we can monitor their health in real-time and provide timely alerts. This will help users stay informed about their health status and take proactive measures to prevent potential health issues.	#7 Privacy and Data	#8 Data Quality	#12 System Security	5 Traceability
Health Monitoring Application -2	Predictive Health Analytics As a data scientist, I want to analyze collected health data to predict potential health issues so that we can provide users with early warnings and personalized health advice. By leveraging predictive analytics, we can help users maintain better health and reduce the risk of serious health problems.	#7 Privacy and Data	#8 Data Quality	#2 Explainability	8 data quality
Health Monitoring Application -3	Personalized Health Recommendations As a health coach, I want to provide personalized health recommendations to users based on their collected data so that they can make informed decisions about their health and lifestyle. By offering tailored advice, we can help users achieve their health goals more effectively.	15 Stakeholder Participation	#7 Privacy and Data	#2 Explainability	15 Stakeholder Participation
Health Monitoring Application -4	Sharing Health Data with Third Parties As a product manager, I want to enable users to share their health data with third-party health providers so that they can receive comprehensive health services. By allowing data sharing, users can benefit from a more holistic approach to health management	#7 Privacy and Data	#9 Access to Data	#19 Ability to Redress	5 Traceability
Health Monitoring Application -5	Health Data Anonymization As a security engineer, I want to anonymize users' health data before storing it so that their privacy is protected while allowing us to perform analytics. Anonymization will help prevent the identification of individual users from the data, ensuring their personal information remains confidential.	#7 Privacy and Data	#8 Data Quality	12 system security	12 system security
Smart Home Automation -1	Voice-Activated Controls As a software developer, I want to implement voice-activated controls for smart home devices so that users can conveniently manage their home environment using voice commands. This feature will enhance user experience by providing a hands-free way to control various devices.	#14 Accessibility	#7 Privacy and Data	#12 System Security	14 Accessibility
Smart Home Automation -2	**Automated Energy Management As an energy manager, I want to develop an automated energy management system for smart homes so that users can optimize their energy consumption and reduce utility costs. The system will monitor energy usage and make adjustments to improve efficiency.**	**#7 Privacy and Data**	**#12 System Security**	**#16 Environmental Impact**	**7 Privacy and Data**

Smart Home Automation -3	Smart Security Surveillance As a security engineer, I want to implement a smart surveillance system for home security so that users can monitor their property in real-time and receive alerts about suspicious activities. The system will use AI to analyze video feeds and detect potential threats.	#7 Privacy and Data	#12 System Security	#2 Explainability	5 Traceability
Smart Home Automation -4	Personalized Home Automation As a smart home developer, I want to create personalized automation routines based on user behavior so that the smart home system can adapt to their preferences and habits. This will provide a more customized and convenient living experience.	#2 Explainability	#7 Privacy and Data	#10 Human Agency	8 data quality
Smart Home Automation -5	Data Sharing with Smart Grid As a smart home product manager, I want to enable data sharing with the local smart grid so that users can participate in demand response programs and benefit from optimized energy distribution. This will help balance energy supply and demand.	#7 Privacy and Data	#9 Access to Data	#16 Environmental Impact	**9 Access to Data**
Online Learning Platform -1	Personalized Learning Paths As an educational software developer, I want to implement personalized learning paths for students so that they can receive tailored content and assessments based on their individual needs and progress. This will help improve learning outcomes by addressing each student's unique strengths and weaknesses.	#10 Human Agency	#2 Explainability	#7 Privacy and Data	8 data quality
Online Learning Platform -2	Student Performance Analytics As a data scientist, I want to analyze student performance data to identify trends and areas for improvement so that educators can make data-driven decisions to enhance the learning experience. By understanding student performance, we can provide targeted support and resources.	#2 Explainability	#7 Privacy and Data	#8 Data Quality	4 Documenting trade-offs
Online Learning Platform -3	Automated Grading System As an instructor, I want to implement an automated grading system to evaluate student assignments so that grading can be done more efficiently and consistently. This will save time and provide quicker feedback to students.	#10 Human Agency	7 Privacy and Data	#18 Auditability	7 Privacy and Data
Online Learning Platform -4	As a platform administrator, I want to implement AI moderation in virtual classrooms to monitor and manage student interactions so that the learning environment remains safe and conducive to learning. The AI will detect and address inappropriate behavior in real-time.	#10 Human Agency	#7 Privacy and Data	#2 Explainability	6 System Reliability
Online Learning Platform -5	Learning Analytics Dashboard for Educators As an educational researcher, I want to develop a learning analytics dashboard for educators so that they can visualize and analyze student performance data to improve teaching strategies. The dashboard will provide insights into student engagement, progress, and areas for improvement.	#10 Human Agency	#7 Privacy and Data	3 communication	3 communication

4.9 Study Replicability

The prototype codes for this research can be found in the replication package uploaded on Github (cinapr 2024a). Meanwhile, the scripts for comparing the possible algorithms (Subsection 4.6.3) are provided in Github (cinapr 2024b). The replication package can be used to replicating and expanding this project.

