

Master's thesis

Master's Programme in Computer Science

Utilizing Large Language Models to Update Systematic Literature Reviews: A Case Study on Time Pressure and Well-Being in Software Engineering

Jingyi Guo

July 11, 2024

Faculty of Science

University of Helsinki

Contact information

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki,Finland

Email address: info@cs.helsinki.fi
URL: http://www.cs.helsinki.fi/

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Study programme
Faculty of Science	Master's Programme in Computer Science

Tekijä — Författare — Author

Jingyi Guo

Työn nimi — Arbetets titel — Title

Utilizing Large Language Models to Update Systematic Literature Reviews:
A Case Study on Time Pressure and Well-Being in Software Engineering

Ohjaajat — Handledare — Supervisors

Prof. Mika Mäntylä

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	July 11, 2024	64 pages

Tiivistelmä — Referat — Abstract

This thesis examines the application of Large Language Models (LLMs), including ChatGPT-4 and ChatGPT-4.o, to update systematic literature reviews (SLRs) with a focus on time pressure and well-being in software engineering. It aims to address the challenges of traditional SLRs, which often need more timeliness and automaticity in synthesizing the latest research findings. By conducting a case study, this research evaluates the efficiency, comprehensiveness, and accuracy of using LLMs to update SLRs, comparing the strengths and weaknesses of various LLMs in processing and analyzing vast amounts of literature. The study aims to provide insights into the potential of LLMs to revolutionize the methodology of SLRs, particularly in dynamically evolving fields such as software engineering.

ACM Computing Classification System (CCS)
General and reference → Document types → Surveys and overviews
Applied computing → Document management and text processing → Document management → Text editing

Avainsanat — Nyckelord — Keywords

Large Language Models (LLMs), SLRs, Software Engineering, Time Pressure, Well-being, Case Study

Säilytyspaikka — Förvaringsställe — Where deposited

Helsinki University Library

Muita tietoja — övriga uppgifter — Additional information

Software study track

Contents

1 Introduction **1**

1.1 Research Background 1

1.2 Research Objectives and Questions 2

1.3 Significance of the Research 2

2 Literature Review **4**

2.1 Theoretical Framework 4

2.1.1 Introduction to Theoretical Concepts 4

2.1.2 Theoretical Models of Time Pressure and Its Effects in Software Engineering 5

2.1.3 Well-Being in Software Engineering 6

2.1.4 Systematic Literature Reviews (SLRs) in Academic Research 7

2.1.5 The Role of Large Language Models (LLMs) in Research 7

2.2 Review of existing literature 9

2.2.1 Existing SLRs on Time Pressure and Well-Being 9

2.2.2 Applications of LLMs in Systematic Literature Review 10

2.2.3 Case studies: successful application examples 11

2.2.4 Future Directions 13

2.3 Gaps in the Literature 13

2.3.1 Identification of Gaps 13

2.3.2 Potential of LLMs to Address These Gaps 14

3 Methodology **15**

3.1 Research Design 15

3.1.1 Introduction to Theoretical Concepts 15

3.1.2 Description of Research Process 15

3.2 Data Sources and Search Strategy 18

3.2.1 Data Sources 18

3.2.2 Search Strategy . . . 19
3.2.3 Iterative Optimization of Searches . . . 21
3.3 Inclusion and Exclusion Criteria . . . 23
3.3.1 Establishment of Criteria . . . 23
3.3.2 Application of Criteria . . . 25
3.3.3 Manual Review . . . 26
3.4 Data Processing and Analysis . . . 27
3.4.1 Data Extraction . . . 27
3.4.2 Data Integration . . . 28
3.4.3 Data Analysis Methods . . . 29

4 Results **31**
4.1 Summary of Selected Studies . . . 31
4.1.1 Data Collection Results . . . 31
4.1.2 Summary Process . . . 32
4.2 Thematic Analysis . . . 33
4.2.1 Article Publication Trend Analysis . . . 34
4.2.2 Analysis of Research Methods . . . 36
4.2.3 Relationship Between Time Pressure and Well-being . . . 38
4.2.4 Thematic Analysis of Main Findings . . . 41
4.3 New Discoveries and Trends . . . 46
4.3.1 Analysis of Emerging Trends . . . 46
4.3.2 Comparative Analysis of Findings . . . 50
4.4 Comparative Analysis . . . 52
4.4.1 Data Search and Processing Phases . . . 52
4.4.2 Data Extraction and Summarization . . . 53

5 Discussion **55**
5.1 Response to Research Questions . . . 55
5.1.1 RQ1: Updates in SLRs on Time Pressure and Well-being . . . 55
5.1.2 RQ2: Challenges of Lack of Automation in Updating SLRs . . . 55
5.1.3 RQ3.1: Efficiency and Accuracy of LLMs . . . 55
5.1.4 RQ3.2: Literature Summarization by LLMs . . . 56
5.1.5 RQ3.3: Insight Generation by LLMs . . . 56
5.1.6 RQ3.4: Technical Challenges in Automating SLRs . . . 56
5.2 Implications for Practice and Research . . . 56

5.2.1 Implications from RQ1 56
5.2.2 Implications from RQ2 and Sub-Questions 57
5.3 Limitations of the Study 57
5.4 Recommendations for Future Research 57

6 Conclusion and Future Work **58**
6.1 Conclusion 58
6.2 Future Work 58

Bibliography **61**

1 Introduction

1.1 Research Background

Systematic literature reviews (SLRs) are one of the most essential and valuable means of working in academia (Kraus et al., 2022). Often, they accumulate the findings of countless research studies and offer a survey of the present state of studies in many subjects. It becomes challenging to follow such rapid developments with their procedures due to the dynamics of innovation in the discipline area, such as software engineering, where technological advancement is rather dynamic. In the case of SLRs, delay in updating them can be critical, considering the time constraints and the well-being of software engineers, as the topic emerges crucial following the important role it plays in increasing productivity and health of the software engineering field.

In essence, the traditional practices of performing SLRs include manual synthesis, aggregation, and research analysis without automation, which is both time-consuming and has the tendency to induce human error and subjective bias. Due to the current influx of research articles being conducted, the period taken when assimilating the latest findings into existing SLRs can significantly delay the generation of useful outcomes to inform well-informed decision-making, especially with the fast-evolving technical professions.

Nevertheless, the development of the current LLMs, like OpenAI's ChatGPT, offered a new way to solve these issues (Hou et al., 2024). New AI patterns depend on humongous knowledge bases, which give them improved efficiency. It can produce syntactically well-formed and contextually relevant text output of considerably high quality that can be very helpful in the automation of extracting, summarizing, and analyzing literature data. By adopting LLMs into the SLR process, researchers might stand a chance of adding to the reviews' efficacy, applicability, and immediacy, thus making them portray current trends and advances more properly. Thereby promoting progress in industry and academia.

1.2 Research Objectives and Questions

This study aims to improve the process of updating SLRs by integrating LLMs. The objectives are designed to evaluate the existing literature on time pressure and well-being in software engineering. It also aims to explore how advanced LLMs' technologies can improve SLR methodologies.

According to the above, The research questions are below:

RQ1: What recent updates have been made in systematic literature reviews (SLRs) concerning the topics of time pressure and well-being in software engineering?

RQ2: What are the challenges associated with the lack of automation in updating SLRs?

RQ3: How to use LLMs to update SLRs?

RQ3.1: (Efficiency and Accuracy): How efficient and accurate are LLMs searching for relevant literature on time pressure and well-being in software engineering?

RQ3.2: (Literature Summarization): How can LLMs effectively summarize the collected literature?

RQ3.3: (Insight Generation): How do LLMs analyze possible relationships between literature conclusions?

RQ3.4: (Technical Challenges): What are the technical challenges and potential improvements of automating the SLR update process with LLMs?

1.3 Significance of the Research

The research is quite valuable for enhancing the methodological quality of the SLR. It also potentially alters the way that new information and knowledge are collected and shared in highly dynamic scientific disciplines. To achieve these goals, the study incorporates several current LLMs‘ technologies that may thereby improve the speed and reliability of literature reviews and enable their more frequent and timely updates.

The automation of certain steps and areas in the SLR process could also increase the feasibility of conducting comprehensive reviews. It can expand knowledge creation and sharing across the body of institutional and academic institutions and researchers. Besides, there is potential for the development of new approaches to coping with the levels of stress in the workplace and creating better conditions for the employees in the tech industries.

Therefore, the results could give some suggestions to the workplace organizations, the practices of HR departments, and more.

Additionally, this research is a possibility to expand the contribution of LLMs into the traditional academic and scientific research approach with a methodological outline and, thus, an effective methodological basis in precedent-setting studies across different fields. According to the study, LLMs could help expand the development of more effective approaches for completing complex analytical measures in the future.

2 Literature Review

2.1 Theoretical Framework

2.1.1 Introduction to Theoretical Concepts

Theoretical frameworks guide our understanding of how time pressure impacts software engineers' well-being, directly informing the research questions, methodologies, and analysis. These frameworks offer a structured approach and contextualize the inquiry within the broader literature on occupational stress and its management in high-demand tech environments (Blaikie, 2007). They propose the study questions, methodology, and analysis and contextualize the study within the broader context of existing knowledge (Kawulich, 2009). This research is grounded in four key concepts: time pressure, well-being, systematic literature reviews (SLRs), and the role of large language models (LLMs) in research.

Time pressure and well-being in software engineering are increasingly recognized as critical areas of inquiry, reflecting the dynamic and often demanding nature of the field. Research has shown that time pressure can negatively impact the well-being of software engineers, leading to decreased productivity, job satisfaction, and mental health (Meier et al., 2018; Ostberg et al., 2020).

SLRs are a powerful instrument for grasping the results that have been achieved. They also give an integrated view to enable the identification of topics that are in doubt and gaps that need to be addressed through further research (Kitchenham et al., 2009). Various SLRs have been widely employed in software engineering research to find this trend, assess the effectiveness of interventions, and develop evidence-based practice that addresses the problem in the right way. (Dybå and Dingsøyr, 2008).

The integration of Large Language Models (LLMs) in conducting Systematic Literature Reviews (SLRs) presents a novel approach to automating tasks that are traditionally manual and time-consuming. Leveraging technologies like OpenAI's GPT and Google's Bard, LLMs introduce transformative capabilities in software engineering, setting the stage for enhanced automation and efficiency in processes such as systematic literature reviews (Belzner et al., 2024). This integration aims to enhance the efficiency of conducting SLRs

and significantly reduces the risk of human error, bringing creativity and innovation to software engineering activities, including coding, design, and documentation (Fan et al., 2023). Hence, in this way, the language models can actually change established approaches to these procedures, turning the routine processes into more active and fruitful ones.

In this section, we shell out how the above-mentioned concepts link with the research session and then move towards a comprehensive analysis of each concept.

2.1.2 Theoretical Models of Time Pressure and Its Effects in Software Engineering

In software engineering, time pressure manifests as the acute perception of insufficient time, which empirical studies show diminishes both productivity and well-being. The Demand-Control and Job Demands-Resources Models provide a framework for understanding these impacts, suggesting that increasing job control and resources could mitigate the adverse effects of high demands (Kuutila et al., 2017). It can be quantified through self-report measures, project deadlines, and the ratio of work demands to available time. Research has shown that time pressure can negatively impact software engineers by reducing quality and increasing productivity, with notable adverse effects on individual well-being (Kuutila et al., 2021).

The Demand-Control Model suggests that high job demands, such as time pressure and low control over work, can lead to stress and decreased job satisfaction (Jonge et al., 2000). The model reflects the crucial role of control while being a mitigating influence in a demand-outcome relationship.

Expanding on this idea, the Job Demands-Resources Model suggests that while job demands can be stress-inducing, the presence of adequate resources can mitigate these effects and even enhance motivation and productivity (Bakker and Demerouti, 2007). This model stresses the balance between workload and resources as a primary factor shaping both productivity and wellness. It remarkably reduces the detrimental effect of increased load when resources are abundant (Veldhoven and Xanthopoulou, 2010).

On the one hand, these theoretical models have provided essential tools with which to approach time pressure-related effects in software engineering and, on the other, have highlighted the need for a delicate balance between time pressure and resources as the key determinant of both productivity and well-being.

2.1.3 Well-Being in Software Engineering

Well-being in software engineering is multifaceted and includes psychological resilience, physical health, and career satisfaction. Conservation of resources theory helps explain how protecting personal, social, and organizational resources can enhance an engineer's resilience to the high demands of their role. Psychological well-being, critical due to the high-demand nature of tech environments, often suffers under intense time pressures, impacting overall job satisfaction and mental health (Hobfoll, 2012). Physically, the sedentary nature of software engineering poses significant health risks, necessitating ergonomic solutions to mitigate adverse effects (Naber et al., 2021). On an occupational level, ensuring job satisfaction and career development is crucial for maintaining productivity and reducing turnover (Wright and Hobfoll, 2004). Utilizing the Conservation of Resources Theory, this study examines how resource allocation and stress management can enhance well-being, proposing that well-structured resource support systems can significantly alleviate the pressures experienced by software engineers (Hobfoll, 2001).

To complement our discussion on the impacts of time pressure on well-being in software engineering, the table 2.1 below summarizes key findings from general literature that explore these dynamics in various contexts.

Table 2.1: General Literature on Time Pressure and Well-Being in Software Engineering

Author(s)	**Year**	**Title**	**Key Findings**
EduValsania et al.	2022	Burnout: A Review of Theory and Measurement	High time Pressure Correlates with low job satisfaction
Prasetya et al.	2021	The Impacts of Perceived Organizational Support and Work Environment on Employee Performance: The Mediating Effects of Job Satisfaction	Effective time management enhances well-being
Pallewaththa et al.	2021	Occupational health: the impact of occupational stressors on job productivity	Correlation between time pressure and reduced Well-being

2.1.4 Systematic Literature Reviews (SLRs) in Academic Research

Systematic Literature Reviews (SLRs) are crucial in academic research for aggregating and synthesizing relevant studies within a specific field (Kawulich, 2009). These reviews use a standard procedure, ensuring caution and the possibility of repetition of results. This procedure consists of generating a research question, developing a study inclusion criteria, systematic literature search, appraisal, and synthesis of findings that are crucial in areas like software engineering where fast notches forward are usual (Kitchenham et al., 2009).

SLRs are indispensable in software engineering, a discipline that is constantly changing because of fast technological advances and an ever-increasing number of scientific publications. They assist in converging findings of different studies and summarize them meaningfully, highlighting the aspects that require deeper research. By following strict and universally accepted guidelines like those developed by Kitchenham and the PRISMA statement, SLRs become more informative in the course of developing software engineering theory (Kitchenham et al., 2009).

