



Emotional Strain and Frustration in LLM Interactions in Software Engineering

Downloaded from: <https://research.chalmers.se>, 2026-05-15 08:30 UTC

Citation for the original published paper (version of record):

Martinez Montes, C., Khojah, R. (2025). Emotional Strain and Frustration in LLM Interactions in Software Engineering. Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering (EASE 2025): 193-204.
<http://dx.doi.org/10.1145/3756681.3756951>

N.B. When citing this work, cite the original published paper.



PDF Download
3756681.3756951.pdf
02 February 2026
Total Citations: 0
Total Downloads: 152

Latest updates: <https://dl.acm.org/doi/10.1145/3756681.3756951>

RESEARCH-ARTICLE

Emotional Strain and Frustration in LLM Interactions in Software Engineering

CRISTINA MARTINEZ MONTES, Chalmers University of Technology, Gothenburg, Vastra Gotaland, Sweden

RANIM KHOJAH, Chalmers University of Technology, Gothenburg, Vastra Gotaland, Sweden

Open Access Support provided by:
Chalmers University of Technology

Published: 17 June 2025

[Citation in BibTeX format](#)

EASE '25: Evaluation and Assessment in Software Engineering
June 17 - 20, 2025
Istanbul, Turkiye

Emotional Strain and Frustration in LLM Interactions in Software Engineering

Cristina Martinez Montes
Chalmers University of Technology
Gothenburg, Sweden
University of Gothenburg
Gothenburg, Sweden
montesc@chalmers.se

Ranim Khojah
Chalmers University of Technology
Gothenburg, Sweden
University of Gothenburg
Gothenburg, Sweden
khojah@chalmers.se

Abstract

Large Language Models (LLMs) are increasingly integrated into various daily tasks in Software Engineering, such as coding and requirement elicitation. Despite their various capabilities and constant use, some interactions can lead to unexpected challenges (e.g. hallucinations or verbose answers) and, in turn, cause emotions that develop into frustration. Frustration can negatively impact engineers' productivity and well-being if it escalates into stress and burnout. In this paper, we assess the impact of LLM interactions on software engineers' emotional responses, specifically strains, and identify common causes of frustration when interacting with LLMs at work. Based on 62 survey responses from software engineers in industry and academia across various companies and universities, we found that a majority of our respondents experience frustrations or other related emotions regardless of the nature of their work. Additionally, our results showed that frustration mainly stemmed from issues with correctness and less critical issues, such as adaptability to context or specific format. While such issues may not cause frustration in general, artefacts that do not follow certain preferences, standards, or best practices can make the output unusable without extensive modification, causing frustration over time. In addition to the frustration triggers, our study offers guidelines to improve the software engineers' experience, aiming to minimise long-term consequences on mental health.

Keywords

Software Engineering, Large Language Models (LLMs), Frustration, Emotions

ACM Reference Format:

Cristina Martinez Montes and Ranim Khojah. 2025. Emotional Strain and Frustration in LLM Interactions in Software Engineering. In *Evaluation and Assessment in Software Engineering (EASE '25)*, June 17–20, 2025, Istanbul, Turkiye. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3756681.3756951>

1 Introduction

Software Engineering (SE) comes with many challenges, from fixing bugs to dealing with changing requirements. Recently, Large Language Models (LLMs) and LLM-powered chatbots like ChatGPT

and GitHub Copilot have been used by software engineers to assist them in performing various tasks, including code generation, and quality assurance [1, 32]. Current research focuses on understanding how practitioners aim to increase their productivity and make their work process more efficient by targeting LLMs to automate the generation of software artefacts or receive guidance on how to solve certain problems [27, 54].

Challenges and limitations of LLMs hinder their effectiveness, such as unhelpful responses, which can lead to frustration among engineers [59]. This frustration contributes to techno-stress, affecting their workflow, well-being, and productivity [40, 55]. While frustrations have gotten little attention in LLM research for software engineering, we argue that understanding the causes of frustrations when using LLMs for software-related tasks is the first step to minimising them and thus improving the productivity and well-being of practitioners in the SE industry and academia. Moreover, revealing such triggers helps the designers of LLM-powered tools (e.g., AI Chatbots) improve the user experience and evolution of such tools.

This exploratory study aims to empirically investigate the causes of frustrations in software engineers' interactions with LLMs (and LLM-powered chatbots) and propose strategies for improvement. We focus on the following research questions:

RQ1: What are the triggers or sources of frustration of software engineers when using LLMs?

Our study presents four main categories that can cause frustrating emotions for the software engineer. We found that the main cause of frustration is when the software engineers receive an unhelpful or incorrect answer, followed by misunderstanding the intention, failing to meet personal preferences, and other limitations of LLMs. We argue that these categories are specific to SE since it is a domain that requires precise, context-aware, and technically accurate responses that can be directly applied, like using LLMs for code generation.

RQ2: How can the frustrating experience impact motivation?

We report that while frustration from unmet expectations can momentarily impact motivation, software engineers typically remain engaged in tasks despite these frustrating interactions with LLMs. This suggests that such interactions are not usually disruptive enough to prevent task completion.

RQ3: How can the user experience be improved to reduce frustration for software engineers?

We provide recommendations for improvements based on software engineers' expectations and lived experiences. Engineers offer practical, grounded recommendations that reflect user needs and



This work is licensed under a Creative Commons Attribution 4.0 International License. *EASE '25, Istanbul, Turkiye*

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1385-9/25/06

<https://doi.org/10.1145/3756681.3756951>

are in the form of interventions [39], [46] or looking for the causes [62].

Furthermore, users' expectations when using technology, particularly LLMs, are crucial to their frustration. Studies have shown that prior experiences with technology can shape users' expectations, influencing their perceptions of performance and trust [13]. When technology does not meet users' expectations, whether based on past interactions or external portrayals, frustration can escalate, hindering effective task completion and satisfaction [51]. By examining these responses, we aim to inform LLM design improvements that enhance productivity and reduce technostress.