4.10 Summary of Implementation

In conclusion, this chapter presents the findings from the development and implementation of a prototype aimed at automating suggestion of ethical implementation method on agile software project management.

The prototype developed integrates the ECCOLA model to automate ethical suggestions in agile software projects. Root from previous researchers, such as Hingle et al. 2023 that promoting the usage of AI technology such as NLP and LLM, the prototype employs a Large Language Model (LLM) to process user stories. Then classify them based on predefined ethical codes, ECCOLA (Halme et al. 2024 and Vakkuri, Kemell, Jantunen, Halme, et al. 2021).

The prototype operates on a local client machine with a lightweight frontend and backend built using Python, and leverages third-party LLMs for natural language processing tasks, ensuring quick and accurate classification of user stories.

[1] Explanation of Table.10 Columns:
IssueKey: The identifier of the user story.
User_Story: The title and description of the user story.
Possible Ethical Dimension 1, 2, and 3: The manual classification of the ethical dimensions.
ECCOLA Code from Prototype: The classification of ECCOLA Code automatically generated by the suggestion prototype.

5 RESULTS AND DISCUSSION

This chapter presents the results of our study, followed by a comprehensive discussion that interprets these findings in the context of existing research, highlighting their implications, significance, and potential impact. We also outline the prospective directions for future research, address the limitations encountered during the study, and critically examine the threats to validity that may impact the findings of this project.

5.1 Results

In this results section, we provide comprehensive responses to the research questions outlined below. These responses are derived from an extensive literature review (Chapter 2), empirical insights gained during experimental research and implementation (Chapter 4), and analysis of the case study (Section 4.8).

5.1.1 Answer to Sub-RQ1

In which ways can ethical suggestion automation in agile software project management be carried out?
Ethics involve nuanced, context-dependent judgments that can vary significantly between situations and individuals (Gibson et al. 2022).

Ethical automation in agile software project management should act as suggestion, instead of final decision-maker. Because of the dynamic nature of ethics, fully automating ethical concerns is not ideal. Therefore, the automated tools will only act as suggestions instead of being enforced as final judgment.

Based on our observations, effective ethical judgment can be reached by a combination of empathic and rational deliberation, which automated systems cannot fully replicate. This is in line with the view of Gogoll et al., who stated that ethical deliberation is a skill learned through practice and cannot be reduced to following a checklist. Hence, automated tools are designed to provide suggestions from a rational point of view, while humans will act as the rational deliberation to ensure comprehensive ethical consideration.

To ensure users understand how the automation work, they will receive a page of consent (acknowledgement notice) that includes information on what kind of ethical questions are covered, where the logical reasoning comes from, how their data will be processed, and importantly, that the result is intended to act only as a suggestion. This transparency helps users understand the limitations and the advisory nature of the tool, ensuring they are equipped to make the final ethical decisions themselves .

5.1.2 Answer to Sub-RQ2

What kind of input is needed for the prototype to have sufficient information to make suggestions?

There are two inputs critical to be receive by the prototpye (Figure 1). First is the ethical baseline model. As the prototpye employing ECCOLA ethical model as its baseline, a list of ECCOLA cards in CSV format are expected to be added. The ECCOLA card list is included with the replicable package, allowing users to include only the cards relevant to their project.

The second aspect is the knowledge customized for each project processed by the prototype. This knowledge informs the prototype about the specific information unique to the uploaded project, distinguishing it from other projects. This information can be sourced from various places, such as manually listing them or exporting from agile tools like JIRA.

5.1.3 Answer to Sub-RQ3

What transformation algorithm is employed to enable the prototype to generate ethical suggestions?

The transformation algorithm employed to make the suggestions in the prototype is based on GPT-3.5, which runs on a server owned by OpenAI (Subsection 4.6.3). This algorithm was chosen because it offers the best performance in terms of speed and minimal CPU and RAM consumption compared to other models evaluated, such as GPT2, BART, and various models from GPT4ALL. Although GPT-3.5 lacks in terms of privacy since it runs on an external server, it was selected due to its superior reasoning capabilities and efficiency in processing time.

However, we recommend that future researchers with better machine or server availability consider deeper fine-tuning of GPT-2 to improve its reasoning skills. Since GPT-2 runs locally, it would be better for user story data privacy.

5.1.4 Answer to Sub-RQ4

What kinds of output formats are useful for the prototype to effectively deliver ethical suggestions?