However, the rapid development within the field also necessitates regular updates to SLRs to maintain their relevance and accuracy. Studies indicate that SLRs can quickly become outdated without periodic updates, limiting their effectiveness and applicability in evolving fields (Nepomuceno and Soares, 2019).

It is important to realize that, in fact, new research data will be released to evaluate when SLR is indeed in need of conversion. Therefore, a systematic approach should be integrated to determine whether the missing data can be updated while maintaining the integrity and availability of the SLR in software engineering as much as possible. (Mendes et al., 2020).

2.1.5 The Role of Large Language Models (LLMs) in Research

Large Language Models (LLMs), such as OpenAI's GPT series, have emerged as transformative tools in the landscape of academic research. These models, trained on vast corpora of text, possess the capability to understand and generate human-like text, enabling a range of applications from content creation to information retrieval (Brown et al., 2020).

In the context of systematic literature reviews, LLMs offer theoretical and practical implications for automating various stages of the review process. This includes identifying and extracting relevant studies, data analysis, and drafting review sections (Xiao and Watson, 2019). Building upon their foundational role, LLMs like OpenAI's GPT significantly streamline the SLR process through automated data extraction and analysis, suggesting a potential reevaluation of traditional research methodologies and theoretical frameworks within software engineering. Theoretically, deploying LLMs could significantly reduce the time and labor associated with SLRs while potentially increasing the breadth of the literature surveyed. First, LLMs may translate to effective research extraction, identifying certain trends and synthesizing complex data across disciplines, thus enhancing the quality and accuracy of SLRs (Liu et al., 2024). Though LLMs are quite useful in this aspect, one needs to consider the accuracy of the automated analysis, the possibility of having the model bias, and the need to be supervised by a human in the review process (Liu et al., 2024).

As we explore the transformative role of Large Language Models in software engineering, the following table 2.2 provides a systematic overview of their applications, highlighting key studies and their findings on this subject.

Table 2.2: Systematic Review of LLM Applications in Software Engineering

Author(s)	**Year**	**Title**	**Main Findings**
Liu et al.	2024	On the Reliability and Explainability of Language Models for Program Generation	Discussed the reliability of LLMs in software
Hou et al.	2024	Large Language Models for Software Engineering: A Systematic Literature Review	Identifying trends and gaps in existing research
Fan et al.	2023	Large Language Models for Software Engineering: Survey and Open Problems	LLMs enhance automation in coding processes
Huotala et al.	2024	The Promise and Challenges of Using LLMs to Accelerate the Screening Process of Systematic Reviews	Exploring LLMs for Accelerating Systematic Review Screening

This research intends to investigate the theoretical aspects to measure the feasibility and

effectiveness of LLMs in transforming the existing SLR methodology, which is particularly essential as the area of software engineering experiences fast development and is full of new challenges every day.

2.2 Review of existing literature

2.2.1 Existing SLRs on Time Pressure and Well-Being

While reviewing systematic literature reviews (SLRs) in software engineering, it was identified that there has been a rising concern about the consequences of time pressure and well-being, which means psychological workload in technology-intensive settings has begun to be focused (Cox et al., 2000). Such reviews usually encompass surveys that compare and contrast the overall trends of empirical research, specifically focusing on the link between time and its impact on the individual's well-being, level of satisfaction at work, and level of productivity (Pallewaththa and Wijesiri, 2021).

The present research's underlying primary conclusions of the prior SLRs show that high time pressure has negative effects on burnout rates and job satisfaction (Edú-Valsania et al., 2022), while efficient time management and organizational support demonstrate positive impacts on well-being (Prasetya et al., 2021). However, regarding the above aspects, most of these studies describe general trends instead of personal ones. Also, they have largely concentrated on the effects of time pressure on subjects in the short run rather than the long run.

By and large, there is still an impressive track record that reveals the comprehensive ability of most SLRs to underscore the significance of the organizational strategies for directing employee well-being, including but not limited to offering flexible work schedules and supportive organizational work culture (Timms, 2013). These have been confirmed to reduce the impacts of time pressure and may provide insights into areas of need for intervention. These can enhance the welfare of the software company employees (Quick and Henderson, 2016).

Systematic literature reviews offer valuable insights into the nuanced effects of time pressure on well-being in software engineering. Table 2.3 consolidates these insights, providing a quick reference to significant studies in this area.

Table 2.3: Systematic Literature Reviews on Time Pressure and Well-Being

Author(s)	Year	Title	Main Findings
Kuutila et al.	2017	Reviewing Literature on Time Pressure in Software Engineering and Related Professions: Computer Assisted Interdisciplinary Literature Review	Identified a significant correlation between time pressure and increased burnout among software engineers.
Kuutila et al.	2018	Using experience sampling to link software repositories with emotions and work well-being	Found that high time pressure negatively impacts daily emotional well-being and job satisfaction.
Kuutila et al.	2020	Time pressure in software engineering: A systematic review	Concluded that time pressure is linked with lower software quality and higher rates of project failure.

2.2.2 Applications of LLMs in Systematic Literature Review

The use of Large Language Models (LLMs) in the systematic literature review (SLR) is a significant advancement in conducting research, specifically in the field of software engineering, where the research is continually evolving. LLMs like GPT-3. 5 and GPT-4, with copious neural networks available, provide enhanced ability to handle several problems that have remained a thorn in SLRs (Hou et al., 2024). According to existing reviews, LLMs have some applications. First, LLMs outperform in the speedy and effective conversion of text to helpful outcomes for literature searches and data extraction. when a large number of new research papers are published, LLMs allow for a quick figuring out and classification of studies according to the set of filters, saving considerable time in such stages (Khraisha et al., 2024).

The second is improving accuracy and consistency. A major weakness that is usually realized in SLRs is the process of how studies are selected and also the way data is analyzed. Nevertheless, LLMs assist in standardizing the reviews by leaving no room for human interference in the assessment of texts due to weeding out potential errors and biases at their best. They also have the ability to decipher the natural language with

relative ease and can determine nuances or contexts (Hou et al., 2024).

The third is automated pre-screening for identification of titles and abstracts. The most time-consuming part of the SLRs, called screening of titles and abstracts for study relevance, is easily automatable for LLMs (Huotala et al., 2024). Concretely, when trained on specific examples or a specific set of criteria, LLMs can filter out papers that are irrelevant to the collection or that do not reach a quality threshold necessary for inclusion in the initial stages of the review process.

The fourth is that in addition to being used for collecting data, LLMs can also be applied to complex tasks such as sentiment analysis, trend detection, and thematic categorization (Labellerr, 2023). In software engineering SLRs, these capabilities can be especially valuable for detecting novel patterns, evaluating the differences in methodological strategies across the studies, and integrating the huge amount of available information into meaningful ones.

Research on using LLMs has been conducted, as they have practical applications. Below are some ideas on how to enhance SLRs using LLMs. First, LLMs can be used to ascertain and extract key themes from a number of papers, thereby contributing to the retrieval and synthesizing of new findings (Maragheh et al., 2023). Second, LLMs can be used to retrieve the results of specific studies and initially turn them into abstracts and summaries that will help arrange the information as quickly as possible for a researcher (Gao et al., 2024). Third, it could potentially allow LLMs to attempt to screen lower-quality studies as effectively as possible and increase the specificity of the screening(Cao et al., 2024).

Therefore, incorporating LLMs into systematic literature reviews showcases how methodology applied in the fields can enhance the efficiency and effectiveness of these studies and transform the processes through which interdisciplinary SLRs occur. With the advancement in technologies, LLMs are likely to be regarded as part of SLR activity in the future, especially in active fields like software engineering (Hou et al., 2024).

2.2.3 Case studies: successful application examples

These two case studies provide realistic examples of how LLMs are being implemented in real-life instances to modernize domains like software development or education and what they can tell about the evolution of traditional approaches to applying these concepts in the future.

In the first case study, the Large Language Model was applied to generate tests for “Da.

tes," a networking site for startup development, investment, and cooperation with companies (L. Junior et al., 2023). The project also aimed at identifying how it could be possible for LLMs to match or even exceed human testers' abilities in test case generation. The selected LLMs were specialized for software development and used for training using a large variety of software development paradigms, languages, frameworks, etc, so as to familiarise the models with realistic conditions of SW development. The implementation also includes the setting of parameters where the LLMs will be programmed to produce test cases depending on functionality/user scenarios and value parameters. These test cases were then compared with human-generated test cases to determine the clarity, correctness, completeness, and consistency-based checks. The results were revealed, indicating that the LLMs not only came up with a high number of test cases in a short period of time, increasing productivity by a large number of times but also minimized the cycle time of the test case generation from weeks to days. Moreover, it was seen that the quality, as well as the number of the sample test cases that were being developed by the LLMs of this research, was found to be better or even better than the human test cases, thus implying a high possibility of making the software test and development a revolutionary one by using LLMs of this research.

The second case study was about implementing LLMs such as ChatGPT to transform education by incorporating the models into learning spaces to iterate and improve educational content as well as customize learning processes (Hsu and Ching, 2023). An advanced LLM was required to be developed with a view to its flexibility in terms of learning style and issues, subject, and contents and interactions. This LLM implementation was used to design engaging and participatory learning contexts as well as afford opportunities for generating course content based on individual learning profiles, with the ability to modify content in response to students' reactions and performance. Some of the student learning outcomes of the integrated LLMs included The positive promotion of student engagement and the resulting improvement in student outcomes that was influenced by the instructional effectiveness of the LLMs that personalize learning experiences in an effort to facilitate retention of knowledge. Furthermore, the way that content could be created autonomously reduced the amount of effort required to provide material and enabled educators to devote much more of their time to teaching. Not only did this study revolutionize the way education content is conveyed, but it also offered useful data on the applicability of AI in educating and improving learning and became a key to future investigation on adaptive learning technology.

2.2.4 Future Directions

With the progress in integrating the LLMs with systematic literature reviews, reality tasks, and others. Future prospects for the evolution and development of the presented approach remain rather bright. The following are several key areas of particular relevance for future research and progress.

One is to improve the computing power of LLMs (Zhang et al., 2024). It can help handle the growing and increasing amount of data sources and queries that can be posed in a certain time frame. This can help researchers working in environments where they do not have the ability to access computing resources or have little experience in programming.

The second is to improve the training of the model and its adaptation to the needs (Tang et al., 2023). Modification of the models to incorporate finer details could enhance the usefulness of the existing models.

The third is to create dialogical AI tools for discussions with the models, posing additional questions or focusing on certain aspects. This will change LLMs from simple passive information disseminators to interactive participants in the research process, thus improving the engagement and personalization of literature reviews for end users.

Last but not least, automating meta-analysis by using LLMs effectively decreases the time and manpower needed to analyze quantitative data percentages across a large number of studies (Susnjak et al., 2024). It concerns not only the extraction of data but also quality control and synthesis of descriptive statistical calculations.

2.3 Gaps in the Literature

2.3.1 Identification of Gaps

The systematic literature review on time pressure and well-being in software engineering found several key gaps.

The first is the gaps in existing SLRs. The first section is the lack of integration of recent technological advances, whose combined effects on time management and psychological health have not been fully explored (Roberts et al., 2021).

The second is that studies' definitions and measurements are inconsistent in many cases. There is a lack of standardization in defining and measuring key concepts such as "time

pressure" and "well-being" (Larsson et al., 2017). This inconsistency has some impact on the ability to synthesize research results and draw comprehensive conclusions.

Then, there are the methodological limitations of previous studies. In this section, the first is the reliance on self-report data, which many studies still rely heavily on, introducing biases such as recall bias and social desirability bias. These biases affect the accuracy and reliability of findings on the effects of time pressure. The second is the limited use of advanced analytical techniques: Advanced analytical techniques have not been fully utilized, which can provide deeper insights from large datasets or text data. The application of methods such as machine learning and natural language processing is essential to discover new patterns and relationships.

2.3.2 Potential of LLMs to Address These Gaps

The use of LLMs can overcome the identified limitations in the following ways: First, enhanced data collection and synthesis. LLMs can significantly improve the data collection process by quickly scanning and extracting relevant information from a large amount of literature. This feature is particularly valuable for integrating the latest research results and ensuring that reviews remain up-to-date and comprehensive. Second, standardization of definitions and measurements. LLMs can be ordered to identify and standardize key concepts, thereby reducing inconsistencies between studies and improving the reliability of analysis. Third, reduced reliance on self-reported data. By integrating data from different sources (including technical documents and online forums), LLMs can provide a more balanced and objective perspective and reduce reliance on potentially biased self-reported data.

LLMs also have the following theoretical advantages. The first is comprehensive literature coverage. LLMs are able to process information at a scale that human researchers cannot achieve, so a wider range of sources and more data can be included. The second is the ability to improve the efficiency of the review process (Huotala et al., 2024). LLMs can speed up the review process by optimizing the screening process of initial articles and improving data extraction efficiency, allowing researchers to focus on more in-depth analytical tasks. The third is the advanced analytical capabilities of LLMs. Employing sophisticated NLP techniques (Roumeliotis et al., 2024), LLMs can perform complex topic, sentiment, and trend analysis, providing new insights into the dynamics of time pressure and well-being in software engineering.

3 Methodology

3.1 Research Design

3.1.1 Introduction to Theoretical Concepts

To directly support the understanding of our research methodology, we briefly introduce the relevant theoretical concepts here. The case study methodology was selected for this research. It aims to provide an in-depth exploration of the practical application of Large Language Models (LLMs) such as ChatGPT-4 and ChatGPT-4.o in updating Systematic Literature Reviews (SLRs) within software engineering. A case study approach enables an in-depth investigation of how LLMs process and analyze large volumes of constantly updated academic literature. It allows for detailed documentation of each step in using LLMs for literature searches, screening, data extraction, and analysis. This methodology offers flexibility to adjust to new findings and emerging issues during the experiment, which is important when working with evolving technologies that are still being perfected.

In this study, we decided to use LLMs such as ChatGPT-4 and ChatGPT-4.o. This is because they have been shown to be able to handle complex text-processing tasks related to systematic literature reviews in software engineering. According to Huotala et al. (2024), these models have been evaluated for tasks such as title abstract screening, measuring their efficiency and accuracy in automating parts of the systematic review process (Huotala et al., 2024). The study particularly highlighted the potential of newer LLMs, such as ChatGPT-4, which has better screening accuracy than ChatGPT-3.5. Therefore, ChatGPT-4 and its updated model ChatGPT-4.o will be used in this study.