2.3 Emotions in Software Engineering Tasks

Emotions related to software engineering tasks have been widely researched. Several studies have identified a wide range of emotions, from anger and frustration to joy and satisfaction, as software engineers perform their tasks and communication channels within the development context [16, 18, 23, 33–35, 48]. Sánchez-Gordón's [50] literature review further analysed the diversity of emotions developers face, identifying 40 discrete emotions, the most frequent being anger, fear, disgust, sadness, joy, love, and happiness. Various situational and contextual factors shape this rich diversity of emotions. Additionally, the impact of affective states has been related to performance and code quality. For example, Graziotin et al. [22] found that positive emotions, such as happiness, are closely linked to improved performance and productivity. Conversely, negative emotions, such as frustration and unhappiness, can reduce motivation, hinder task completion, and increase the likelihood of turnover [21]. Furthermore, studies have found that developers report higher productivity when in a state of flow and often experience frustration due to being stuck, technical difficulties or unfulfilled information needs [19, 41].

Moreover, specific triggers of negative emotions have been extensively studied. For example, unhappiness, Graziotin et al. [20] found that everyday sources of unhappiness are time pressure, bad code quality, repetitive tasks, and inadequate decision-making. Regarding frustration, Ford and Parnin [15] identified program comprehension challenges, poor tooling, and fear of failure as common causes.

Despite significant research on emotions associated with traditional software engineering tasks, there is still a gap in investigating the emotional impact of interactions with emerging tools like Large Language Models (LLMs). This gap leaves unexplored how emotions such as frustration evolve in response to LLM use, which has critical implications for improving their design and effectiveness in supporting software engineering tasks.

3 Methodology

Our study aims to understand software engineers' frustrations and emotions when interacting with LLMs, identifying causes and potential solutions. This study implements an exploratory design to gather initial insights and detect interaction patterns [52]. We used surveys following Stol and Fitzgerald guidelines [53] since they are particularly effective for exploratory studies aiming to generalise findings across a population. We collected qualitative and quantitative data through open and closed questions, covering

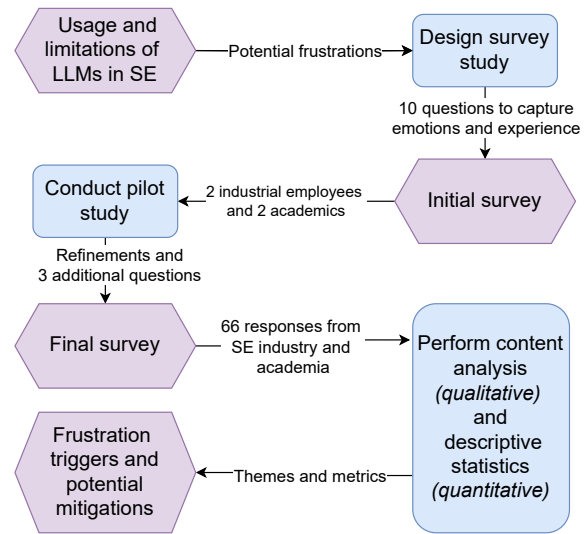


Figure 2: The exploratory study process that we followed to design our survey and analyse the responses qualitatively and quantitatively.

specific and broader aspects of LLM usage. Figure 2 illustrates our methodology.

3.1 Target Population

We surveyed software engineers in academia and industry to get a broader exploration of how LLMs are used in software engineering tasks. While industry engineers provided information about practical applications such as coding, debugging, and meeting production demands, academics, on the other hand, added LLM usage in research-related tasks such as programming, testing and general research on software processes and quality. Despite differences in their environments, both groups share core practices and challenges, such as dealing with tool limitations and managing cognitive load. Including both populations ensured a comprehensive analysis of how LLMs impact software engineering workflows, making our findings broadly relevant to diverse real-world and research contexts.

3.2 Data Collection

We used a questionnaire as our main data collection method. We designed the survey questions based on previous research about the usage of LLMs in SE and how it impacts their experience and productivity [27] as well as general limitations of LLMs [24].

After a pilot study with two PhD students (academia) and two software engineers (industry), we added more questions and refined the survey (available in our Zenodo package [37]). The survey started asking for consent to participate and included (i) four demographic questions and queries about participants' experience and familiarity with LLMs, (ii) two open-ended and one closed-ended question about their usage of LLMs during their work and their

expectations from these LLMs, (iii) four open-ended questions and one closed-ended question about the participants' emotions when certain expectations are not met and about frustrations specifically that the participants experienced when using LLMs, and (iv) five Likert-scales of the level of importance of different LLMs abilities and aspects to minimise frustrating experiences.

The survey was created on Google Forms and distributed across social platforms (LinkedIn and Facebook). We used stratified random sampling to get software engineers from academia and industry, drawing on Baltes and Ralph's [4] work as a framework for our sampling approach. We sent approximately 20 personal invitations to employees across 10 software organisations of different sizes (Startups and large companies) and domains (e.g., automotive and eLearning). We also invited researchers and academics in SE conferences (RE'24 and FSE'24) to allow our sample of software engineers to be diverse in terms of countries and domains. All data was anonymised, and we did not ask for personal data or identification. We followed our university's ethical regulations and guidelines. We believe that it is important that our sample was diverse in terms of areas and domains, particularly that it included software engineering academics. Software engineering practices often overlap between academia and industry, with some differences in priorities and contexts. This diversity enabled us to capture a broader range of perspectives and demonstrate that academics and practitioners share emotions and challenges related to using LLMs in the field.