From the user's experience point of view, there are several options for how the suggestions can be delivered (Subsection 4.6.2). Instead of a website, we chose to use a lightweight desktop application to ensure that all client data is fully processed locally, thereby ensuring data privacy. A lightweight desktop application was selected over a modern-style desktop application to enhance performance by cutting down on aesthetic features.

For the content, as we are employing ECCOLA, the prototype delivers suggestions following the ECCOLA flow (Section 2.3). The client receives a list of questions sequentially. These questions are annotated with information on which user stories that are relevant to them, how relevant they are, and guidance on how to expand the user stories further. However, we found that instead of listing each user story under the most relevant ECCOLA type, it is more beneficial to list a user story under all relevant ECCOLA types, as a user story can have more than one ethical concern.

5.1.5 Answer to Sub-RQ5

What kind of actions should be taken to address the privacy concerns of clients' private data?
To address the privacy concerns of clients' private data, four strategies are taken.

First, to minimize the risk of data breaches during transfer, the prototype is designed to execute operations on the client side as much as possible, thereby guaranteeing data security and privacy.

Second, an acknowledgment and consent section is implemented within the prototype that outlines the system's data processing flow. By informing users about how their data will be handled and processed, the prototype builds trust and ensures users are aware of privacy measures.

Third, the system provides clear documentation and rationalization for each decision made. This transparency aids in auditing the system's decisions and ensures alignment with ethical principles.

Finally, local LLM processing is recommended for future enhancements. While the prototype currently uses a third-party LLM due to machine limitations, moving the LLM process to a fully local setup will further enhance data protection by eliminating the need to share client data externally.

These measures collectively ensure that client data is protected, enhancing trust and maintaining the integrity of the system.

5.1.6 Answer to Sub-RQ6

How can the prototype be evaluated?
Ethics are inherently dynamic, leading to varying and potentially overlapping judgments

among individuals, which makes statistical comparison, commonly used as the standard for AI testing, not applicable in this scenario. Therefore, three main strategies are structured to evaluated the result of this project (Subsection 4.4.3).

First, we did consistency checks with the ethical baseline model to maintain alignment with established ethical principles. We ensured and documented that the system architecture of this prototype is inline with the flow of the baseline framework, ECCOLA.

Second, we provided detailed documentation and rationalization for each decision made by the system. This aids in auditing the system's decisions and ensures they align with ECCOLA principles, thus maintaining transparency and accountability.

Finally, we implemented a structured feedback loop to identify and address any logical inconsistencies or gaps in the decision-making process. This involves collecting the system's decisions, allowing users to give feedback regarding these decisions, and using this feedback to refine and improve the system.

Additionally, future researchers are recommended to explore techniques for statistical testing of emotion-dependent AI as a dedicated research project.

5.1.7 Answer to Main-RQ

To what extent can ethical decision-making processes in software development be automated?

With the current available technology of chosen transformation algorithm, ethical decision-making processes in software development can be automated only as suggestions, not as final decisions. The dynamic nature of ethics, which involves subjective and context-specific judgments, makes it unsuitable for full automation.

Automated tools can provide valuable insights and recommendations and are recommended to prevent human oversight, but they cannot replace the nuanced and empathetic judgment that human can offers. Therefore, while automation can support ethical decision-making by offering logical suggestions, the final decisions must remain with human practitioners to ensure comprehensive and contextually appropriate ethical deliberation.

5.2 Discussion

This study aimed to explore the feasibility of automating ethical decision-making processes in agile software project management by leveraging the ECCOLA ethical model and the GPT-3.5 language model. The developed prototype was designed to provide ethical suggestions based on user stories, offering a new approach to integrating ethics into software development. This discussion interprets the results, examines their implications, and addresses the strengths, limitations, and practical implications of the findings.

The findings of this study align with existing literature that underscores the importance of ethical considerations in software development. The use of the ECCOLA model as a baseline for ethical assessment builds on the methodologies proposed by Vakkuri, Kemell, Jantunen, Halme, et al. 2021 and Halme et al. 2024, providing a structured and comprehensive approach to ethical decision-making. The decision to use GPT-3.5 was driven by its superior performance in terms of speed and minimal CPU and RAM consumption, despite privacy concerns associated with running on an external server. This choice highlights the ongoing trade-offs between performance and data privacy in AI applications.

The prototype demonstrated that while full automation of ethical decision-making is not feasible due to the nuanced and context-dependent nature of ethics, AI can effectively provide valuable suggestions. These suggestions can aid human practitioners in making more informed and ethically sound decisions. This based on the view of Gogoll et al. 2021 that ethical deliberation is a skill that requires practice and cannot be fully automated.

5.2.1 Strengths and Limitations

One of the main strengths of the prototype is its a novel approach in project management ethical assessment. By automating the suggestion process, the prototype can help identify potential ethical issues in every stages of project life-cycle, by minimizing human oversight. The integration of ECCOLA ensures that the suggestions are grounded in a peer-reviewd ethical model, which enhances the credibility and reliability of the recommendations.