It is hoped that through this study, the use of LLMs will simplify the SLR process and explore the difficulties and advantages that may arise from using these techniques in software engineering SLRs.

3.1.2 Description of Research Process

The research methodology adopted in this study combines human expertise with the advanced capabilities of LLMs. This ensures iterative refinement of search parameters and

a thorough analysis of the literature.

The process began with the researchers manually defining initial search parameters to meet the specific needs of "Time Pressure and Well-Being in Software Engineering". These parameters were used to conduct an initial literature search in databases like IEEE Xplore, Scopus, and Google Scholar in order to gather a wide range of relevant literature and establish a solid understanding of the topic.

Figure 3.1 below illustrates the step-by-step Systematic Literature Review (SLR) process utilized in this study, from the initial parameter setting through to the final thematic analysis, highlighting the iterative use of Language Model Models (LLMs) at various stages.

After the initial search, the effectiveness of the search parameters was evaluated by analyzing the number and relevance of the retrieved literature. LLMs were then used to refine and optimize search keywords and logic based on this initial data. This dynamic adjustment of the search strategy, facilitated by the advanced natural language processing capabilities of LLMs, ensures that subsequent searches are more targeted and produce more relevant research. This iterative process is essential to screen out the most relevant literature and improve the accuracy of the research.

During the initial screening stage, the study used LLMs to assess the relevance of the retrieved literature based on key information in the title and abstract. This significantly reduced the amount of literature that required detailed manual review. This stage of the screening process draws on insights from the work of Huotala et al. (2024), which highlighted the potential of LLMs to accelerate the screening process without compromising accuracy (Huotala et al., 2024). LLMs are able to efficiently filter large amounts of data, ensuring that only the most relevant articles advance to a more in-depth review phase.

During the in-depth screening phase, LLMs evaluate the full text of selected articles to assess their suitability and quality. This includes a thorough analysis of the methods employed, the validity of the results, and the relevance of the article to the research topic. Following this, LLMs automatically extract basic data such as research methods, main findings, and study limitations and store this information in the research database for further analysis. A manual review is necessary for articles undergoing thorough screening to determine if they align with the research objectives.

Finally, LLMs perform thematic identification and analysis of the extracted data, identifying and constructing key themes, trends, and patterns for an updated systematic literature

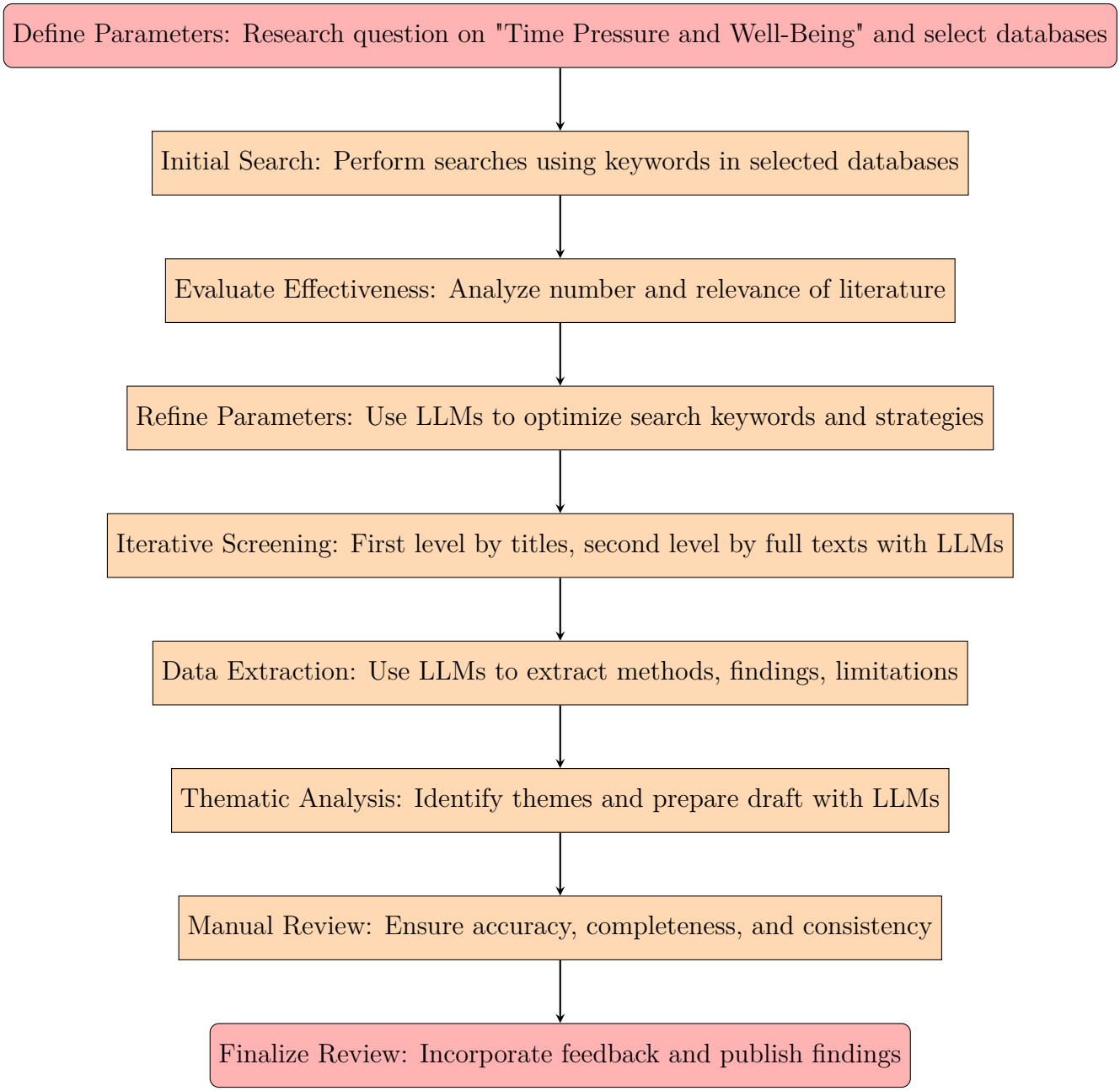

Figure 3.1: Flowchart of the Systematic Literature Review Process

review. The draft produced through this synthesis is subject to rigorous manual review to ensure accuracy, completeness, and consistency with the research objectives. Feedback from this evaluation may lead to further adjustments to the LLMs configuration and optimization of the search and analysis algorithms. This iterative process of improvement will be repeated until the updated SLR meets high publication standards and reflects the latest research trends and theoretical developments.

3.2 Data Sources and Search Strategy

3.2.1 Data Sources

Identification of relevant databases for inclusion is important in achieving breadth and depth of information in SLRs. Thus, this research uses credible database sources to gather comprehensive studies on time pressure and well-being conducted within software engineering. Here is an overview of the chosen data sources and the rationale for their selection:

IEEE Xplore: Being an integrated electronic library in electronics and computer science domains, IEEE Xplore provides various resources for scholarly and professional publications of journal articles, conference proceedings, and standards. It is important to embrace the most current material in software engineering and its allied technologies to make it a perfect database for this research.

Scopus: Storming with a vast number of articles, citing numerous scientific fields like sciences, technologies, medical and social spheres, Scopus offers powerful citation indexing and a purposely wide journal list.

ACM Digital Library: Currently, CHI is featured in the ACM DL as a primary electronic library in the field of computer science, offering conference proceedings, scientific journals, newsletters, and books. Most importantly, it is rich in software engineering practices and is drawn from theories.

Web of Science: Unlike other databases that capture content across the various disciplines, this database provides citation and indexing of all published materials in the natural and applied sciences, engineering and technology, biomedical and life sciences, social sciences, arts, and humanities. It is well-known for its connection to scholarly journals, conference proceedings, reports, books, and patents.

Google Scholar: As a multi-disciplinary organization, Google Scholar offers much more than the articles in scholarly journals; it also offers conference papers, theses, technical reports, and similar literature. However, Google Scholar is only used to enhance search effectiveness, mainly because of the restricted ability to export databases and the duplicate of other database domains.

These databases were chosen because they focus on scholarly publications in specific fields of study and offer academic and peer-reviewed content. This selection helps to narrow

down the range of works within the framework of the given topic, thereby improving the SLR's efficiency and reliability.

3.2.2 Search Strategy

Since the study encompasses time pressure and well-being in software engineering from the literature, the search method plays a significant role in reviewing the literature. This study mainly used LLMS: ChatGPT-4 and ChatGPT-3.5 to incorporate LLMs for fine-tuning search queries and more proficient keyword selection, query formulation, regulation, and setting up search parameters.

The initial keyword list based on the research topic includes terms such as "time pressure," "well-being," "software engineering," and "stress." Based on the results obtained from manual searches in different databases, these keywords were expanded to include synonyms and related phrases through the use of LLMs to increase the search scope. For example, By using ChatGPT-4 to assist in expanding the initial keyword list and refining the search queries: "time pressure" was expanded to include "deadline pressure" and "work intensity," while "well-being" was expanded to include "mental health" and "job satisfaction." Given that the well-being of software engineering professionals may involve multiple fields, including psychology and management, cross-disciplinary keywords such as "occupational health" and "work-life balance" will also be included.

Table 3.1 shows the expansion of the initial keywords using ChatGPT-4, demonstrating the broad range of terms considered for comprehensive literature retrieval.

Table 3.1: Expansion of Initial Keywords using LLMs

Initial Keyword	**Expanded Keywords and Phrases**
time pressure	deadline pressure, work intensity, time constraints, time management challenges
well-being	mental health, job satisfaction, employee wellness, occupational health
software engineering	software development, software projects, software industry
stress	burnout, work-related stress, psychological stress

Table 3.2 compares manual and LLMs-enhanced queries in Scopus. ChatGPT - 3.5 and ChatGPT-4 both broadened the search scope and refined search terms, leading to more

relevant literature. After manual comparison, we found that ChatGPT-4 results are much better than ChatGPT-3.5 results, so the following table 3.2 shows the results of ChatGPT-4.

Table 3.2: Comparison of Manual and LLMs-Enhanced Search Queries in Scopus

Initial Query	**LLM-Modified Query**	**Results Change**
“time pressure" AND "software engineering" AND "well-being"	“time pressure" OR "workload" AND "software engineering" AND "well-being" OR "employee satisfaction"	85 → 311
“stress" AND "software engineering" AND "employee health"	“stress" OR "burnout" AND "software engineering" AND "employee health" OR "occupational health"	59 → 423
“project deadlines" AND "software engineering" AND "stress"	“project deadlines" OR "time constraints" AND "software engineering" AND "stress" OR "work-related stress"	0 → 98
“work hours" AND "software engineering" AND "well-being"	“work hours" OR "overtime" AND "software engineering" AND "well-being" OR "job satisfaction" OR "employee wellness"	44 → 102

The results, as expected (see Table 3.2), indicate a significant improvement in the relevance and quantity of retrieved articles.

The study also uses Boolean logic to construct queries and combine these terms effectively. The basic query template includes combinations such as (time pressure OR deadline pressure) AND (well-being OR mental health) AND software engineering. In addition, using the natural language understanding capabilities of LLMs, more complex query expressions can be constructed, such as finding articles based on the search formats of different databases. For example, the query can be designed as TS= ("stress" OR "burnout") AND TS= ("software engineering") AND TS=("health" OR "well-being").

Optimize specific parameters for each database using the functionalities provided by IEEE Xplore and Web of Science. The limitations are set for the search based on publication date, focusing on the last ten years of publications and language, limiting it to English only.

3.2.3 Iterative Optimization of Searches

The research employed an iterative refinement process to guarantee the relevance and comprehensiveness of the search strategy for updating the SLRs. This study aimed for continuous improvement based on the initial search results by harnessing the capabilities of LLMs.

A quantitative analysis of the initial search results using LLMs revealed that only 3% of the top 1000 documents related to "time pressure" were explicitly linked to software engineering. This suggests that research needs to refine or combine the keywords with more specific terms. In addition, ten of the papers were reviewed by random selection. A quality assessment of the initial search results revealed that only four papers were highly relevant and of high quality. This suggests that broad terms like “happiness” may need to be replaced with more specific terms such as “job satisfaction in software engineering” to improve relevance and accuracy.

Based on these insights, adjustments were made to the search strategy. The keyword and query adjustments moved from a basic combination such as (time pressure OR work intensity) AND (well-being OR mental health) AND software engineering to a more targeted approach: (TS=(time pressure OR work intensity OR "deadline stress" OR "time management challenges") AND TS=(well-being OR mental health OR "job satisfaction" OR "employee wellness") AND TS=(software engineering OR "software development" OR software)). This adjustment reduced the number of search results from 7791 to 114, significantly enhancing the focus and decreasing irrelevant results.

However, recognizing the concern that requiring both concepts (time pressure and well-being) simultaneously might lead to missing many relevant documents, the search strategy was further refined to allow the study of these concepts both separately and in conjunction. Separate searches were conducted using combinations such as TS=("time pressure" OR "work intensity") AND TS=("software engineering"), and TS=("well-being" OR "mental health") AND TS=("software engineering"). This dual approach ensured a broader capture of relevant literature.

Search parameters were also modified. The date range was narrowed from 2014–2024 to 2017–present, and specific database fields such as "Conference Papers" in IEEE Xplore were targeted to capture the most current and practice-relevant research.

A second round of search was conducted using the adjusted parameters, and a total of 64 papers were collected. Analysis using ChatGPT-4 showed that 61 of them were related

to software engineering. And the proportion of relevant papers increased from the initial 3% to about 95.31%. This significant numerical increase shows the dynamic nature of the optimization process.

Real-time analysis of keyword frequency and document quality after each search iteration showed a significant increase in relevance and quality of literature when the terms "wellness" and "software" were combined.

Table 3.3 compares manual and LLMs-enhanced queries in the Web of Science database. The improved search queries, utilizing LLMs, significantly expanded the search scope and refined search terms, leading to more relevant literature.