3.3 Data Analysis

We used content analysis with an inductive approach following the steps by Erlingsson and Brysiewicz [11] to analyse the open questions. Both authors carried out the whole data analysis together systematically. The analysis started by reviewing each answer in detail and discussing it to ensure a shared understanding. This allowed us to identify initial patterns and codes. The codes were then categorised based on their similarities and differences [11], with the categories refined iteratively to ensure accuracy. Themes emerged organically from the data, reflecting the participants' perspectives and providing meaningful insights. We used the emotion classification and feeling wheel by Willcox [61] to categorise and identify the range of emotions expressed by participants. For example, one participant's comment, "I acknowledge that it might give incorrect answers so it is indifferent for me unless it happens often" was coded as Indifferent'. Enabling a deeper understanding of their emotional responses during interactions with LLMs. For the Likert and closed questions, we use descriptive statistics and visualisations.

4 Results

This section presents the results of the visualisation of the closed questions and the content analysis of the open questions. We distinguished between academics and practitioners when their results differed, such as in LLM usage. We combined the results when their patterns were similar, like emotional responses or frustration triggers.

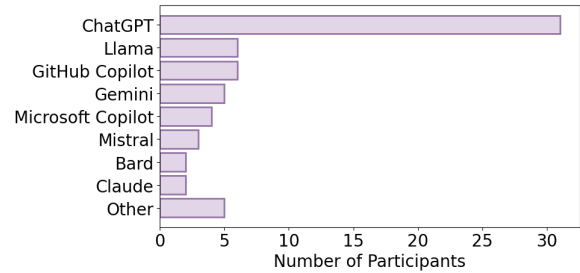


Figure 3: LLMs used by our participants.

4.1 Respondents Demographics

Our survey sample included software engineers in diverse roles (see Table 1) with a median age of 32. Participants represented organisations from seven countries across three continents, spanning aviation, automotive, game design, infotainment, eLearning, cybersecurity, telecommunications, trade, and SE research and education domains.

Table 1: Participants' areas and roles.

Area	Roles	# Participants	Total
Academia	PhD Student	15	27
	Researcher	7	
	Professor	5	
Industry	Software Developer	10	35
	Software Engineer	6	
	Manager	5	
	AI Engineer	4	
	Researcher	3	
	Tech Lead	2	
	Software Designer	2	
	Software Tester	1	
	Application Specialist	1	
	Applied Scientist	1	

Most participants (58 of 62) described themselves as "familiar" or "very familiar" with LLMs. Figure 3 shows the range of LLMs they use at work, with ChatGPT being the most popular.

4.2 Usage of LLMs in Software Engineering Industry and Academia

The results show that 56 participants (94.9%) use LLMs occasionally, out of which 38 participants (66%) use them on a weekly or daily basis. Only 3 (5.1%) participants indicated that they rarely use LLMs at work. Table 2 shows the tasks for which academia and industry respondents apply them. We considered more than one answer per participant.

In **industry**, respondents use LLMs for **programming tasks** like code generation, debugging, and optimisation. They also employ LLMs for **creative and communication tasks**, such as drafting emails and brainstorming ideas, and for generating and improving text. Additionally, LLMs help users on **learning new technologies and research** by providing starting points, best practices, and summaries of lengthy information. Lastly, respondents view LLMs as digital assistants for **task management and problem-solving**, streamlining workflows and enhancing productivity.

Table 2: LLM Usage Across Industrial and Academic Tasks

Usage Group	Area	Total Count
Writing Tasks	Academia	13
Programming Tasks	Academia	8
Educational Tasks	Academia	8
Research Related tasks	Academia	5
Non-Critical Tasks	Academia	5
Programming Tasks	Industry	18
Written Communication Tasks	Industry	10
Learning New Concepts	Industry	8
Task Management	Industry	5

In **academia**, LLMs are primarily used for **writing tasks**, including generating drafts, checking grammar, and providing content suggestions. Users find them helpful for managing busy work, such as email writing and idea generation, and for creating initial drafts for refinement. Additionally, participants view LLMs as **educational tools**, using them to understand new technologies and programming concepts or to assist in teaching. For **programming tasks**, LLMs help write simple code, debug, and learn new coding concepts, offering initial code snippets and quick insights into technologies.

LLMs are also employed for **research-related tasks** such as summarising academic papers, generating ideas, and finding references. They assist in translating data, cleaning datasets, and extracting key information from research.

4.3 Emotions During LLM Interaction

Due to the complexity of the emotional responses to unexpected LLM interactions, participants often described multiple, layered feelings in their experiences. We mapped these feelings described to Willcox’s emotions wheel (see Figure 1) to assess on a more fine-grained level how evolved the emotions were. Following this framework, we could trace how initial feelings, such as anger (first-level emotion), might progress into more specific emotions, like annoyance (second level) and then frustration (third level), and quantify how often these emotions occurred at each stage.

In Figure 4, we show the different emotions that were expressed by our participants for each category that maps to Willcox’s emotions wheel. For instance, we found that the most common emotions (54.6%) are frustration or emotions that can develop into frustration, such as annoyance and anger. This poses potential challenges to well-being and smooth workflows, with a risk of cumulative emotional strain over time.

Many respondents (27.8%) have also reported sadness-related emotions such as disappointment, indifference, or even guilt. In contrast to anger-related emotions, where respondents primarily blamed the LLM for its limitations, those who felt disappointed, sad, or guilty often turned the blame inward and criticised themselves for not being able to write the right prompt or meet their own expectations. When expectations were lower, the disappointment turned into indifference.

Participant: 29 “Knowing how LLM[s] work, I typically have lower expectations. So I [don’t] feel as frustrated or disappointed, particularly if I know that the task I asked is not trivial.”

Such expectations come from building knowledge about the LLM and understanding its capabilities and limitations based on previous interactions.