However, the study also has several limitations that arise from the decisions made during the research design (Section 4.6) or were discovered after testing (Section.4.8). The reliance on external LLMs like GPT-3.5 raises significant privacy concerns, as sensitive project data must be transmitted to and processed by an external server. While the use of GPT-3.5 was necessary due to resource constraints, it highlights the need for more secure and privacy-preserving AI solutions in future research. Additionally, the prototype's suggestions are currently limited to a single ethical dimension per user story, which may not fully capture the complexity of real-world ethical dilemmas.

5.2.2 Implications for Practice

The practical implications of these findings are significant for software project managers and developers. The prototype offers a valuable tool for integrating ethical considerations into the agile development process, promoting a more proactive approach to ethical risk management.

For software project managers, this tool can serve as a supplementary resource for ethical decision-making, ensuring that ethical considerations are consistently integrated into project workflows. Developers can use the prototype to gain a better understanding of the ethical implications of their work, fostering a culture of ethical awareness and responsibility within the development team.

In conclusion, this study demonstrates the potential of AI to support ethical decision-making in agile software project management. While fully automation of ethical judgments is not feasible, the developed prototype provides valuable insights and suggestions that can enhance the ethical quality of software projects. Future research should focus on improving the privacy and accuracy, as well as exploring ways to integrate multiple ethical dimensions into the suggestion process. By addressing these challenges, we can move closer to a more ethically conscious and responsible software development environment.

5.3 Limitations

In this section, we will discuss the inherent limitations that arise from the nature of the research method (Prototyping) and research object (Ethics in Information Technology).

This project focuses on prototyping rather than developing a full-fledged system, which limits the scope and depth of ethical analysis. In the current prototype, each user story can be given a suggestion for only one ethical concern, and recommendations are provided at a generic level.

From the point of view of ethical nature, ethics are inherently dynamic, leading to varying and potentially overlapping judgments among individuals. With the existing technology, this empathy-dependent automation prevents our prototype from being statistically evaluated. Therefore, we have incorporated a feedback loop into the prototype instead of statistical evaluation, due to project time and resource limitations. Future projects can include statistical testing, following the expansion of the LLM model beforehand (Section 5.5).

5.4 Threats to Validity

The project's validity has been assessed using the four classification types established by Campbell and Cook 1979, which include internal, external, construct, and conclusion validity.

5.4.1 Internal Validity

Internal validity refers to the extent to which the observed effects can be attributed to the experimental treatment or manipulation rather than to other factors. In this research, potential threats to internal validity include:

Bias in Data Collection:

There might be biases in the user stories and ethical concerns data used for fine-tuning. This could influence the findings and the effectiveness of the ethical suggestion prototype. Therefore, we would suggest for future researcher to expand the fine-tuning dataset (Section 5.5).

Algorithmic Limitations:

The chosen transformation algorithm (GPT-3.5) relies on a third-party server (OpenAI), which might introduce biases based on its training data and operational.

Testing with a Small Amount of Data:

The current study tested the prototype with a limited dataset, which might not capture the full variability and complexity of real-world data. Future research should include larger datasets to enhance the reliability and validity of the findings (Section 5.5).

5.4.2 External Validity

External validity refers to the extent to which the findings of the study can be generalized to other populations, settings and times. Potential threats to external validity in this research may include:

Ethical Context Variability:

Ethical considerations are highly context-dependent. The ethical suggestions generated might not be applicable or appropriate in different cultural, legal, or organizational settings. Because of this nature, we would suggest the automation acts only as suggestion not as final decision-maker.

5.4.3 Construct Validity

Construct validity refers to the degree to which a test or instrument accurately measures the theoretical construct it is intended to measure. For this project, construct validity would concern how well the prototype and its suggestions genuinely capture and reflect ethical considerations as defined by relevant ethical model, ECCOLA.

Definition of Ethical Concerns:

The way ethical concerns are defined and operationalized in the prototype might not capture the full complexity of ethical issues in software development. To mitigate this concern, we built the prototype based on an ethical baseline model that has been peer-reviewed and adjusted over time. Additionally, we did not hard-code the ethical questions; instead, we separated them into a different CSV file that can be edited by users. This approach allows users to easily adjust the questions as needed in the future.

Measurement Tools and Statistical Limitation:

Due to the dynamic nature of ethics and the limitations of timeframe and resources, statistical evaluation is not feasible. Therefore, to address this challenge, we employ a feedback loop instead.

Prototype and Data Security Limitations:

The prototype is currently limited to providing suggestions based on a single ethical dimension per user story, which might not fully address the multifaceted nature of ethical issues. Additionally, the use of external servers for processing could expose sensitive data to privacy risks. Future enhancements should focus on local processing to mitigate these risks (Section 5.5).