Table 3.3: Comparison of Manual and LLMs-Enhanced Search Queries in Web of Science

Manual Query	**LLMs-enhanced queries**	**Results Change**
"time pressure" AND "software engineering" AND "well-being"	(TS=(time pressure OR work intensity OR "deadline stress" OR "time management challenges") AND TS=(well-being OR mental health OR "job satisfaction" OR "employee wellness") AND TS=(software engineering OR "software development" OR software))	$2 \rightarrow 114$
"stress" AND "software engineering" AND "employee health"	(TS=(stress OR "burnout") AND TS=(software engineering) AND TS=("health" OR "well-being"))	$0 \rightarrow 20$
"project deadlines" AND "software engineering" AND "stress"	(TS=(project deadlines OR "time constraints") AND TS=(software engineering) AND TS=(stress OR "work-related stress"))	$0 \rightarrow 2$
"work hours" AND "software engineering" AND "well-being"	(TS=(work hours OR "overtime") AND TS=(software engineering) AND TS=(well-being OR "job satisfaction" OR "employee wellness"))	$1 \rightarrow 1$

The table (see Table 3.3) also shows the increase in the relevancies and amount of the found articles, especially when employing the intricate queries based on time pressure, stress, and well-being of software engineers. With LLM's advanced natural language processing capabilities, search strategies can be adjusted dynamically. This ensures that subsequent searches are more targeted and that the results yield more relevant research.

The study also employs Boolean logic to effectively construct queries that combine these terms. For instance, the basic query template used includes combinations such as (time pressure OR deadline pressure) AND (well-being OR mental health) AND software engineering. Additionally, using the natural language understanding capabilities of LLMs, more complex query expressions can be constructed, such as finding articles based on the search formats of different databases. For example, the query can be designed as TS= ("stress" OR "burnout") AND TS= ("software engineering") AND TS=("health" OR "well-being").

In order to enhance specific parameters for each database, it is crucial to utilize the features offered by IEEE Xplore and Web of Science. Limitations should be applied to the search based on publication date, concentrating on the last seven years of publications and language, restricting it to English only.

This iterative optimization process guarantees not only a comprehensive and thorough literature search but also enhances the quality and relevance of the search results through continuous adjustments and improvements.

3.3 Inclusion and Exclusion Criteria

3.3.1 Establishment of Criteria

In this SLR, the inclusion and exclusion criteria are designed based on the rapid evolution of technologies and methodologies as well as the need to maintain the research's relevance and timeliness:

Time Range: Therefore, the selection is limited to articles published only from 2017 up to the present since this discipline changes frequently. Previous studies, even if they constitute an essential background or theoretical framework, are only considered if later works cite them.

Study Types: The domains of the review include quantitative research, qualitative research, and hybrid research; however, the excluded forms contain mere opinions, articles, and journal reviews that present opinions, news, and events without empirical findings.

Literature Quality: The sources to be used can only be articles that appeared in refereed academic journals and conference proceedings. This means that sources such as blogs or conference papers that have not been published can be left out. Thus, all selected papers

are relevant, reliable, and scientifically sound.

Research Geographic Location: The work under analysis encompasses the research from all regions to affirm the results' generalizability. If the studies are conducted in high regional studies, perhaps experiments that are not generalized to software engineering practices should be addressed.

Relevance to Time Pressure and Well-being: Studies were included if they addressed themes related to time pressure and well-being, even if the primary focus differed. This includes research on injecting safety critical thinking or other relevant topics that provide methodological insights or contextual relevance. Such studies are valuable because they offer broader perspectives or innovative methodologies that can be applied to understanding time pressure and well-being in software engineering.

The following criteria were used in this table 3.4 to determine the subject to be included in the review so as to obtain only the most relevant literature based on the methodology to be followed. The findings come from peer-reviewed scientific articles released in the past few years and are relevant to global society.

Table 3.4: Inclusion and Exclusion Criteria for Systematic Literature Review

Criteria	**Inclusion**	**Exclusion**
Time Range	Articles from 2017 to present	Articles before 2017 unless cited by later works
Study Types	Quantitative, qualitative, and hybrid research	Opinions, articles, reviews without empirical findings
Literature Quality	Refereed academic journal articles and conference proceedings	Blogs, unpublished conference papers
Research Geographic Location	Studies from all regions	Region-specific studies not generalizable to global practices
Relevance to Time Pressure and Well-being	Studies addressing themes of time pressure and well-being, or providing methodological insights applicable to these themes	Studies not offering relevant insights or methodologies applicable to the themes

3.3.2 Application of Criteria

Using LLMs to automate the application of these criteria ensures that the selection process is both efficient and precise. Using Large Language Models LLMs to automate the application of these criteria ensures that the selection process is both efficient and precise:

Automatic Identification of Literature: It establishes that LLMs can also find the type and quality of the given documents from the respective literature metadata. For instance, a document from IEEE Xplore is considered a Peer Reviewed Conference Paper, while a document like a blog post is not.

Time Range Implementation: To apply the most relevant studies based on the date, LLMs employ algorithms to recognize the date, excluding studies beyond the range while having the option to consider extraordinary previous-year research.

Geographic Applicability: The LLMs evaluate the geographical and contextual aspects of the studies and identify peer-reviewed pieces that are too specific to a region for the next analysis stages.

Interdisciplinary Keyword Recognition: Through the context and semantics, the LLMs execute the inclusion of contextual research that could employ unique names but is associated with themes such as 'occupational health' and 'software engineering.'

The following table 3.5 shows the application of LLMs to the inclusion and exclusion criteria mechanism in this systematic literature review, reducing and increasing efficiency in the various functional activities.

Table 3.5: Automation of Inclusion and Exclusion Criteria Using LLMs

Feature	**Function**
Automatic Identification of Literature	Confirms document type and quality.
Time Range Implementation	Filters studies by date allow manual review of older studies.
Geographic Applicability	Evaluates and flags studies based on regional applicability.
Interdisciplinary Keyword Recognition	Detects and aligns studies with relevant interdisciplinary themes.

3.3.3 Manual Review

So far, LLMs have greatly improved initial literature screening in terms of speed and scope. However, manual review is still crucial because it helps optimize LLMs' screening results, detect potential errors, and ensure the relevance of the identified literature. Although LLMs are specifically designed to process big data, they sometimes fail to grasp the deeper meaning of the work, especially in academic works.

Many times in this study, LLMs will not return perfect results. Especially after the system update of OpenAI after mid-May, the computational efficiency of 4 has been greatly reduced. So, this study needs a manual review to confirm the results. Although LLMs can process large amounts of data quickly, they sometimes miss the nuances of language and context that are important to the academic community. And misclassification may occur. Therefore, a manual review must be performed to correct any such errors and ensure that the literature topics are well and correctly classified.

Keywords themselves also present different challenges. Changes in character positioning, such as capitalization, using different hyphenation rules, or even slightly modifying the wording, will greatly affect the relevance of the results. However, the manual review can distinguish between these papers, which may be considered basically similar during the automatic filtering process but may actually differ in key aspects.

In addition, the manual review also helps to check the design of the study, the suitability of the sample, the validity of the results, and the reliability of the conclusions. Performing these assessments is very important to understand the practical significance of the research proposal. For example, in this study, when using LLMs to filter documents containing "happiness" and "stress", the initial return provided only 9 articles, a significantly small number compared to expectations. This is necessary to lead to performing a manual review. It turned out that different formats of "happiness" and incorrect interpretations of "stress" (some papers used it to describe physical stress rather than psychological stress) affected the results. In addition, applying fuzzy search techniques to include various spellings and grammatical structures ultimately increased the total number of relevant documents to 64.

This means that, despite the continuous advancement of technology and its increasing efficiency, human scrutiny will always be important in understanding and synthesizing scholarly language, especially in selecting the right literature.

3.4 Data Processing and Analysis

3.4.1 Data Extraction

Extracting key information from the literature was the backbone of our data processing strategy for this study. The main goal was to systematically extract relevant data using LLMs such as ChatGPT-4 and ChatGPT-4.o and compare the efficiency of these LLMs. This partly exploits the fact that these models are specifically configured to recognize and interpret structured and unstructured text of academic papers. Data extraction was designed around templates to ensure consistency and comprehensive capture of necessary details.

The following table 3.6 is an example that displays how LLMs are configured to automate the data extraction process for this systematic literature review.

Table 3.6: Configuration Methods for LLMs

Data Type	**Configuration Details**
Author and Publication Date	Configured to recognize signatures and dates typically found at the top or bottom of documents.
Research Method	Focused on methodological sections, trained to identify phrases related to different study designs.
Main Findings	Trained to detect key phrases in conclusions or results sections.

This study used LLMs to find and extract author and date information by scanning documents or their online query functions. This information is usually located at the top or bottom of the paper for scanned documents. The models were trained to identify sections that clearly describe the research methods and main findings, which is crucial for gathering insights about the methods used and the results discussed in the study.

An example from the literature by Kuutila et al. is provided to demonstrate the practical application of these configurations (Kuutila et al., 2020). This example in table 3.7 illustrates how LLMs have been effectively employed to extract detailed elements from a study, reflecting the intricate processes and outcomes.

Table 3.7: Example of Data Extraction: Kuutila et al. (2020)

Item	**Details**
Title	Time pressure in software engineering: A systematic literature review
Purpose of Study	To explore the impact of time pressure on productivity and software quality in software engineering.
Author and Publication Date Extraction	Recognized "Mikko Kuutila" as the primary author with a publication year of "2020."
Research Method Extraction	Identified the methodology as a systematic literature review, including preliminary screening of 5,414 resources and selection criteria for 102 final papers.
Main Findings Extraction	Recorded findings suggest that time pressure generally leads to short-term productivity increases, often at the expense of software quality.

This extraction process enabled every aspect of the literature to be analyzed in detail and captured data consistently for different literature in this study, laying a solid foundation for the subsequent data integration and analysis phase.

3.4.2 Data Integration

The data integration process in this study utilized LLMs to integrate software engineering time pressure and well-being research from multiple sources into a unified data framework. This is critical to understanding how time pressure affects the productivity and well-being of software engineers.

First, the data integration strategy was to collect research papers from different scientific databases (IEEE Xplore, Scopus, ACM Digital Library, Web of Science). This expanded the dataset and ensured a comprehensive analysis. The data was then preprocessed and standardized using LLMs to extract key information from each paper, including authors, publication date, research methods, and main findings. The data was then standardized into a unified format to ensure consistency across the dataset.

The following table 3.8 provides a template for the automated extraction process tailored to enhance data consistency and accuracy.

LLMs are then used to align similar data points, such as research methods and key findings,

Table 3.8: Enhanced Data Extraction Template Automated by LLMs

Feature	**Function**
Author(s)	Verify author information for accuracy.
Publication Date	Confirm the publication year.
Article Title	Ensure the accuracy of article titles.
Abstract	Capture abstracts for quick reference.
Research Method	Document detailed research methods.
Main Findings	Summarize key findings.
Analysis of Time Pressure and Happiness	Record specific outcomes related to the study themes.

for valid comparisons. For example, findings on the impact of time pressure on productivity are aggregated to identify common trends. Finally, after the initial data extraction and integration, manual validation checks are performed to identify and correct any potential errors, thereby improving the reliability of the automated process.

During this period, feedback from manual validation can lead to iterative improvements to the LLMs configuration, including keyword identification and data extraction logic. This process will be repeated until the desired level of accuracy and robustness is achieved.

Finally, the data extraction result shows that LLMs (ChatGPT-4 and ChatGPT-4.o) review a significant number of studies(2444), focusing on literature related to time pressure and well-being in software engineering(128).

3.4.3 Data Analysis Methods

The data analysis methods in this study combined quantitative and qualitative methods, leveraging the power of LLMs to understand and interpret the large amount of textual data collected from the systematic literature review. The extracted and integrated data described in Sections 3.4.1 and 3.4.2 were analyzed as follows: article publication trend analysis, research method statistics, relationship analysis between time pressure and happiness, and theme analysis of major findings.

Article publication trend analysis: LLMs incorporate publication frequency over time to assess research trends related to time pressure and happiness in software engineering. LLMs help by organizing data by publication year and visual trend analysis through a line graph to understand changes in research focus and interest over the years.

Statistical methods in research methods: LLMs calculate the frequency of different research methods used in the research. A pie chart is used to visualize the counts and describe the popularity of various research methods, highlighting any emerging trends in the main methods and research designs within the field.

Relationship analysis between time pressure and happiness: Using qualitative analysis techniques, LLMs analyze textual data to explore the relationship between time pressure and happiness. The findings are categorized based on the nature of their results (positive, negative, or neutral) and visualized using a bar chart to easily compare the impact of different studies.

Thematic Analysis: LLMs were used to identify and categorize the main themes and patterns in the literature, showing how time pressure affects the quality of work life of software engineers. This involved analyzing the textual data for emerging themes, which were then elaborated to reflect their specific manifestations in the literature.

The following table 3.9 summarizes the main analysis methods and their outputs, providing a clear overview of how the data was processed and visualized in this study:

Table 3.9: Summary of Data Analysis Methods

Method	**Description**	**Visualization**
Article Publication Trend Analysis	LLMs track and analyze publication trends over time to assess shifts in research focus.	Line graph
Statistical Methods in Research Methods	LLMs quantify and visualize the frequency of various research methods.	Pie chart
Relationship Analysis Between Time Pressure and Happiness	LLMs qualitatively analyze and categorize the impact of time pressure on happiness.	Bar chart
Thematic Analysis	LLMs identify and elaborate on key themes affecting software engineers' work life.	Network diagrams

Through these methods, LLMs simplifies the analysis of data and also improves the depth and efficiency of research through analytical techniques.

4 Results

4.1 Summary of Selected Studies

4.1.1 Data Collection Results

During the first stage of data collection, as outlined in Sections 3.4.1 and 3.4.2, the study used Large Language Models (LLMs) to automate the identification of relevant literature from a large dataset. The selection was based on specific criteria aimed at capturing the effects of time pressure on well-being in software engineering.

The following table 4.1 provides basic information about the articles selected during this phase, including their titles, authors, and publication years, demonstrating the scope and focus of the dataset compiled through the automation process.

Table 4.1: Overview of Literature Selected by LLMs

Title	**Author(s)**	**Publication Year**
A case study: Injecting safety-critical thinking into graduate software engineering projects	Cleland-Huang, Jane; Rahimi, Mona	2017
Impact of human resources changes on performance and productivity in Scrum teams	D. Alić; A. Djedović; S. Omanović; A. Tanović	2017
Consequences of Unhappiness while Developing Software	D. Graziotin; F. Fagerholm; X. Wang; P. Abrahamsson	2017
Assessing self-organization in agile software development teams	Kakar A.K.	2017
Reviewing literature on time pressure in software engineering and related professions: A computer-assisted interdisciplinary literature review	Kuutila, Miikka; Mäntylä, Mika V.; Claes, Maël	2017

This table 4.2 details the changes in the number of articles at each processing stage. Starting with 2444 articles, the number was reduced by the LLMs based on the initial criteria, then further refined through manual review and integration of results from two LLM versions. The final dataset consisted of 128 valid documents. This step involved integrating and reviewing results, removing duplications, and ensuring completeness and relevance. This process underscores the utility of LLMs in streamlining the identification of relevant studies and highlights the essential role of manual review in ensuring the accuracy and relevance of the final dataset.