Less frequent reactions included positive emotions like calmness, thoughtfulness, playfulness, and curiosity, as well as negative emotions such as confusion and fear. This shows the varied and sometimes unexpected emotional responses that emerged. These emotions suggest that interacting with LLMs is not just a functional exchange but an exploratory experience for some software engineers. Curiosity, for example, has attached an investigative mindset often aligned with a trial-and-error approach. Playfulness, meanwhile, shows a willingness to engage with the LLM on a more open-ended basis. Fear introduces a new angle, suggesting that some engineers may feel a sense of responsibility if the interaction does not go as expected.

4.4 Expectations When Interacting with LLMs

As shown in Table 3, participants’ expectations for LLMs extend beyond functionality to quality, usability, and versatility, which are important factors in developing effective and trustworthy products in software development and design. Regarding *quality* and *performance*, engineers expect LLMs to consistently deliver correct and unhallucinated information without error or delay, as inconsistency can erode trust in the development workflow.

Participant: 19 “I rely on the LLM to provide accurate, relevant information that I can trust for both coding tasks and daily life management. It’s like having an expert who gets it right the first time!”

In terms of *understanding*, engineers value that LLMs understand context and intent, preferring models that ask clarification questions when necessary. They expect the LLM to understand the context without the need for a detailed context description in the initial prompt. This expectation is important since software research and development happen in dynamic and complex environments that require a lot of context, such as best practices, policies, and relevant software artefacts. With the lack of consideration for such contexts, the outcome can become unusable and hard to integrate into the solution.

Participant: 19 “I need an LLM that can seamlessly adapt to different contexts. Whether it’s helping me with technical jargon, understanding project management lingo, [...], the LLM should be versatile enough to handle it all.”

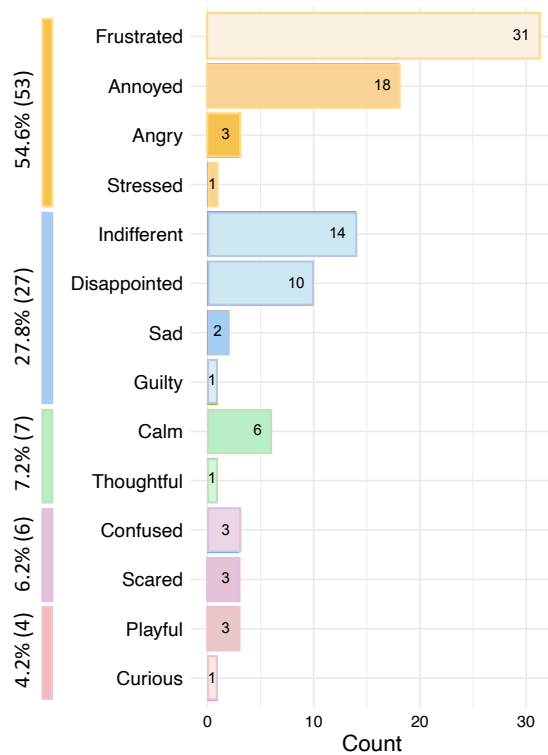


Figure 4: Emotional responses when receiving an incorrect answer. The colours map to Willcox’s emotions wheel in Figure 1.

Furthermore, since software engineering is a broad field with researchers from various sub-domains and specialities, researchers need to organise their texts and use language that fits the targeted audience. Therefore, software researchers both in academic and industrial organisations emphasised the importance of clarity and organisation of LLM-generated text, and that tailoring the answer to the *structure* that the task needs is crucial. For code-related tasks, engineers prefer responses starting with code snippets followed by explanations. When using LLMs to learn new concepts or explain artefacts, participants preferred elaborative answers.

Additionally, they expect LLMs to be transparent by providing the source of the information and confidence estimation of the output accuracy to ensure *information integrity* and support decision-making throughout the development lifecycle. This should also come with a need to protect the shared information in the chat.

Finally, in terms of *versatility*, engineers in industry increasingly expect LLMs to integrate with other tools and adapt to diverse workflows, reflecting the growing need for flexibility in software engineering environments. This also requires a *usable* LLM, ideally with a user-friendly interface that enables intuitive interactions and seamless integration of the LLM in software-related tasks.

Table 3: Users’ Expectations when Using LLMs

Themes	Categories
Quality	Accurate and Correct (39)
	Reliability and Consistency (13)
	No Hallucination (4)
Answer Structure	Conciseness vs. thoroughness (14)
	Complete with Examples (7)
	Organised and Good Grammar (6)
Performance	Efficiency (response time) (20)
Information Integrity	Transparency (10)
	Up-to-Date Information (4)
	Data Security and Confidentiality (5)
Understanding	Intent Understanding (10)
	Domain/Contextual Understanding (2)
Usability	Ease of Use (5)
Versatility	Integration with Other Tools (1)
	Adaptability to Workflow (1)
	Adaptability to Communication (2)

4.5 Frustration Triggers in Software Engineering

After exploring emotions in general related to receiving unexpected answers from LLMs, we focus on frustration-related emotions. We asked participants to describe specific situations where they felt frustrated when interacting with LLMs. From this, we identified patterns of triggers that cause LLMs to fail to meet engineers’ expectations, leading to strains such as frustration. We outline the frustration triggers below.

Repeated inaccuracies and hallucinations: One of the most common causes of frustration was receiving repeated incorrect or hallucinated responses from the LLM.

The definition of incorrectness varied among participants. Some examples of incorrect answers were uncompileable or buggy code, incorrect explanations of error messages, or incorrect factual information that was verified using other sources (e.g., an expert or a search engine). Hallucination was described as nonsense explanations, references that do not exist, made-up packages, and invalid syntax in a programming language.

Participants explicitly described that frustration arose when these issues persisted despite attempts to rephrase prompts or correct the LLM. For instance, participants referred to such situations as “annoying” or “disappointing” initially. However, they noted that *repeated failures* led to frustration, describing the LLM as “stubborn” and “insisting on an incorrect or hallucinated answer”.