Generic Output Based on Predefined ECCOLA Suggestions:

The prototype currently provides generic suggestions based on predefined ECCOLA guidelines. In the future, adding customized suggestions tailored to the specific nature of each project related to each of the inputted user story could enhance the relevance and applicability of the recommendations (Section 5.5).

5.4.4 Conclusion Validity

Conclusion validity refers to the degree to which the conclusions we reach about relationships in our data are reasonable. It addresses whether the observed effects can genuinely be attributed to the interventions or variables tested rather than other factors or errors.

Prototype Scope and Ethical Dimension Coverage:

The study's focus on prototyping rather than developing a full-fledged system limits the depth of ethical analysis and the robustness of the findings. Currently, the prototype provides suggestions based on a single ethical dimension per user story, which may not fully capture the multifaceted nature of ethical issues

Data Privacy and Security Concerns:

The use of external servers for processing could expose sensitive data to privacy risks, which might influence the validity of the conclusions. Future enhancements should focus on local processing to mitigate these risks (Section 5.5).

5.5 Future Works

In this section, we list all the limitations identified in our previous work and propose solutions for future researchers. Additionally, we suggest ideas for expanding this thesis in the future.

There are three main areas suggested to be the focus for future development of this thesis: **implementation enhancement, converting the prototype to a real system, and improving evaluation methods**.

The suggestion system currently relies on third-party large language models (LLMs). Although the data protection policies of these third parties have been thoroughly reviewed, client data is still transferred to external entities. Therefore, to enhance data protection, future enhancements will **move the LLM process to a fully local setup**, minimizing the need to share client data externally. This shift is the most important **enhancement for the implementation** to ensure greater data privacy and security, which is essential for building trust and safeguarding sensitive information.

Second, the current version only classifies each user story under a single ECCOLA code. However, each user story can be viewed from multiple ethical perspectives. An enhancement can be made by **allowing each user story to receive suggestions from several ethical dimensions**. This enhancement aligns with how different humans might have different perceptions toward the most relevant ethical dimension of a user story.

Furthermore, as this project represents an initial effort to automate ethical risk assessments, the system is currently in a prototype stage in terms of its capabilities, user experience, and dataset size. **Transitioning from a prototype** to a fully operational system will require extensive refinement, including **dataset expansion, transparency, and enhancing the user experiences**.

The **expansion of the datasets** is also instrumental, because the limited availability of testing data constrains the robustness of the system. Engaging the general public with diverse backgrounds and robust feedback loop datasets is also very useful for improving the capability of the system.

An important consideration for future enhancements is the critical nature of user data. Therefore, all processed data will be **transparently disclosed on an acknowledgment page**. This transparency will adhere to principles of public engagement and awareness, data privacy, and will serve to educate software practitioners about the importance of transparency in data handling (Gray et al. 2021, and Nicodeme 2020).

Enhancements to the user interface involve both cosmetic improvements and user experience. Cosmetic enhancements can be achieved by **redesigning the front-end**. For further usability, adding **personalized ethical suggestions** with reasoning for each user story, instead of the current pre-coded suggestions, will be beneficial.

Finally, we suggest improving the **evaluation** technique. Statistical comparison, commonly used as the standard for AI testing, is not applicable to this scenario. The dynamic nature of AI might prioritize different ethical dimensions for the same user story. As the current prototype only addresses one dimension, the dimension addressed might not be the prioritized one and could therefore be marked as incorrect by statistical comparison, instead of reflecting the true nature of ethics, where different assessors may have different ethical dimension priorities. Coupled with the limitation of datasets size, due to time and resource constraints, statistical comparison at this stage will not be accurate. We recommend that future researchers explore methods for **statistically testing empathy-dependent AI** as a dedicated research project, following the expansion of the datasets and allowing multiple dimensions to be addressed. One way to approach this is by surveying the users of the final system, but this does not preclude other more suitable approaches that future researchers might develop as technologies evolve.

6 CONCLUSION

There are two differing viewpoints regarding the feasibility of automating ethical assessment in project management. One perspective asserts that automating risk assessment will enhance the efficiency of the project management process (Mitchell et al. 2022), particularly in fields that still lack public awareness, such as ethical IT (McNamara et al. 2018). However, other researchers argue that ethical design and decision-making processes cannot be automated because machines are not moral agents (Zuber et al. 2022).

This study explored the feasibility of automating ethical decision-making processes in Agile software project management through narrative literature review and the development of a prototype based on the ECCOLA model and GPT-3.5. This thesis work can be considered the initial stage of research in this field, as only two prior studies exist for automating ethical assessment, but neither of which specialize in IT project management. The prototype aimed to provide automated ethical suggestions for user stories in Agile projects, utilizing large-language model (LLM) and ECCOLA ethical model to classify and give suggestion to enhance the ethical standard for each of the user story.