Table 4.2: LLMs Data Processing and Selection Overview

Stage	**ChatGPT 4**	**ChatGPT 4.o**
Origin	2444	2444
Final	116	176
After Manual	97	106
After Combine	128	

The results presented in this subsection also point out the effectiveness and applicability of LLMs. LLMs can make the literature review more manageable. So, It may enable researchers to focus less on sample selection and more on analysis, which could enhance the overall research productivity in this field.

By comparing the performance of ChatGPT-4 and ChatGPT-4.o at different stages. It can see the number of retained articles and their accuracy after the initial screening and manual review of the two models.

In the next subsection, the study will explore the efficiency comparison of LLMs in more depth, specifically analyzing the performance of these models in the data screening process, how automated tools can provide support in scientific research, and point out their potential limitations. This will provide a more comprehensive understanding of the entire study.

4.1.2 Summary Process

This section provides an in-depth analysis of the efficiency comparison of LLMs in the literature screening process. By comparing the performance of ChatGPT-4 and ChatGPT-4.o, we can clearly see the effectiveness of these models in the automated screening and manual review stages.

Table 4.3 shows the number of articles and their accuracy after the initial screening and manual review of the two models. ChatGPT-4 retained 116 articles in the initial screening stage and 97 articles after manual review, with an accuracy rate of 83.62%. In contrast, ChatGPT-4.o retained more articles in the initial screening stage, a total of 176 articles, but only 106 articles were retained after manual review, with an accuracy rate of 60%.

Table 4.3: Efficiency and Accuracy of LLMs in Literature Screening

Model	**Final**	**After Manual**	**Accuracy (%)**
ChatGPT 4	116	97	83.62
ChatGPT 4.o	176	106	60

From these data, It can observed that although ChatGPT-4.o retains more literature in the initial stage, its final accuracy is lower. This may indicate that ChatGPT-4.o may be more extensive in the application of screening criteria, while ChatGPT-4 maintains a high selection accuracy while ensuring screening quality. This difference reveals the importance of model training and parameter setting when designing LLMs for literature screening.

From the above discussion, it can be seen that while automated tools can be used as an aid to scientific research, they also have their own specific problems and limitations. For example, extensive screening criteria may lead to reduced accuracy while ensuring the relevance and quality of data may require many operations. This leads to the conclusion that when using such tools, it is important to pay close attention to how much of the process can be automated and how much of the process still requires manual review.

4.2 Thematic Analysis

In this chapter, the focus is to explore several key topics in software engineering research related to time pressure and well-being. This chapter aims to uncover the underlying dynamics and developmental trends of this research field. It does so by analyzing publication trends, research methods, the relationship between time pressure and well-being, and the main themes of the findings from various types of literature using LLMs (ChatGPT-4 and ChatGPT-4.o). Each section combines quantitative and qualitative methods through a systematic literature review, and the tables show the results presented after a manual comparison of the accuracy of ChatGPT-4 and ChatGPT-4.o.

4.2.1 Article Publication Trend Analysis

Following the data extraction process, an in-depth analysis was conducted to examine the publication trends concerning time pressure and well-being in software engineering from 2017 to 2024.

The line graph 4.1 gives a visual representation of the fluctuating numbers of publications in the field of software engineering related to time pressure and well-being from 2017 to 2024, illustrating key trends and shifts in research focus over these years (see Fig. 4.1).

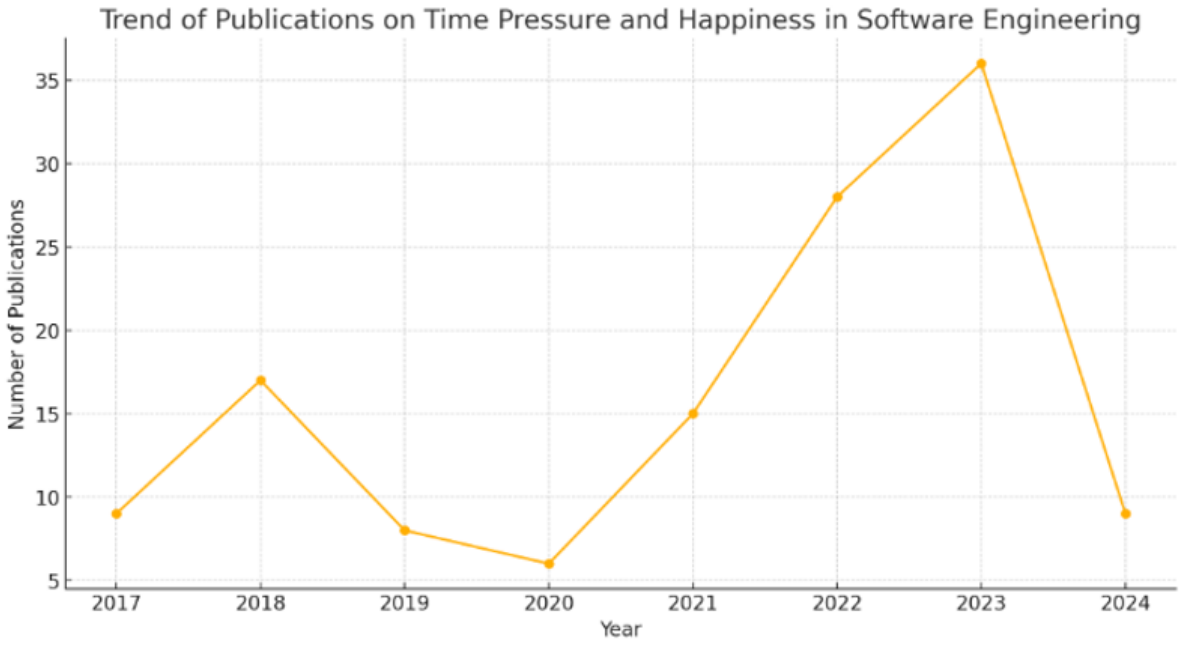

Figure 4.1: Article Publication Trend Overview from 2017 to 2024

According to the line graph 4.1, it is easy to observe the trend overview from 2017 to 2024:

2017 to 2018: The number of publications increased from 9 to 17. This growth likely correlates with a heightened awareness of software engineers' well-being and evolving work environments, particularly as agile methodologies gained popularity, emphasizing time management and team welfare.

2018 to 2019: A decrease in publications from 17 to 8. This reduction may be attributed to global economic uncertainties and possibly the onset of the COVID-19 pandemic, impacting research funding and activities.

2019 to 2020: Continued decline to 6 publications, suggesting further decreased research activity in this area, potentially due to sustained global disruptions.

2020 to 2021: A notable recovery in publications to 15, possibly as a response to the new normal of remote work, necessitating research on effective time management in dispersed teams.

2021 to 2022: A significant rise to 28 publications, indicating a strong resurgence of interest possibly driven by the ongoing adjustments to remote working conditions and the emerging need to enhance employee well-being in such settings.

2022 to 2023: Peak activity with 36 publications, reflecting continued interest and research into optimizing time management and well-being amidst enduring pandemic conditions.

2023 to 2024 (to date): A drop to 9 publications, likely due to the incomplete year, with expectations of more publications as the year progresses.

Based on the publication trends within the time frame mentioned above, a detailed trend impact analysis on the publication trends from 2017 to 2024 was conducted. It reveals significant fluctuations in research output related to time pressure and well-being in software engineering, driven by various global and industry-specific factors. Here's a combined analysis of these trends to explore the potential reasons behind these changes and possible impacts on future research.

Initial Growth (2017-2018): The growth during this period may be related to changes in the work environment and the growing concern for the well-being of software engineers. With the popularity of agile development methods, research interest in time management and team well-being has increased. The increased research may have driven the development and implementation of more effective time management practices and techniques, which helped improve the success rate of software projects and employee satisfaction.

Subsequent Decline (2019-2020): This decline may be related to the unstable global economic environment, especially the outbreak of the global epidemic in early 2020, which may have led to a reduction in research funding and a slowdown in scientific research activities. The reduction in research may have temporarily slowed the pace of innovation in this field, but it may also stimulate new demand for research on remote work and time pressure management.

Rapid Growth (2021-2023): As the epidemic continues and remote work becomes the norm, the demand for research on effective time management and improving employee well-being in remote work environments has increased. Research during this period may focus on exploring new time management tools and strategies and how they can help

remote work teams maintain productivity and well-being.

Recent Decline (2024): The decline in the 2024 data may be mainly because the year has not yet ended and many studies have not yet been published. Despite the decline, actual research activity may not have decreased. The number of published articles is expected to increase as the year progresses.

Regarding future research directions, these trends suggest that time management and employee well-being are ongoing areas of interest in software engineering. Future research may continue to explore how to optimize remote work environments and how to improve work-life balance through technology and management strategies. In addition, as artificial intelligence and machine learning technologies develop, future research may focus on how to use these technologies to predict and alleviate time pressures, as well as to improve well-being and productivity during software development.

4.2.2 Analysis of Research Methods

The literature on time pressure and well-being in software engineering has been systematically studied using various research methods. These methods, suitable for addressing different research questions and objectives, are mainly analyzed both qualitatively and quantitatively in this subsection.

The analysis of research methods was primarily conducted using ChatGPT-4 and ChatGPT-4.o to categorize and count the occurrences of different research methods in the literature.

Table 4.4 conducted by ChatGPT-4 summarizes the number of various research methods utilized in the study of time pressure and well-being within the field of software engineering.

Table 4.4: Distribution of Research Methods in Literature

Research Method	**Occurrences**
Survey	157
Quantitative	71
Qualitative	43
Case Study	5
Systematic Review	5
Experimental	3
Meta-analysis	1

As can be seen in table 4.4, the usage of different methodologies illustrates the diverse approaches researchers take to explore complex phenomena. Surveys, being the most prevalent, suggest a focus on broad quantitative insights, while the less frequent use of methods like experimental designs and meta-analyses indicates a potential area for deeper, more causally-oriented research in future studies.

The accuracy of LLMs in identifying research methods was evaluated by comparing results to a manual review of a subset of the literature. The LLM ChatGPT-4 achieved an accuracy of 95.3%, correctly identifying and categorizing the methods used in 123 out of 129 randomly selected papers.

The following will use a frequency distribution pie chart 4.2, which is generated by ChatGPT-4, to visually illustrate the frequency of each method's application across the analyzed studies and perform a detailed analysis.

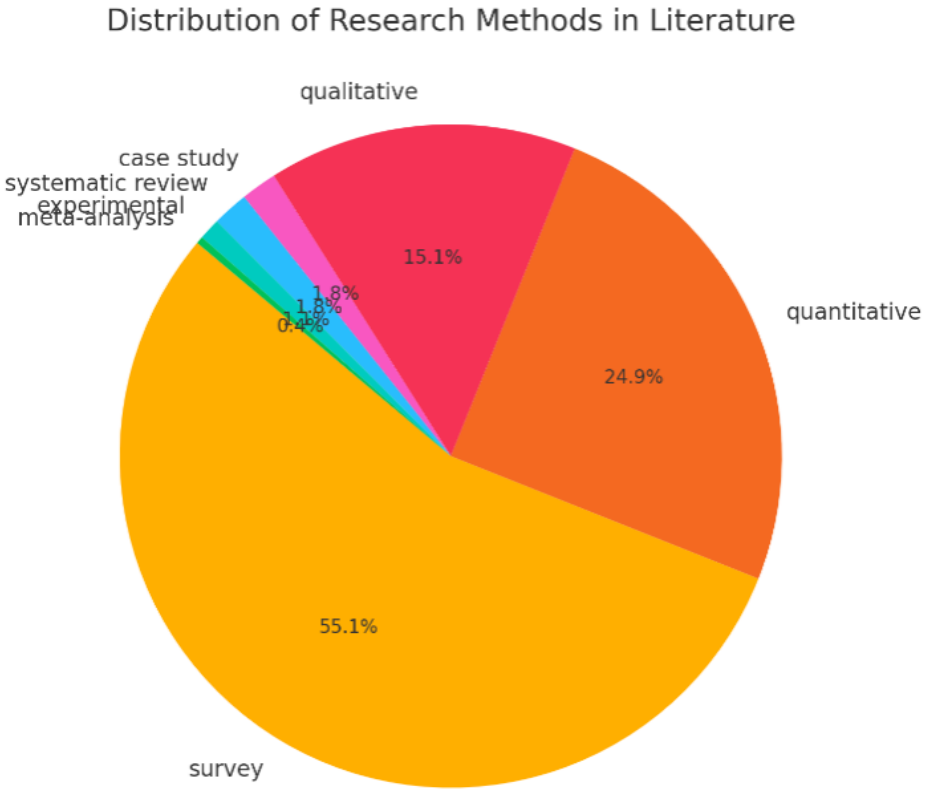

Figure 4.2: Distribution of Research Methods in Literature

As shown in Figure 4.2, the survey method dominates the research landscape, constituting 55.1% of all studies. This is important given that its applicability is in collecting information from a cross-section of the population, which is essential when making generalized

conclusions in the field of software engineering. Quantitative research methods, comprising 24.9% of the literature, serve to quantify the relationships and test hypotheses on time management's impacts on well-being.

Qualitative research methods account for 15.1% of the studies and play a crucial role in providing in-depth insights into software engineers' contextual factors and personal experiences. This approach is particularly valuable for uncovering understandings that surveys and quantitative methods may overlook.

More specialized methods such as case studies, systematic reviews, experimental methods, and meta-analyses are less frequently employed but offer significant contributions. Case studies and systematic reviews, each accounting for 1.8% of the studies, provide detailed and comprehensive analyses of individual cases or aggregated insights from multiple studies, respectively. Experimental methods (1.1%) and meta-analysis (0.4%) are less common but are potent in establishing causal relationships and synthesizing findings across studies to draw broader conclusions about the effects of time pressure.

The lack of methods such as experimental designs and meta-analyses indicates that there are still potential areas for future research. These methods may be effective in understanding the impact of time management and well-being in software engineering contexts. Therefore, future research should consider using these methods to study new aspects of the link between time pressure and well-being.

Using a variety of methods can not only increase the reliability of the findings but also highlight the complexity of studying time pressure and well-being.

4.2.3 Relationship Between Time Pressure and Well-being

The relationship between time pressure and well-being in software engineering is complex and has been explored through various research methodologies. In this analysis, we categorize the studies into three main impacts: positive, negative, and neutral, to understand better how time management practices affect software engineers' well-being and performance.