Participant: 3 “After several corrections, and repeating the prompt in different manners, it decided to reiterate the same wrong response.”

This pattern of repeated failures can disrupt workflows in software engineering, where development cycles are often fast-paced

and agile, which requires more reliable and stable tools. For example, when the LLM provides a code that imports hallucinated libraries, it renders the code unusable, which leads to wasted time fixing, debugging, or rewriting the entire implementation.

Participant: 38 “During a coding problem, I was looking for the usage of a specific function in a library. I was frustrated when it provided a different [function] (which did not work or even exist).”

Intent not understood: Frustrations (or related emotions) are also caused when the software engineers feel that the LLM did not understand their prompt. Not understanding can be reflected in a response that is irrelevant to the initial question. Intent understanding was a common frustration trigger among our participants from industry and academia, since in practice, engineers’ queries are often highly technical and domain-specific and deal with complex software artefacts. Similarly, researchers and academics deal with novel techniques and niche problems that may cause the LLM to misunderstand the intent. Note that such misunderstandings are more common in general-purpose LLMs than in fine-tuned and specialised LLMs.

Participant: 29 “I was trying to ask [LLM] to fill one specific cell in a notebook based on the others but it kept returning the same generic code for two cells instead of one. I had to talk to [LLM] like a child and say don’t do that and do this, and only this.”

On another note, some participants pointed out their perceived usefulness of prompt programming and carefully constructing a prompt that would minimise such misunderstandings that can be caused by poor phrasing or the lack of context in the prompt.

Participant: 5 “I provide short and maybe unclear prompts [then] I usually get irrelevant responses. The better the prompt, the better response.”

While prompt programming has shown a high potential in enhancing the LLM outcome, it remains unclear whether it is effective in software engineering-related tasks.

Personal preferences unmet: Many of our participants pointed out that they get frustrated with reasons related to their personal preferences. Some aspects of certain LLMs (e.g. answer structure) can be annoying to some engineers, and when the frequency of interactions with the LLM increases, the annoyance turns into frustration. For example, two participants pointed out that an LLM apologising every time they tried to correct the LLM was a source of frustration since it can disrupt the workflow, especially during a refactoring or debugging process with the LLM, which results in a long conversation with many follow-up prompts. Others were frustrated with how the LLM they use only provides answers that are long or only in bullet points. Forcing the LLM to structure an answer that aligns with their preferences required specifying many requirements and constraints in the prompt.

Participant: 35 “[I get frustrated] when [LLM] gives too long answers. I quite often ask things that can be answered with a short sentence, but still I get half a page of ramblings back.”

Such preferences depend on the task that the software engineer is solving. For example, important details when debugging might be hidden in long responses, while overly concise bullet points might omit crucial information needed to understand a system’s architecture.

Limitations of the LLM: LLM limitations (e.g., inability to perform specific tasks) or configuration constraints (e.g., context window size) are frustration triggers for software engineers. When an engineer attempts to force the LLM to overcome these limitations by prompting, it often leads to more frustration. For example, trying to generate a large application code in one prompt can result in missing lines and errors when the LLM hits its maximum token limit [25]. The participants highlighted that there are some of the many tasks in software engineering research and practice that LLMs are just “not good at”.

Participant: 56 “[I got frustrated when] formatting of a table in latex, [had to] move to another LLM”

Finally, as a verification question, we asked the participants to rate aspects of LLMs (correctness, lack of hallucinations, understanding, performance, and ability to answer) on a 5-point Likert scale based on how important they are in an LLM in order to ensure a better user experience. The results in Figure 5 resonate with the previous results, where correctness, lack of hallucination, and correctness are the most important. In comparison, performance (e.g., response time) and ability to answer were seen as less critical.

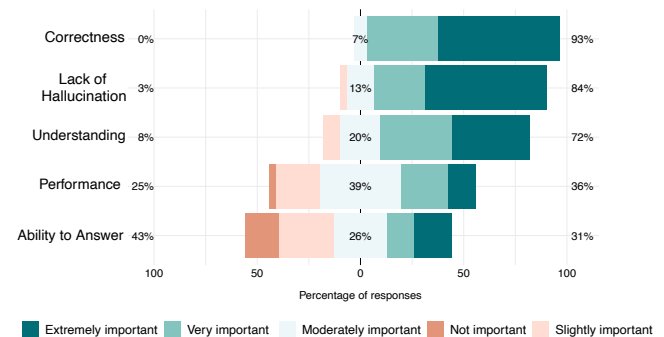


Figure 5: Likert-scale results of the importance level of different aspects of the LLMs that can impact the user experience. The scale ranges from Not Important (left - red) to Extremely Important (right - green).

4.6 Unmet Expectations’ Impact on Motivation

When the LLM failed to assist our participants, the participants’ motivation to complete their task was influenced mainly in three ways.

21.3% (13 out of 62) of our respondents reported that their **motivation decreased** when LLMs did not give the correct answer. Responses expressed frustration, stress, or disappointment, impacting participants’ willingness to continue. Other participants commented that they eventually gave up. These respondents mentioned that after some effort, they decided to abandon the LLM and move on to other methods or stop altogether.

Participant: 29 “Eventually I give up [on the task], or report negative experimental results.”

Another group was formed of 77% (48 out of 62) of the participants who were unfazed by the LLM’s failure, treating it as a non-critical tool and continuing with the task with their **motivation not being affected**.

Participant 8: “My motivation is not affected, I just realise that the task will take longer.”

In an interesting case, one participant mentioned that their **motivation increased** which was due to perceiving the interaction as a learning opportunity.