Ethics encompass both empathy and logical reasoning. Even two individuals can possess differing perspectives on what constitutes ”good ethics.” Consequently, discussion and empathy are essential components of ethical assessment. The findings of this study indicate that fully automated decision-making is not feasible using the method we have chosen (LLM), due to the subjective and context-specific nature of ethics. Automation can only serve as a suggestion to address the intelligence aspect; human perspectives are necessary to incorporate the empathy component. However, automated tools can offer significant support to human judgment by providing logical and consistent ethical suggestions.

The prototype developed in this study demonstrates the potential to provide ethical suggestions automatically, acting as supportive insights rather than final decisions. This distinction is crucial, as ethical decision-making involves nuanced, context-dependent judgments that cannot be entirely captured by automated systems.

One of the key findings is the performance and limitations of using large language models (LLMs) like GPT-3.5 for generating ethical suggestions. The study selected GPT-3.5 for its superior reasoning capabilities and efficiency compared to other models evaluated, such as GPT-2 and BART. Despite the reliance on an external server, GPT-3.5 provided the best balance between performance and resource consumption, making it a suitable choice for the prototype. However, the reliance on third-party servers underscores the importance of addressing data privacy and security concerns. Future enhancements should focus on achieving full local processing to ensure data security and mitigate the risks associated with external data processing.

This research contributes to the field by providing a functional prototype that integrates ethical guidelines into Agile project management, offering a practical tool for ethical deliberation. Additionally, the study provides insights into the performance of LLMs in ethical suggestion generation, highlighting both their potential and limitations.

To enhance the robustness and practicality of the prototype, future research should concentrate on several key areas. Expanding the dataset utilized for fine-tuning and testing the prototype will boost its reliability and validity. Incorporating a more extensive and diverse dataset will ensure the prototype can handle a wider range of scenarios and ethical concerns. Another crucial area for future research is the personalization of ethical suggestions. The prototype currently offers generic suggestions based on predefined ECCOLA guidelines. Personalizing recommendations tailored to the specific context of each project and user story will enhance the relevance and applicability of the suggestions. This personalization will enable the prototype to effectively address the unique ethical challenges encountered by different projects.

Further developing the prototype to consider multiple ethical dimensions simultaneously for each user story is another vital direction for future research. The current prototype is limited to offering suggestions for a single ethical dimension per user story, which may not fully capture the multifaceted nature of ethical issues in software development. By addressing multiple ethical dimensions, the prototype will provide a more comprehensive and holistic ethical evaluation.

Lastly, statistical comparison, commonly used as the standard for AI testing, is not applicable to this scenario, given the dynamic nature of ethic combined with current available time-frame and resource. We suggest that future researchers explore methods for statistically testing empathy-dependent AI as a dedicated research project.

In conclusion, while full automation of ethical decision-making in software development is not feasible, automated tools can play a vital role in supporting human decision-making by providing logical and consistent ethical suggestions. The balance between automated support and human oversight is essential for maintaining high ethical standards in Agile software project management. By addressing the identified limitations and building on the proposed future research directions, the integration of automated ethical considerations in software development can be further realized, contributing to more ethically aware and responsible software engineering practices.

REFERENCES

Abdulhalim, Hisham, Yotam Lurie, and Shlomo Mark (2018). “Ethics as a quality driver in agile software projects”. In: *Journal of Service Science and Management* 11.1, pp. 13–25.

Amugongo, Lameck Mbangula et al. (2023). “Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications”. In: *AI and Ethics*, pp. 1–18.

Artificial Intelligence Set Up by the European Commission, Independent High-level Expert Group on (2019). “Ethics Guidelines for Trustworthy AI”. In: 6.

Barletta, Vita Santa et al. (2023). “A Rapid Review of Responsible AI frameworks: How to guide the development of ethical AI”. In: *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, pp. 358–367.

Brey, Philip et al. (July 2022). *SIENNA D6.1: Generalised methodology for ethical assessment of emerging technologies*. English. Zenodo. DOI: `10.5281/zenodo.7266895`.

Campbell, Donald T. and Thomas D. Cook (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

cinapr (2024a). *Ethical Automated Suggestion*. `https://github.com/cinapr/EthicalAutomatedSug` Accessed: 16.07.2024.

– (2024b). *LLM-Comparison-ForEthicSuggestiton*. `https://github.com/cinapr/LLM-Comparison-ForEthicSuggestiton`. Accessed: 16.07.2024.

Dam, Hoa Khanh et al. (2019). “Towards effective AI-powered agile project management”. In: *2019 IEEE/ACM 41st international conference on software engineering: new ideas and emerging results (ICSE-NIER)*. IEEE, pp. 41–44.