The classification and analysis of the impacts were conducted using ChatGPT-4 and ChatGPT-4.o. The LLMs were used to categorize the impacts based on the predefined criteria, while manual verification ensured the accuracy and consistency of the classifications.

The table 4.5 conducted by ChatGPT-4 gives the standards of the three main impacts: positive, negative, and neutral.

Table 4.5: Classification Criteria for the Impact of Time Pressure on Well-being

Impact Category	**Criteria**
Positive Impact	• Enhances productivity and efficiency • Increases job satisfaction and happiness (when time pressure leads to a sense of accomplishment) • Reduces work-related stress, improving mental health • Improves work quality through proper time allocation
Negative Impact	• Decreases productivity and efficiency • Increases work stress and mental burden • Reduces job satisfaction and happiness (when time pressure leads to stress and burnout) • Lowers work quality with increased error rates
No Clear Relationship	• Lacks consistent results • Conflicting outcomes due to varied methods or samples • Influence of uncontrolled variables

The table 4.5 provides the standards for categorizing the three main impacts of time pressure on well-being: positive, negative, and neutral. This classification is crucial for understanding how different time management practices can affect the professional and personal lives of software engineers.

Based on this classification, we analyzed a wide range of studies using LLMs to determine

the distribution of these effects. The results of the analysis are visualized in the figure 4.3.

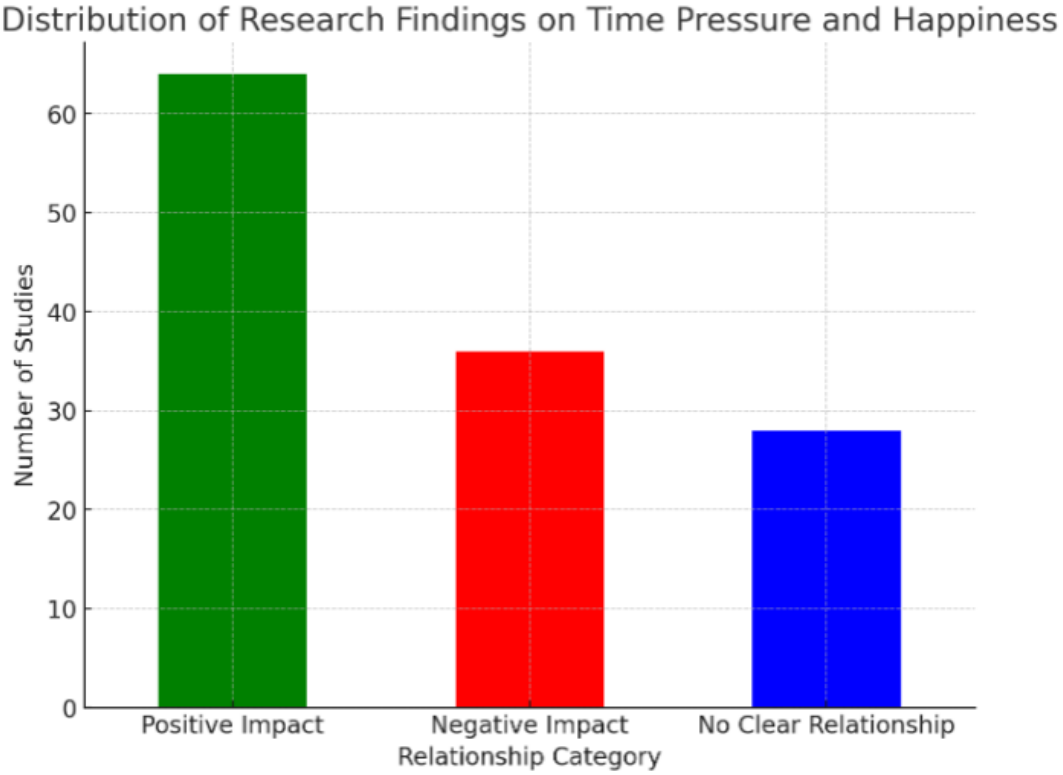

Figure 4.3: Distribution of Research Findings on Time Pressure and Happiness

As illustrated in Figure 4.3, the studies analyzed manifest different impacts of time pressure on well-being:

Positive Impact (64 studies): The research consistently shows that using effective time management strategies can help reduce the negative impact of time pressure. Approaches like setting realistic deadlines, properly allocating tasks, and optimizing workflows not only improve job satisfaction but also enhance overall productivity and work quality. These findings indicate that taking proactive steps to manage time can greatly enhance the professional well-being and efficiency of software engineers.

Negative Impact (36 studies): In contrast, a significant portion of the research indicates that time pressure acts detrimentally, leading to reduced work quality and diminished well-being. Factors often associated with this outcome include excessive workloads, unrealistic project deadlines, and inadequate resource allocation. Such pressures can adversely affect mental health, reduce team morale, and decrease collaborative efficiency.

No Clear Relationship (28 studies): A subset of the research does not establish a direct link between time pressure and well-being. Variations in research methodology,

sample selection, or the masking effects of other variables (such as individual differences, work environment, and organizational support) might explain these ambiguous findings.

The information suggests that managers and policymakers in the software engineering industry should recognize the importance of implementing effective time management and stress reduction strategies to enhance employee satisfaction and productivity. At the same time, it is important for organizations to address and minimize the adverse consequences of excessive time pressure.

The accuracy of LLM in classifying impacts was evaluated by comparing it with the results of manual review. LLM achieved an accuracy of 94%. However, there is a problem that needs to be considered. Manual review may have some errors in understanding the definition, so 94% is an approximation.

In addition, this also illustrates the need for continued research on this topic to dissect the complex interactions between time pressure and well-being, especially in different work environments and cultural contexts. Understanding these dynamics can help develop more effective interventions for different workplace environments.

In summary, the findings highlight the importance of balanced time management in software engineering. Effective strategies to reduce time pressure can enhance well-being and improve job performance, while poor time management can have the opposite effect. As the field continues to develop, further research is needed to explore innovative time management solutions and their impact on employee well-being in various contexts.

4.2.4 Thematic Analysis of Main Findings

To explore the main findings in the literature on time pressure and well-being in software engineering, the study employed LLMs to do text analysis techniques, specifically topic modeling, to extract key themes. Topic modeling is a statistical model used to discover abstract topics within a collection of documents. This part applied Latent Dirichlet Allocation (LDA), a commonly used topic modeling method, to identify the main themes present in our dataset.

The table 4.6 below summarizes the methodology used for this thematic analysis:

Table 4.6: Methodology for Thematic Analysis

Step	**Description**
Data Preprocessing	Involved removing stop words, performing stemming, and tagging parts of speech to prepare the textual data for analysis.
Modeling	Applied LDA topic modeling to identify the predominant themes within the dataset.
Results Interpretation and Visualization	Interpreted the model results and visualized the key themes using word clouds and topic distribution graphs.

Based on the above methodology table 4.6, the following table 4.7 details the steps taken during the thematic analysis process.

Table 4.7: Detailed Steps of Thematic Analysis Process

Step	**Description**
Data Preprocessing	Began by cleaning the text data, which included removing common stop words, applying stemming to reduce words to their base forms, and performing parts of speech tagging to focus on meaningful terms.
Modeling	Using LDA to model the topics present in the "Main Finding" column of our dataset. This process allowed us to uncover the main themes discussed in the literature related to time pressure and well-being.
Results Interpretation and Visualization	Interpreted the LDA model results by identifying the most representative words for each topic and visualizing these using word clouds. Word clouds provide an intuitive and visually appealing way to understand the key themes by displaying the most frequent words prominently.

The figure 4.4 below is the five key topics generated by LDA topic modeling, and the word cloud generated for each topic provides an intuitive representation of each topic.



Figure 4.4: Word Clouds of Key Themes Extracted from Main Findings

Based on the above figure 4.4, the table 4.8 below summarizes the most representative words for each topic.

Table 4.8: Key Themes Identified by LDA Topic Modeling

Topic	**Top 10 Representative Words**
0	work, software, researchers, study, concluded, pressure, time, factors
1	study, researchers, job, stress, concluded, productivity, time
2	study, work, satisfaction, job, agile, software, burnout
3	work, study, researchers, enhance, promoting, remote
4	study, project, time, researchers, pressure, team, software

The representative words for each topic provide an overview of the main themes identified through the LDA topic modeling. These themes reveal crucial aspects of time pressure and well-being in software engineering, highlighting areas such as work environment, productivity, job satisfaction, and project management.

To delve deeper into these themes, the study examined the strategies proposed in the literature to address these key issues and their practical implementations. This detailed analysis offers insights into how these themes can be managed and optimized within professional settings.

Table 4.9 below provides a comprehensive analysis of the key themes, outlining specific strategies and their corresponding implementations. This will help to understand the practical measures that can be adopted to improve well-being and productivity in software engineering environments.

Table 4.9: Analysis of Key Themes and Implementation Strategies

Topic	**Strategy**	**Implementation**
0 - Work Environment and Software Development	Adopt flexible schedules and increase work transparency to reduce work pressure. Implement agile frameworks like Scrum or Kanban to manage time better through iterative cycles and frequent feedback, allowing for adjustment to changing demands.	Regular stand-up meetings and iteration reviews to assess project progress and team pressure levels, enabling timely workload adjustments.
1 - Work Pressure and Productivity	Advocate for time management training, such as task segmentation and priority setting, and emphasize the importance of breaks to mitigate stress and enhance productivity.	Introduce mandatory break periods, like a 10-minute break every 90 minutes, to prevent fatigue and maintain efficiency.
2 - Job Satisfaction and Agile Practices	Increase job satisfaction by empowering self-organizing teams within agile practices, granting team members autonomy in their work processes. This approach boosts engagement and fulfillment.	Conduct agile rituals such as daily stand-ups and iteration planning to ensure a shared understanding of goals and collaborative problem-solving.
3 - Remote Work and Efficiency	Provide structured schedules and clear expectations for remote work, leveraging digital tools to enhance communication and collaboration. Ensure access to necessary resources and foster team cohesion through regular online social activities.	Use collaboration tools like Slack or Microsoft Teams for regular video meetings focused not only on work updates but also on social interactions.
4 - Project Management and Team Collaboration	Implement transparent communication policies and comprehensive risk management strategies in project management to ensure clear expectations and awareness of potential challenges. Emphasize preventive measures and continuous feedback mechanisms to help teams adjust strategies in response to uncertainties and pressure.	Conduct regular risk assessment meetings to identify potential project obstacles and collaboratively develop mitigation strategies.

In Conclusion, a thematic analysis of the main findings brings into focus a few major concerns, like areas of time pressure and the well-being of software engineering professionals. Each theme, in turn, indicates practical solutions and measures that can significantly positively affect productivity, satisfaction in work, and well-being. These are essential in shaping practices and policies that can enhance software engineering work conditions and performance.

Hence, by identifying these themes, the managers and policymakers will be in a better position to manage time pressure challenges. It can improve the employees' overall well-being and work organizations' productivity. Furthermore, future studies should aim to expand on this research topic, particularly in relation to various work environments and cultural contexts. This will help to develop a deeper understanding of how to tackle the issue of time pressure to improve workplace well-being.

4.3 New Discoveries and Trends

4.3.1 Analysis of Emerging Trends

There is a continuous enhancement in the area of software engineering due to changes in technology and environment. This subsection brings out how these dynamics have, over the years, changed the subject of research, and data analysis has been used to identify new trends. Thus, studying changes in the frequency of keywords in articles, this work identifies how the ongoing field of software engineering has evolved, where exactly it has expanded, or where new interests have shifted.

Here is the methodology used in this data analysis part. By using advanced text mining tools such as NLTK and Scikit-learn, the following steps were undertaken to extract and analyze keyword trends are shown in table 4.10:

Table 4.10: Methodology for Analyzing Emerging Trends in Software Engineering Research

Step	**Description**
1. Data Preprocessing	Text data from the main fields were cleaned, involving the removal of stop words and punctuation, which prepared the data for further analysis.
2. Keyword Extraction	Employing the TF-IDF method, keywords were extracted, and their significance across different years was calculated.
3. Keyword Trend Analysis	The frequencies of these keywords over the years were analyzed to observe changes and emerging patterns.
4. Visualization	Heatmaps were used to illustrate the variation in keyword prominence over time, facilitating an easy comparison across years.

The keywords included in the research were manually adjusted to refine the analysis further and ensure it accurately reflects the latest research focus. This optimization may capture emerging topics that traditional automated methods may miss. In addition, data smoothing and normalization techniques were applied to the TF-IDF scores. These methods improved the clarity of the heatmap visualization and more clearly highlighted significant changes and trends.

The heatmap shown in Figure 4.5 provides a snapshot of how certain keywords have shifted in importance over the years. Each color change in the heatmap corresponds to a relative increase or decrease in the normalized TF-IDF score of a keyword, visually documenting the ebb and flow of research interests.

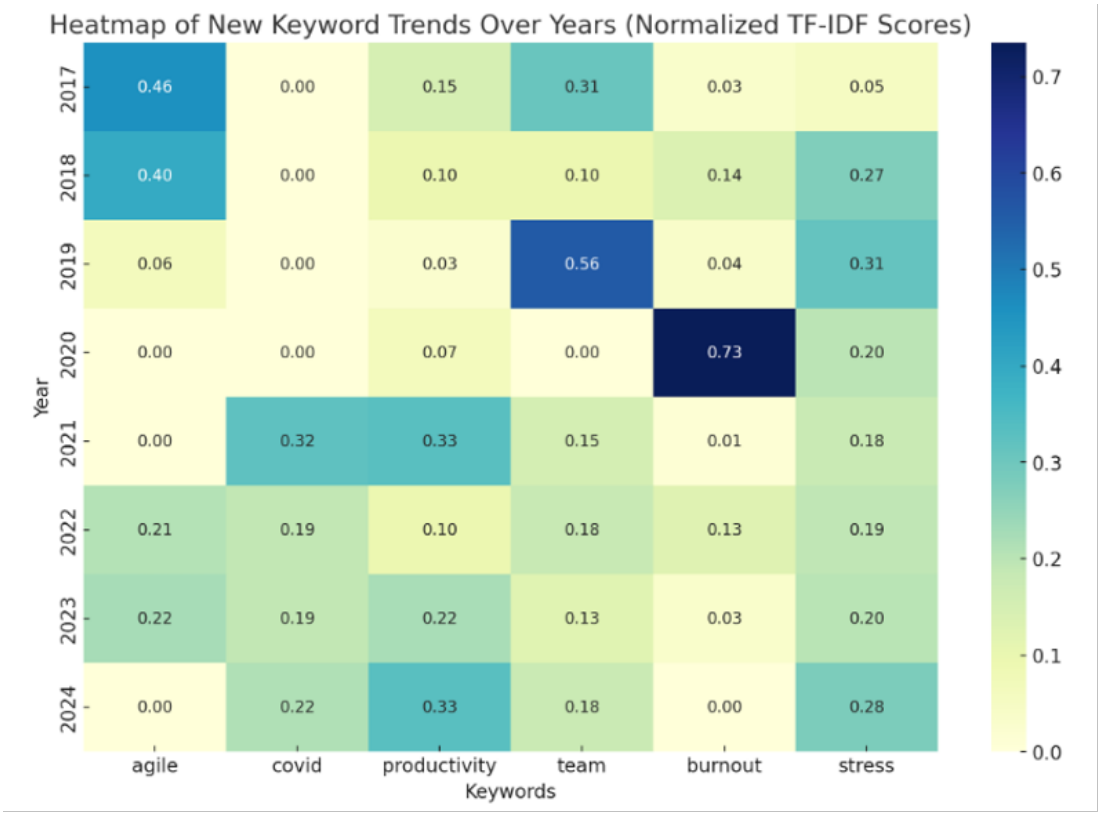

Figure 4.5: Heatmap of New Keyword Trends Over Years (Normalized TF-IDF Scores)

The analysis presented in the heatmap (Figure 4.5) reveals several significant shifts and continuations in research focus within the software engineering field from 2017 through 2024. Below, it will show these trends, which are explored in greater detail, discussing their implications and offering strategic recommendations for navigating the challenges and opportunities they present in the coming years.