Participant 16: “I usually understand the problem a lot more so I want to complete the task more”

4.7 User Actions for Improving LLM Interactions

When asked what actions participants typically take after receiving an unexpected answer from an LLM, the majority (41 out of 62 participants) said they changed the prompt to try again. A smaller group (12 participants) reported providing feedback to the LLM, while a few (2 participants) said they did nothing (see Figure 6). In the “Other” category, participants described various strategies. Some combine multiple actions, such as changing the prompt, switching to another LLM, or even reverting to traditional search methods like Google or Wikipedia. A few participants mentioned disengaging from the LLM entirely or constructing the solution themselves.

Participant: 39 “Sometimes the LLM hallucinates and puts me in loops. When I realise this, I resolve it myself using my human knowledge and [X] years of experience in the development field”

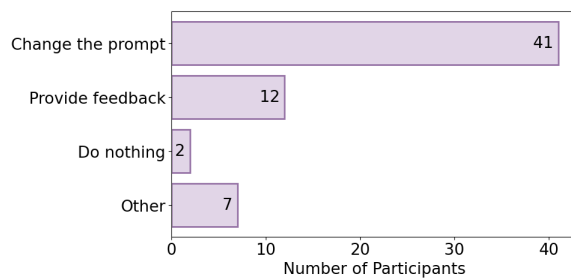


Figure 6: Actions done after receiving an unexpected answer from LLMs

We asked our participants about the improvements they would like to see that would enhance their experience when using LLMs. They identified several areas that we grouped and sorted in Table 4 based on the number of times they were mentioned. The suggestions mainly align with the engineers’ expectations (See Table 3), though with a variation in the emphasis and distribution.

Regarding the *quality* of the outcome, the participants emphasised the importance of reducing hallucinations and improving response accuracy. They also commented on the need for enhanced analytical capacity to tackle more complex problems. However, the participants highlighted that while they expect an LLM to be accurate, transparency is more crucial for a less frustrating user experience. The transparency was described by the participants concerned **information integrity**, particularly specifying the data sources and the confidence level of the LLM’s responses. This also includes stating any assumptions and reasoning by the LLM before providing an answer. This helps practitioners decide whether to rely on the LLM’s suggestions, trust their own judgment, or seek assistance from a colleague instead. Establishing trust in this context is crucial, as it determines the main flow of the interaction, and hence how the LLM output will be eventually used by the software engineer [26].

Regarding the *answer structure*, participants preferred concise and short answers (unless they were prompted otherwise). While in terms of *understanding*, prompt comprehension was considered essential, with users expecting LLMs to ask clarifying questions when necessary. However, some participants mentioned that this could be a *human-related* improvement where training that is designed specifically to learn to “talk” with the LLM is important to make communication more effective.

Finally, *versatility* was another improvement that was consistently mentioned among our participants; they indicated the importance of handling more complex tasks and adapting to different user needs. The increased functionality to support a broader range of tasks was also crucial for improving overall user satisfaction. This is especially critical as software engineers, researchers, and practitioners often work with a variety of tasks that range from learning new concepts and tasks to implementing products and using tools. Furthermore, practitioners particularly stressed the importance of LLM *performance* in terms of increased memory and learning capabilities, allowing LLMs to retain relevant information and build on past interactions with a faster processing time.

Overall, these requirements guide chatbot designers to understand practitioners’ and researchers’ needs and priorities in software engineering. For instance, the next-generation LLMs may need to direct efforts toward transparency rather than performance when aiming for a better user experience.

5 Discussion

In this section, we answer our research questions, highlighting the main takeaway per question.

5.1 Frustrations in the Context of LLM Interaction

We identified four main frustration triggers for software engineers using LLMs, accuracy issues, hallucinations, misunderstandings, unmet preferences, and LLM limitations, pointing to significant challenges with potential long-term repercussions. For example, repeatedly dealing with inaccurate, buggy or non-standard code disrupts workflow, forcing engineers to spend extra time debugging [14] or reworking tasks that an efficient tool should minimise. This is particularly important if the code produced by the LLM is less

Table 4: Improvements for Better User Experience

Themes	Categories
Information Integrity	Transparency (source, confidence) (16) Trust and Data Security (4) Relevant information (Up-to-date) (2) State reasoning and assumptions (2)
Versatility	Adaptability in communication (11) Integration in project environment (2) Extended Functionality (2)
Quality	Improved Response Accuracy (6) Reduced Hallucinations (4) Less creative (3)
Understanding	Context understanding (5) Clarification questions (3) Intent Understanding (2)
Answer Structure	Elaborative answers on-demand (4) Consistency in Responses (3)
Performance	Higher memory utilisation (4) Efficiency (processing time) (1)
Human-related	Training for engineers (2)

maintainable [31] and can be more frustrating than writing the code with no LLM assistance [59]. This extra workload impacts project timelines and creates a deeper frustration that could enforce frustration and emotional strain [7, 29] as engineers may feel that they have spent longer than planned to reach their goal (e.g., completing a task). Additionally, participants' frustration resulted from other emotions, including anger and annoyance (see Figure 4).

On Wilcox's emotion wheel, these frustrations can be linked to various emotions within the frustration spectrum, including anger, annoyance, and confusion. For instance, frustration over inaccurate or faulty code can easily evolve into anger when engineers feel a lack of control over the situation, especially if the tool is supposed to improve efficiency. Similarly, when an LLM produces outputs that deviate from expectations, it can lead to another emotion within the spectrum, for example, annoyance, particularly when the tool fails to meet personal preferences or the engineer's standards. In addition, misunderstandings or incorrect outputs might lead to confusion as engineers try to reconcile the LLM's output with their original intent.

Studies on GitHub Copilot [47, 47] showed similar emotions, especially around data privacy concerns, intrusive code suggestions and usability, particularly unnecessary large code suggestions. Eshraghian et al. [12] explained that frustration and anger can come from feeling a threat without being able to control it. In the previous example, the threat of leaking confidential data with very little control over it (i.e., to use the LLM, they need to accept the policy) was the trigger for frustration.