Dybå, Tore, Torgeir Dingsøyr, and Nils Brede Moe (2014). “Agile project management”. In: *Software project management in a changing world*, pp. 277–300.

Gibson, L, A Milligan, and C Peck (2022). “Addressing the elephant in the room: Ethics as an organising concept in history education”. In: *Historical Encounters* 9.2, pp. 45–63.

Gogoll, Jan et al. (2021). “Ethics in the software development process: from codes of conduct to ethical deliberation”. In: *Philosophy & Technology* 34.4, pp. 1085–1108.

Google (2024a). *Gemini*. `https://gemini.google.com/`. Accessed: 2024-08-09.

– (2024b). *Google Scholar*. `https://scholar.google.com/`. Accessed: 2024-08-09.

Gray, Colin M et al. (2021). “Dark patterns and the legal requirements of consent banners: An interaction criticism perspective”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–18.

Halme, Erika et al. (2024). "Making ethics practical: User stories as a way of implementing ethical consideration in Software Engineering". In: *Information and Software Technology* 167, p. 107379.

Hingle, Ashish, Andrew Katz, and Aditya Johri (2023). "Exploring NLP-Based Methods for Generating Engineering Ethics Assessment Qualitative Codebooks". In: *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, pp. 1–8.

Hoda, Rashina et al. (2023). "Augmented Agile: Human-Centered AI-Assisted Software Management". In: *IEEE Software* 40.4, pp. 106–109.

Holt, Deborah (2020). "Chapter 10 - What is a Moral (Ethical) Dilemma?" In: *Ethics and Society*. `https://viva.pressbooks.pub/phi220ethics/chapter/what-is-an-ethical-dilemma/`. Pressbooks/VIVA Open Publishing.

Hung, Le et al. (2018). "A risk assessment framework for construction project using artificial neural network". In: *Journal of Science and Technology in Civil Engineering (JSTCE)-HUCE* 12.5, pp. 51–62.

Ibáñez, Javier Camacho and Mónica Villas Olmeda (2022). "Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study". In: *AI & Society* 37.4, pp. 1663–1687.

Kemell, Kai-Kristian, Ville Vakkuri, and Erika Halme (2022). "Utilizing User Stories to Bring AI Ethics into Practice in Software Engineering". In: *International Conference on Product-Focused Software Process Improvement*. Springer, pp. 553–558.

Laudon, Kenneth C. and Jane P. Laudon (2020). "Management Information Systems: Managing the Digital Firm". In: *Management Information Systems: Managing the Digital Firm*. Sixteenth. Ethical and Social Issues in Information Systems. Boston: Pearson. Chap. 4, pp. 120–160.

Lucivero, Federica and Federica Lucivero (2016). "Scenarios as "Grounded Explorations". Designing Tools for Discussing the Desirability of Emerging Technologies". In: *Ethical Assessments of Emerging Technologies: Appraising the moral plausibility of technological visions*, pp. 155–190.

Lurie, Yotam and Shlomo Mark (2016). "Professional ethics of software engineers: An ethical framework". In: *Science and engineering ethics* 22, pp. 417–434.

LUT University (2023). *AI-based tools: Policies*. `https://elut.lut.fi/en/completing-studies/rules-and-regulations/ai-based-tools-policies`. Accessed: 2024-08-09.

– (2024). *LUT Primo*. `https://lut.primo.exlibrisgroup.com/discovery/search?vid=358FIN_LUT:LUT`. Accessed: 2024-08-09.

Matthews, George (2020). "Chapter 5 - Ethics: A Discipline Within Philosophy". In: *Ethics and Society*. `https://viva.pressbooks.pub/phi220ethics/chapter/introduction-to-ethics/`. Pressbooks/VIVA Open Publishing.

Max Rehkopf (n.d.). *User Stories: Examples and Template*. https://www.atlassian.com/agile/project-management/user-stories. Accessed: 2024-07-20. Atlassian.

McNamara, Andrew, Justin Smith, and Emerson Murphy-Hill (2018). "Does ACM's code of ethics change ethical decision making in software development?" In: *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*, pp. 729–733.

Microsoft (2024). *Bing*. https://www.bing.com/chat. Accessed: 2024-08-09.

Mitchell, Anna, Dharini Balasubramaniam, and Jade Fletcher (2022). "Incorporating Ethics in Software Engineering: Challenges and Opportunities". In: *2022 29th Asia-Pacific Software Engineering Conference (APSEC)*, pp. 90–98. DOI: 10.1109/APSEC57359.2022.00021.

Mökander, Jakob et al. (2021). "Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations". In: *Science and Engineering Ethics* 27.4, p. 44.