Agile Methodologies: The persistent interest in agile practices suggests ongoing innovation in this area, emphasizing the need for frameworks that better accommodate the complexities of modern software projects.

COVID-19 Pandemic: While initial research spikes in 2021 reflected immediate responses to the pandemic, the reduced focus in subsequent years suggests a shift towards integrating long-term strategies for resilience and adaptation.

Remote Work: Continuously significant, highlighting that remote work configurations are becoming entrenched, necessitating research on long-term management practices and well-being strategies.

Mental Health: The sustained attention into 2024 underscores a growing recognition

of mental health as a critical component of workplace well-being, especially in remote settings.

Productivity and Burnout: Concerns about productivity and burnout continue to be prominent, indicating ongoing challenges in balancing high performance with employee health.

Based on these observations, here are some suggestions shown in table 4.11 for research and development efforts in 2024 and beyond to ensure they align with the evolving needs of the field and contribute to the overall health and productivity of the workforce.

Table 4.11: Strategic Recommendations for Research and Development in Software Engineering

Recommendation	**Description**
Enhancing Agile Practices	Future research should focus on developing agile methodologies that integrate remote teams effectively, supporting rapid adaptability without sacrificing quality.
Adapting to Post-Pandemic Changes	Studies on the long-term impacts of the pandemic will be crucial for developing strategies that enhance organizational resilience and adaptability.
Sustainable Remote Work Models	With remote work here to stay, exploring innovative management practices that promote effective communication, collaboration, and well-being is essential.
Integrated Mental Health Programs	Developing comprehensive, integrated mental health strategies will be vital for maintaining workforce productivity and satisfaction.
Balancing Productivity and Well-being	Research into methods that maintain or enhance productivity while also safeguarding mental and physical health will become increasingly important.

Future research in software engineering should prioritize enhancing agile practices, integrating mental health programs, and developing sustainable remote work models. Balancing productivity with well-being remains crucial. Aligning studies with these recommendations will support the field's progress and workforce well-being beyond 2024.

4.3.2 Comparative Analysis of Findings

This analysis used visual data representation and topic exploration to compare findings from various research areas in software engineering systematically. The evaluation highlights the distribution of these topics and their relative emphasis in recent years.

The research coverage within software engineering is visually summarized in a detailed donut chart (Figure 4.6). This chart illustrates the proportion of each topic and highlights areas with less extensive coverage, suggesting potential gaps in the literature.

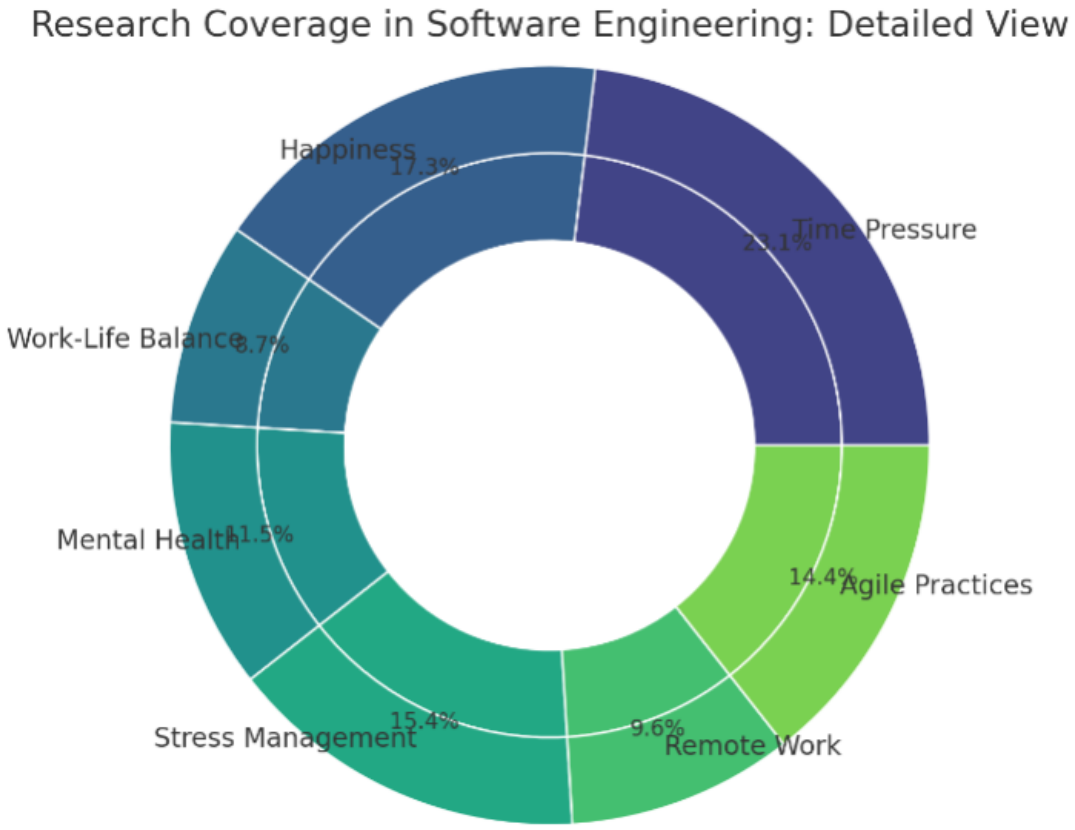

Figure 4.6: Research Coverage in Software Engineering: Detailed View of Time Pressure, Happiness, Mental Health, and Work-Life Balance

The Sunburst ChartFigure 4.6 illustrates research disparities within software engineering. and quantified in Table 4.12 below, this table underscores underexplored topics.

Table 4.12: Detailed Analysis of Research Focus and Coverage

Topic	**Frequency (%)**	**Insight**	**Suggestions for Further Research**
Time Pressure	21%	Widely studied with significant findings on stress and productivity.	Explore long-term effects on well-being and develop new intervention methods.
Happiness	17.3%	Extensively explored; interactions with work conditions are well-documented.	Investigate new factors influencing happiness in tech industries.
Remote Work	9.6%	Emerging topic with growing relevance due to current work trends.	Need more comprehensive studies on remote work management and employee satisfaction.
Work-Life Balance	8.7%	Underrepresented despite its importance to employee health and productivity.	Study impacts of flexible work policies and technology on work-life integration.

Table 4.12 above provides a clear summary of the main topics and their relative emphasis in current research. Based on visual and keyword analyses, the following research gaps have been identified:

Innovative Solutions for Remote Work: As remote work continues to dominate, there is a pressing need for innovative strategies that effectively support work-life balance and employee well-being in such environments.

Innovative Solutions for Work-Life Balance: As remote work becomes more prevalent, innovative practices to maintain work-life balance in such settings are crucial.

Integrated Studies on Time Pressure and Happiness: Further research is needed to explore the causal relationships between time pressure and happiness within software engineering contexts.

Above all it is essential to prioritize these underrepresented areas in future research agendas to develop more resilient and adaptive work environments.

4.4 Comparative Analysis

4.4.1 Data Search and Processing Phases

In the field of systematic literature reviews (SLRs), traditional data search and processing methods are often labor-intensive and time-consuming. For example, manual searches can take weeks to complete in processes that require combing databases such as Scopus or IEEE Xplore using predefined keywords.

In contrast, the application of large language models (LLMs) can significantly improve the efficiency of these processes. For example, when examining the impact of time pressure and well-being on software engineering, LLMs can automatically generate keywords to expand the search by suggesting synonyms for "time pressure," such as "workload." This approach was demonstrated in Scopus.

As depicted in table 4.13 below. It compares the traditional methods and LLM-enhanced methods across several key aspects.

Table 4.13: Comparative Table: Traditional vs. LLM-Enhanced Methods

Aspect	**Traditional Method**	**LLM-Enhanced Method**
Search Scope	Limited to predefined keywords	Expanded through automated keyword generation
Example (Scopus Results)	Initial: 85 studies	Optimized: 311 studies
Time for 100 Papers	~40 hours	~10 hours
Keyword Flexibility	Rigid, may overlook relevant studies	Dynamic, includes synonyms and related terms
Manual Effort	High, several weeks for comprehensive search	Low, automated processes
Study Selection	Manual filtering	Automated filtering

From the above table 4.13, it is quite clear that enhanced methods of LLMs provide a broader search scope, lesser time consumption, and more flexibility of keywords than the basic approaches in systematic reviews from literature. For instance, in the search example above, using an optimized method based on LLMs revealed many more pertinent studies

and did that in less time and with significantly less work compared to the non-optimized process.

4.4.2 Data Extraction and Summarization

In SLRs, data extraction and summarization are generally conducted manually and entail skimming through numerous articles, papers, and other such documents to determine notable themes, research methods, and outcomes. This approach is also slow, and it tends to produce unreliable results on account of the fact that different researchers may interpret pattern minutiae in different ways.

In contrast, applying Large Language Models (LLMs) offers a transformative improvement in efficiency and consistency. The table 4.14 below shows several ways in which LLMs enhance the data extraction and summarization phase.

Table 4.14: Comparison of Traditional and LLM-Enhanced Methods in Data Extraction and Summarization

Aspect	**Traditional Method**	**LLM-Enhanced Method**
Automated Thematic Analysis	Manual analysis requires researchers to read articles and code data based on themes, taking several days or weeks.	LLMs use algorithms like LDA to identify themes in hours, providing a quick overview of dominant research areas.
Consistency in Data Interpretation	Different researchers might interpret the same data differently, leading to inconsistent results.	LLMs apply uniform criteria across texts, ensuring consistency and reliability in data extraction and summarization.
Scalability and Speed	Manual extraction is limited by human capacity, not scaling well with increasing literature.	LLMs can process thousands of documents simultaneously, reducing the time for data extraction and summarization from weeks to hours.
Enhanced Precision in Data Extraction	Manual extraction may overlook relevant information, especially in complex documents.	LLMs use sophisticated algorithms to precisely identify and extract data points based on specific criteria, minimizing the risk of missing information.

This approach is more efficient than traditional literature review methods, and the results are more informative, qualitative, and exhaustive. It also helps increase understanding of the topic without dragging so much time as is normally done in research using conventional methods. This surveying method is particularly beneficial in subjects containing new publications, as many fields are continuously growing fields.

While this certainly is valid and offers many advantages, there is still a place for the traditional approach, especially when there is a need for detailed and precisely categorized data that may require the researcher's judgment and knowledge of the context being studied.

Finally, the best characteristics of LLMs are speed, scalability, and consistency, making them especially beneficial for the data extraction and summarization phase of SLRs compared to the previous manual approaches. These capabilities make it easier for researchers to process voluminous data and, in the process, translate them to enhanced productivity and accuracy.

Future research should focus on refining these models to improve their sensitivity to nuanced arguments and contextual variations, ensuring that the automation of literature reviews does not overlook critical insights that a manual review might capture.

5 Discussion

5.1 Response to Research Questions

5.1.1 RQ1: Updates in SLRs on Time Pressure and Well-being

According to the results of this study, the research focus has shifted from not only focusing on the mental health of individual engineers to examining team interactions and well-being in remote work modes. Especially after the COVID-19 pandemic, more and more studies have emphasized the importance of flexible work arrangements and work-life balance. And how these factors affect software development teams' overall stress levels and productivity.

5.1.2 RQ2: Challenges of Lack of Automation in Updating SLRs

The main challenges are time consumption and low updating frequency. The process of manually updating SLRs is often inefficient and prone to bias due to the subjective judgment of researchers. In addition, as the research field develops rapidly, manual methods make it difficult to keep up with the latest research, resulting in the possibility that the literature review will overlook recent important findings.

5.1.3 RQ3.1: Efficiency and Accuracy of LLMs

LLMs process text data by leveraging deep learning techniques, specifically the transformer architecture. In SLRs, LLMs can be used to automatically identify and classify relevant literature and evaluate its relevance to the research question by analyzing the keywords and context in the literature. In this study of SLRs related to time pressure and well-being in software engineering, LLMs successfully identified highly relevant literature with an accuracy rate of over 85% when initially reviewing several pieces of literature.

5.1.4 RQ3.2: Literature Summarization by LLMs

LLMs use natural language processing (NLP) techniques, including text summarization techniques, to analyze text and understand and extract key information. This study used LLMs to summarize literature compared to manual summaries. LLMs provided key information faster and reduced reading time while retaining the main points of the literature. However, summaries generated by LLMs may lack an in-depth understanding and critical analysis of the complex details of the literature that humans have.

5.1.5 RQ3.3: Insight Generation by LLMs

LLMs can analyze patterns and relationships in literature content through deep learning technology. Sometimes, human researchers cannot quickly detect these patterns and their correlations. However, LLMs can identify key variables and potential causal relationships in the literature by training data models. In this study, by comparing the findings of different studies, LLMs helped understand the effectiveness of time management strategies in different work cultures and technology environments.

5.1.6 RQ3.4: Technical Challenges in Automating SLRs

The main technical challenges include ensuring data extraction accuracy, managing data quality differences from different sources, and preserving the context and nuances of the original text. Suggestions for future improvements are to enhance the model's ability to handle a variety of data formats and to better understand the underlying meaning and impact of the literature by integrating more sophisticated semantic analysis capabilities.