Unlike other domains where frustration often stems from performance issues (e.g., system crashes) [29], our participants did

not report such frustrations, likely due to recent LLMs being stable and fast (Table 3). Other frustrations in the medical domain arise from the emotionally exhausting work environment along with their dependence on the technology (e.g., to document patient data) [45, 56]. In contrast, software engineers can still rely on their expertise or alternative tools, as LLMs are not essential for task completion (Section 4.6).

Takeaway: Frustration triggers studied in software engineer literature (including ours) come from spending extra time refining output, unlike in existing studies in other domains, where it is due to performance and usability issues.

5.2 Impact of Frustrating Experiences on Motivation

While most of our participants expressed frustration (or similar emotions), their motivation to complete the task was not necessarily affected. This can suggest that although frustration may reside in the emotion wheel's 'anger' or 'irritation' sections, the engineers' resilience and coping mechanisms allowed them to manage these emotions without diminishing motivation.

Our results showed that participants often felt demotivated when an LLM failed to meet their expectations, such as when it did not assist them as intended, frequently leading to frustration. However, most of those who felt frustrated reported that their motivation remained intact, likely due to perceiving the LLM as merely one tool among others in achieving their goal. When an LLM could not provide the necessary assistance, participants commonly resorted to alternative solutions, such as using search engines like Google or relying on their expertise to complete the task independently. These observations align with findings by Franca et al. [17], who explored the connection between motivation and the satisfaction and happiness of software engineers, revealing that happiness slightly overlaps with but does not correlate with motivation.

These findings suggest that, despite facing emotional challenges, software engineers maintain their motivation to persist with demanding tasks. Although frustration may not directly impact motivation, it remains a critical factor to consider due to its known connection with burnout. Sustained frustration is still significant as it contributes to emotional strain [56], a known precursor to burnout in high-demand professions like SE. Thus, understanding and managing frustration, even when the motivation appears unaffected, is essential in supporting the long-term well-being of software engineers.

Takeaway: Although frustration occurs when LLMs fail to meet expectations, it generally does not diminish the motivation to complete tasks.

5.3 Towards a Less-Frustrating User Experience

When designing chatbots and LLMs, it is essential to prioritise not only high accuracy and performance but also emotional intelligence, such as recognising user frustration. Wilcox's emotion wheel provides a nuanced view of emotional states, categorising emotions into primary and secondary feelings. Recognising and responding to these emotions in real-time is key.

The emotional intelligence of LLMs has been explored by Wang et al. [58] where they saw that newer-generation LLMs such as GPT-4 (at the time of the study) show a better ability to understand user emotions that compared to humans' emotional intelligence. However, recognising emotions is insufficient as the LLM should also know how to act accordingly. In Section 4.5, we saw that even minor LLM behaviours such as apologising after receiving feedback were making the engineers' experience more frustrating. This was also described by Erlenhov et al [10] about ideal development bots adapting communication to different individuals. While several studies focus on enhancing LLMs, we recognise that these systems will always have room for improvement. Therefore, our focus in this study shifts to the human element. We propose enabling users with the knowledge and skills to navigate LLMs effectively to reduce frustration and improve overall user experience. The goal is to address frustration triggers that arise during their use. Specifically, triggers such as misunderstanding of the intent, unsatisfying personal preferences, or even getting incorrect answers can be minimised by prompt engineering the query before sending it to the LLM. Prompt engineering can involve incorporating prompt techniques (such as Few-shot learning), relevant contextual information (e.g., system description), or constraints about the output (e.g., the output structure). Other frustration triggers, such as hallucinations and limitations of LLMs, can be minimised if the engineers use the LLM according to its capabilities and limitations. Since unmet expectations are among the primary frustration triggers[30] (see Section 4.4), providing software engineers with training on effective usage and a clear understanding of LLM capabilities to set realistic expectations can help reduce disappointment and enhance user satisfaction.

Additionally, raising awareness about potential frustration triggers is important; engineers can manage their reactions accordingly if they recognise the likelihood of frustration in certain situations. For instance, using coping strategies rather than repeatedly attempting to elicit a perfect answer from the LLM. Therefore, we suggest that software engineers need training on how to use LLMs safely. This idea was also discussed by Barman et al. [5] where they propose providing guidelines for LLM users to know how to interact with different LLMs, for instance, if it is appropriate to generate artefacts or only to get some guidance. Our participants commented that they were familiar with LLMs; however, most reported frustration, raising questions about whether they truly understand how to leverage LLMs effectively. Familiarity does not necessarily equate to proficiency[28], stressing the need for improved training and guidance on optimal usage strategies. A complete understanding of LLM capabilities and limitations can help users to manage their expectations.

Takeaway: A less frustrating experience arises from combining “emotionally intelligent” LLMs with engineers' awareness to manage their expectations and reactions to frustration triggers.

5.4 Threats to Validity

In this section, we explain the strategies to address this study's threats to validity.

Internal Validity: To ensure internal validity, we considered several biases and employed mitigations accordingly. Self-selection bias: Participation in the survey was voluntary; hence, individuals with particularly strong positive or negative experiences with LLMs might be overrepresented, skewing the data. To mitigate this, we tried to recruit diverse participants across different experience levels, regions, and fields. To address social desirability bias, we collected anonymous data by including a statement at the beginning of the survey and avoiding personal questions. This approach aimed to prevent participants from feeling pressured to align their responses with what they perceived as socially or professionally acceptable. This could lead to underreporting frustration to appear more competent with new technologies.

External Validity: Our sample size of software practitioners and academics can limit to extent to which our findings can be generalized to the broader population of software engineers, to minimize this threat while avoiding overrepresenting certain groups or regions, we targeted respondents from different countries and several fields within SE. Similarly, we employed stratified sampling to ensure a balanced representation across demographics, skill levels, and industries.