Nicodeme, Claire (2020). "Build confidence and acceptance of AI-based decision support systems-Explainable and liable AI". In: *2020 13th international conference on human system interaction (HSI)*. IEEE, pp. 20–23.

OpenAI (2023). *Privacy Policy*. https://openai.com/policies/privacy-policy/. Accessed: 2024-06-26.

– (2024a). *ChatGPT*. https://chatgpt.com/. Accessed: 2024-08-09.

– (2024b). *GPT3.5 API*. https://platform.openai.com/docs/overview. Accessed: 2024-08-09.

– (n.d.[a]). *Better language models and their implications*. https://openai.com/index/better-language-models/. Accessed: 2024-07-20.

– (n.d.[b]). *Enterprise Privacy*. https://openai.com/enterprise-privacy/. Accessed: 2024-06-26.

– (n.d.[c]). *Privacy Policy*. https://privacy.openai.com/policies. Accessed: 2024-06-26.

– (n.d.[d]). *Security and Privacy*. https://openai.com/security-and-privacy/. Accessed: 2024-06-26.

Perplexity (2024). *Perplexity*. https://www.perplexity.ai/. Accessed: 2024-08-09.

Piras, Luca et al. (2019). "DEFeND architecture: a privacy by design platform for GDPR compliance". In: *Trust, Privacy and Security in Digital Business: 16th International Conference, TrustBus 2019, Linz, Austria, August 26–29, 2019, Proceedings 16*. Springer, pp. 78–93.

Poplazarova, Tatjana, Claar van der Zee, and Thomas Breuer (2020). "Ethical decision-making in biopharmaceutical research and development: applying values using the TRIP & TIPP model". In: *Human Vaccines & Immunotherapeutics* 16.8, pp. 1981–1988.

Raab, Charles D (2020). “Information privacy, impact assessment, and the place of ethics”. In: *Computer Law & Security Review* 37, p. 105404.

Radford University (2020). “Chapter 11 - The Basics of Ethical Reasoning”. In: *Ethics and Society*. https://viva.pressbooks.pub/phi220ethics/chapter/the-basics-of-ethical-reasoning/. Pressbooks/VIVA Open Publishing.

Raith, Florian, Ingo Richter, and Robert Lindermeier (2017). “How project-management-tools are used in agile practice: Benefits, drawbacks and potentials”. In: *Proceedings of the 21st International Database Engineering & Applications Symposium*, pp. 30–39.

Richardson, George P (2011). “System dynamics, the basic elements of”. In: *Complex systems in finance and econometrics*.

Saltz, Jeffrey et al. (2019). “Integrating ethics within machine learning courses”. In: *ACM Transactions on Computing Education (TOCE)* 19.4, pp. 1–26.

Settembre-Blundo, Davide et al. (2021). “Flexibility and resilience in corporate decision making: a new sustainability-based risk management system in uncertain times”. In: *Global Journal of Flexible Systems Management* 22.Suppl 2, pp. 107–132.

Thompson, Paul B et al. (2021). “Early ethical assessment: An application to the sustainability of swine body scanners”. In: *Sustainability* 13.24, p. 14003.

Trantidou, Tatiana et al. (2022). “SENTINEL-Approachable, tailor-made cybersecurity and data protection for small enterprises”. In: *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*. IEEE, pp. 112–117.

Vakkuri, Ville, Kai-Kristian Kemell, Marianna Jantunen, and Pekka Abrahamsson (2020). ““This is just a prototype”: How ethics are ignored in software startup-like environments”. In: *International Conference on Agile Software Development*. Springer International Publishing Cham, pp. 195–210.

Vakkuri, Ville, Kai-Kristian Kemell, Marianna Jantunen, Erika Halme, et al. (2021). “ECCOLA—A method for implementing ethically aligned AI systems”. In: *Journal of Systems and Software* 182, p. 111067.

Vrije Universiteit Amsterdam (n.d.). *Generative AI: Your use, our expectations*. https://vu.nl/en/student/examinations/generative-ai-your-use-our-expectations. Accessed: 2024-08-09.

Wagner, Ben (2020). “Accountability by design in technology research”. In: *Computer Law & Security Review* 37, p. 105398.

Wang, Jianqiao (2023). “Intelligent Decision Support System for Building Project Management Based on Artificial Intelligence”. In: *Journal of Physics: Conference Series*. Vol. 2665. 1. IOP Publishing, p. 012022.

Zahedi, Mohammad Hadi, Alireza Rabiei Kashanaki, and Elham Farahani (2023). “Risk management framework in Agile software development methodology.” In: *International Journal of Electrical & Computer Engineering (2088-8708)* 13.4.

Zuber, Niina et al. (2022). “Empowered and embedded: ethics and agile processes”. In: *Humanities and Social Sciences Communications* 9.1, pp. 1–13.