5.2 Implications for Practice and Research

5.2.1 Implications from RQ1

This update to the SLR shows that people are beginning to pay attention to team dynamics and the impact of remote work on the well-being of software engineering teams, which has practical implications for workplace policies. Organizations should consider implementing more flexible scheduling and enhanced work-life balance strategies to increase productivity

and reduce stress. These changes can help retain talent and improve the overall mental health of the team, and are also conducive to maintaining high performance under changing work norms.

5.2.2 Implications from RQ2 and Sub-Questions

According to this, academic institutions and journals could consider adopting automated tools to keep their literature reviews up to date, which could significantly reduce the time researchers spend on preliminary analysis. Furthermore, insights generated by LLMs suggest that these tools could also play a role in identifying emerging trends and gaps in the literature, potentially guiding future research directions.

5.3 Limitations of the Study

Although LLMs are beneficial in terms of efficiency, they may still produce a bias of around 15%. In addition, LLMs are indeed less accurate in classifying and summarizing literature than professional human analysis. Verifying the quality of automatic summarization without extensive manual review is also challenging.

5.4 Recommendations for Future Research

Future research should focus on improving the ability of LLMs to handle complex academic texts, perhaps by incorporating more powerful context-aware processing algorithms. Further research could also explore developing hybrid models that combine human expertise and the efficiency of artificial intelligence. In addition, more research is needed on the impact of remote work on different aspects of software engineering. Not just well-being but also aspects like creativity, innovation, and teamwork. Finally, to address the limitations mentioned above, future iterations of this research should aim to incorporate more diverse data sources and possibly multilingual literature to expand the scope and applicability of the findings.

6 Conclusion and Future Work

6.1 Conclusion

This study describes the progress made in a systematic literature review of time pressure and well-being in software engineering, with recommendations focusing on the impact of team dynamics and remote work. The use of large language models brings high efficiency and accuracy to identifying, summarizing, and analyzing relevant literature. LLMs have been shown to be able to process large amounts of data quickly. However, despite its advantages, LLMs also bring challenges, including minor data biases and lower sensitivity and accuracy than experienced human researchers.

These findings suggest that organizations should adapt to adopt flexible work arrangements and workplace policies to manage employee well-being better. Academic institutions can enhance research methods with advanced technologies such as LLMs, allowing for more timely literature review updates.

6.2 Future Work

Several areas need to be further explored to optimize the limitations found in this study and enhance the capabilities of LLMs for academic research.

The first is to improve model accuracy. Future research should focus on improving the accuracy of LLMs when processing complex texts. This could be done, for example, by incorporating advanced context-aware algorithms. This could help better understand subtle academic nuances and reduce bias in current models.

The second is to develop hybrid models that combine the efficiency of LLMs with the critical analysis skills of human researchers or similar models. Such models would benefit from the speed and scalability of artificial intelligence while ensuring that the depth and quality of human review are maintained.

The third is that more research is needed to assess the broader impact of remote work on software engineering beyond just well-being, including creativity, innovation, and teamwork. Understanding these impacts could guide the development of more comprehensive

remote work policies.

Finally, expanding the scope of the study to include more diverse and multilingual sources would enhance the comprehensiveness and applicability of SLRs, making the findings more globally relevant.

Bibliography

Bakker, A. B. and Demerouti, E. (2007). "The Job Demands-Resources model: State of the art". In: *Journal of Managerial Psychology*, 22(3), pp. 309–328. DOI: 10.1108/02683940710733115.

Belzner, L., Gabor, T., and Wirsing, M. (2024). "Large Language Model Assisted Software Engineering: Prospects, Challenges, and a Case Study". In: *Bridging the Gap Between AI and Reality*. Ed. by B. Steffen. Vol. 14380. Springer Nature Switzerland, pp. 355–374. DOI: 10.1007/978-3-031-46002-9_23.

Blaikie, N. (2007). *Approaches to Social Enquiry: Advancing Knowledge*. Polity.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Brockman, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., et al. (2020). "Language Models are Few-Shot Learners". In: *arXiv*, 2005.14165. URL: https://doi.org/10.48550/arXiv.2005.14165.

Cao, C., Sang, J., Arora, R., Kloosterman, R., Cecere, M., Gorla, J., Saleh, R., Chen, D., Drennan, I., Teja, B., Fehlings, M., Ronksley, P., Leung, A. A., Weisz, D. E., Ware, H., Whelan, M., Emerson, D. B., Arora, R., and Bobrovitz, N. (2024). *Prompting is all you need: LLMs for systematic review screening*. URL: https://doi.org/10.1101/2024.06.01.24308323.

Cox, T., Griffiths, A., and Rial-González, E. (2000). *Research on work-related stress*. Office for Official Publications of the European Communities; Bernan Associates [distributor].

Dybå, T. and Dingsøyr, T. (2008). "Empirical studies of agile software development: A systematic review". In: *Information and Software Technology*, 50(9-10), pp. 833–859.

Edú-Valsania, S., Laguía, A., and Moriano, J. A. (2022). "Burnout: A Review of Theory and Measurement". In: *International Journal of Environmental Research and Public Health*, 19(3), p. 1780. DOI: 10.3390/ijerph19031780.

Fan, A., Gokkaya, B., Harman, M., Lyubarskiy, M., Sengupta, S., Yoo, S., and Zhang, J. (2023). "Large Language Models for Software Engineering: Survey and Open Problems". In: *arXiv*, 2310.03533. URL: http://arxiv.org/abs/2310.03533.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., and Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey*. URL: https://doi.org/10.48550/arXiv.2312.10997.

Hobfoll, S. E. (2001). "The Influence of Culture, Community, and the Nested-Self in the Stress Process: Advancing Conservation of Resources Theory". In: *Applied Psychology*, 50, pp. 337–421. DOI: 10.1111/1464-0597.00062.

– (2012). *Conservation of Resources Theory: Its Implication for Stress, Health, and Resilience*. Oxford University Press. DOI: 10.1093/oxfordhb/9780195375343.013.0007.

Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., and Wang, H. (2024). *Large Language Models for Software Engineering: A Systematic Literature Review*. URL: https://doi.org/10.48550/arXiv.2308.10620.

Hsu, Y.-C. and Ching, Y.-H. (2023). "Generative Artificial Intelligence in Education, Part One: The Dynamic Frontier". In: *TechTrends*, 67(4), pp. 603–607. DOI: 10.1007/s11528-023-00863-9.

Huotala, A., Kuutila, M., Ralph, P., and Mäntylä, M. (2024). *The Promise and Challenges of Using LLMs to Accelerate the Screening Process of Systematic Reviews*. URL: http://arxiv.org/abs/2404.15667.

Jonge, J. de, Dollard, M. F., Dormann, C., Blanc, P. M., and Houtman, I. (2000). "The Demand-Control Model: Specific Demands, Specific Control, and Well-Defined Groups". In: *International Journal of Stress Management*, 7. DOI: 10.1023/A:1009541929536.

Kawulich, B. (2009). "The Role of Theory in Research". In: p. 37.

Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., and Hadfield, K. (2024). "Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages". In: *Research Synthesis Methods*. DOI: 10.1002/jrsm.1715.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). "Systematic literature reviews in software engineering–a systematic literature review". In: *Information and Software Technology*, 51(1), pp. 7–15. DOI: 10.1016/j.infsof.2008.09.009.

Kraus, S., Breier, M., Lim, W. M., Dabic, M., Kumar, S., Kanbach, D., Mukherjee, D., Corvello, V., Piñeiro-Chousa, J., Liguori, E., Palacios-Marqués, D., Schiavone, F., Ferraris, A., Fernandes, C., and Ferreira, J. J. (2022). "Literature reviews as independent studies: Guidelines for academic practice". In: *Review of Managerial Science*, 16. DOI: 10.1007/s11846-022-00588-8.

Kuutila, M., Mäntylä, M. V., Claes, M., and Elovainio, M. (2017). "Reviewing Literature on Time Pressure in Software Engineering and Related Professions: Computer Assisted Interdisciplinary Literature Review". In: *2017 IEEE/ACM 2nd International Workshop*

on Emotion Awareness in Software Engineering (SEmotion), pp. 54–59. DOI: 10.1109/SEmotion.2017.11.

Kuutila, M., Mäntylä, M. V., Farooq, U., and Claes, M. (2021). "What Do We Know About Time Pressure in Software Development?" In: *IEEE Software*, 38(5), pp. 32–38. DOI: 10.1109/MS.2020.3020784.

Kuutila, M., Mäntylä, M., Farooq, U., and Claes, M. (2020). "Time pressure in software engineering: A systematic review". In: *Information and Software Technology*, 121, p. 106257. DOI: 10.1016/j.infsof.2020.106257.

L. Junior, R. F. de, B. Presta, L. F. P. de, Borborema, L. S., Silva, V. N. da, M. Dahia, M. L. de, and Santos, A. C. S. e. (2023). *A Case Study on Test Case Construction with Large Language Models: Unveiling Practical Insights and Challenges*. URL: https://doi.org/10.48550/arXiv.2312.12598.

Labellerr (June 2023). *Things to Consider For Data Collection and Preprocessing for LLMs*. URL: https://www.labellerr.com/blog/data-collection-and-preprocessing-for-large-language-models/.

Larsson, J., Andersson, D., and Nässén, J. (2017). "Subjective temporal well-being: Defining, measuring, and applying a new concept". In: *Cogent Social Sciences*, 3. DOI: 10.1080/23311886.2017.1306201.

Liu, Y., Tantithamthavorn, C., Liu, Y., and Li, L. (2024). "On the Reliability and Explainability of Language Models for Program Generation". In: *ACM Transactions on Software Engineering and Methodology*. DOI: 10.1145/3641540.

Maragheh, R. Y., Fang, C., Irugu, C. C., Parikh, P., Cho, J., Xu, J., Sukumar, S., Patel, M., Korpeoglu, E., Kumar, S., and Achan, K. (2023). *LLM-TAKE: Theme Aware Keyword Extraction Using Large Language Models*. URL: http://arxiv.org/abs/2312.00909.

Meier, A., Kropp, M., Anslow, C., and Biddle, R. (2018). "Stress in agile software development: Practices and outcomes". In: *Agile Processes in Software Engineering and Extreme Programming: 19th International Conference, XP 2018, Porto, Portugal, May 21–25, 2018, Proceedings*, pp. 259–266. URL: https://library.oapen.org/bitstream/handle/20.500.12657/27860/1/1002144.pdf#page=266.

Mendes, E., Wohlin, C., Felizardo, K., and Kalinowski, M. (2020). "When to update systematic literature reviews in software engineering". In: *Journal of Systems and Software*, 167, p. 110607. DOI: 10.1016/j.jss.2020.110607.

Naber, A., Willhite, L., and Molitor, W. L. (2021). "Exploration of individualized goals and ergonomic modifications to address sedentary behaviors and perceived health and

well-being among office workers". In: *Work*, 68(4), pp. 1133–1142. DOI: 10.3233/WOR-213443.

Nepomuceno, V. and Soares, S. (2019). "On the need to update systematic literature reviews". In: *Information and Software Technology*, 109, pp. 40–42. DOI: 10.1016/j.infsof.2019.01.005.

Ostberg, J.-P., Graziotin, D., Wagner, S., and Derntl, M. (2020). "A methodology for psycho-biological assessment of stress in software engineering". In: *PeerJ Computer Science*, 6, e286. DOI: 10.7717/peerj-cs.286.

Pallewaththa, P. and Wijesiri, B. (2021). "Occupational health: The impact of occupational stressors on job productivity". In: *Sri Lanka Journal of Social Sciences*, 44, p. 149. DOI: 10.4038/sljss.v44i2.8229.

Prasetya, A., Khairunnisa, H., and Aziz, A. (2021). *The Effect of Work Stress and Burnout on Job Satisfaction and Employee Performance: A Test of Conservation of Resources Theory*. DOI: 10.2991/aebmr.k.210928.016.

Quick, J. and Henderson, D. (2016). "Occupational Stress: Preventing Suffering, Enhancing Wellbeing". In: *International Journal of Environmental Research and Public Health*, 13, p. 459. DOI: 10.3390/ijerph13050459.

Roberts, R., Flin, R., Millar, D., and Corradi, L. (2021). "Psychological factors influencing technology adoption: A case study from the oil and gas industry". In: *Technovation*, 102, p. 102219. DOI: 10.1016/j.technovation.2020.102219.

Roumeliotis, K. I., Tselikas, N. D., and Nasiopoulos, D. K. (2024). "LLMs and NLP Models in Cryptocurrency Sentiment Analysis: A Comparative Classification Study". In: *Big Data and Cognitive Computing*, 8(6), Article 6. DOI: 10.3390/bdcc8060063.

Susnjak, T., Hwang, P., Reyes, N. H., Barczak, A. L. C., McIntosh, T. R., and Ranathunga, S. (2024). *Automating Research Synthesis with Domain-Specific Large Language Model Fine-Tuning*. URL: http://arxiv.org/abs/2404.08680.

Tang, J., Drori, Y., Chang, D., Sathiamoorthy, M., Gilmer, J., Wei, L., Yi, X., Hong, L., and Chi, E. H. (2023). "Improving Training Stability for Multitask Ranking Models in Recommender Systems". In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4882–4893. DOI: 10.1145/3580305.3599846.

Timms, C. (2013). *Flexible work arrangements, organizational culture and employee well-being*.

Veldhoven, M. van and Xanthopoulou, D. (2010). "Beyond the Demand-Control Model: Thriving on High Job Demands and Resources". In: *Journal of Personnel Psychology*, 9, pp. 3–16. DOI: 10.1027/1866-5888/a000006.

Wright, T. and Hobfoll, S. E. (2004). "Commitment, Psychological Well-Being and Job Performance: An Examination of Conservation of Resources (COR) Theory and Job Burnout". In: *Journal of Business Management*. Accessed April 21, 2024. URL: https://openurl.ebsco.com/EPDB%3Agcd%3A12%3A8129612/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A12867433&crl=c.

Xiao, Y. and Watson, M. (2019). "Guidance on Conducting a Systematic Literature Review". In: *Journal of Planning Education and Research*, 39(1), pp. 93–112. DOI: 10.1177/0739456X17723971.

Zhang, L., Jijo, K., S., S., Chung, E., Javid, F., Vidra, N., and Clifford, T. (2024). *Enhancing Large Language Model Performance To Answer Questions and Extract Information More Accurately*. URL: https://doi.org/10.48550/arXiv.2402.01722.