Construct Validity: We operationalised key concepts like frustration and hallucination to guarantee construct validity, adding their definitions. Additionally, we provided examples throughout the survey to clarify the scenarios we were exploring. Finally, we explained the Likert scale by adding information on how to measure each level. Further, we piloted the survey to ensure clarity. We used the feedback to fix ambiguous questions, clarify terms, and minimise confusion.

6 Conclusion

This study focused on the emotional strains, particularly frustration, experienced by software engineers when interacting with LLMs and not being assisted as intended. By identifying main triggers, such as the correctness and reliability of responses and issues related to personalisation, we emphasise that understanding the emotional impacts of LLM use in SE is important. This study's insights bring attention to the potential risks to productivity and mental health if emotional responses go unaddressed. Future research should further explore the psychological implications of LLM use, focusing on sustainable strategies to support the well-being of software engineers and optimise their user experience with AI tools.

References

- [1] Owura Asare, Meiyappan Nagappan, and N Asokan. 2023. Is github's copilot as bad as humans at introducing vulnerabilities in code? *Empirical Software Engineering* 28, 6 (2023), 129.
- [2] American Psychological Association. n.d.. Emotion. (n.d.). <https://dictionary.apa.org/emotion> Accessed: 2024-10-05.
- [3] American Psychological Association. n.d.. Frustration. (n.d.). <https://dictionary.apa.org/frustration> Accessed: 2025-04-12.
- [4] Sebastian Baltes and Paul Ralph. 2022. Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering* 27, 4 (2022), 94.
- [5] Kristian González Barman, Nathan Wood, and Pawel Pawlowski. 2024. Beyond transparency and explainability: on the need for adequate and contextualized user guidelines for LLM use. *Ethics and Information Technology* 26, 3 (2024), 47.
- [6] Rafael A Calvo and Dorian Peters. 2014. *Positive computing: technology for wellbeing and human potential*. MIT press.
- [7] Irina Ceaparu, Jonathan Lazar, Katie Bessiere, John Robinson, and Ben Shneiderman. 2004. Determining causes and severity of end-user frustration. *International journal of human-computer interaction* 17, 3 (2004), 333–356.

- [8] Chu-Hsiang Chang, Russell E Johnson, and Liu-Qin Yang. 2007. Emotional strain and organizational citizenship behaviours: A meta-analysis and review. *Work & Stress* 21, 4 (2007), 312–332.
- [9] Arghavan Moradi Dakhel, Wahid Majdinasab, Amin Nikanjam, Foutse Khomh, Michel C Desmarais, and Zhen Ming Jack Jiang. 2023. Github copilot ai pair programmer: Asset or liability? *Journal of Systems and Software* 203 (2023), 111734.
- [10] Linda Erlenhov, Francisco Gomes de Oliveira Neto, Riccardo Scandariato, and Philipp Leitner. 2019. Current and future bots in software development. In *2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering (BotSE)*. IEEE, 7–11.
- [11] Christen Erlingsson and Petra Brysiewicz. 2017. A hands-on guide to doing content analysis. *African journal of emergency medicine* 7, 3 (2017), 93–99.
- [12] Farjam Eshraghian, Najmeh Hafezieh, Farveh Farivar, and Sergio de Cesare. 2024. AI in software programming: understanding emotional responses to GitHub Copilot. *Information Technology & People* (2024).
- [13] Nina R Ferreri and Christopher B Mayhorn. 2023. Identifying and understanding individual differences in frustration with technology. *Theoretical Issues in Ergonomics Science* 24, 4 (2023), 461–479.
- [14] Sue Fitzgerald, Gary Lewandowski, Renee McCauley, Laurie Murphy, Beth Simon, Lynda Thomas, and Carol Zander. 2008. Debugging: finding, fixing and flailing, a multi-institutional study of novice debuggers. *Computer Science Education* 18, 2 (2008), 93–116.
- [15] Dena Ford and Chris Pamin. 2015. Exploring causes of frustration for software developers. In *2015 IEEE/ACM 8th international workshop on cooperative and human aspects of software engineering*. IEEE, 115–116.
- [16] Clare Foster and Janet Sayers. 2012. Exploring physiotherapists' emotion work in private practice. *New Zealand Journal of Physiotherapy* 40, 1 (2012), 17–23.
- [17] César França, Helen Sharp, and Fabio QB Da Silva. 2014. Motivated software engineers are engaged and focused, while satisfied ones are happy. (2014), 8.
- [18] Daniela Girardi, Filippo Lanubile, Nicole Novielli, and Alexander Serebrenik. 2022. Emotions and Perceived Productivity of Software Developers at the Workplace. *IEEE Transactions on Software Engineering* 48, 9 (2022), 3326–3341. doi:10.1109/TSE.2021.3087906
- [19] Daniela Girardi, Nicole Novielli, Davide Fucci, and Filippo Lanubile. 2020. Recognizing developers' emotions while programming. In *Proceedings of the ACM/IEEE 42nd international conference on software engineering*. 666–677.
- [20] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. 2017. On the unhappiness of software developers. In *Proceedings of the 21st international conference on evaluation and assessment in software engineering*. 324–333.
- [21] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. 2018. What happens when software developers are (un) happy. *Journal of Systems and Software* 140 (2018), 32–47.
- [22] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2013. Are happy developers more productive? The correlation of affective states of software developers and their self-assessed productivity. In *Product-Focused Software Process Improvement: 14th International Conference, PROFES 2013, Paphos, Cyprus, June 12-14, 2013. Proceedings 14*. Springer, 50–64.
- [23] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2014. Software developers, moods, emotions, and performance. *arXiv preprint arXiv:1405.4422* (2014).
- [24] Muhammad Usman Hadi, Rizwan Qureshi, Abbas Shah, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, Seydali Mirjalili, et al. 2023. A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints* 3 (2023).
- [25] HuggingFace. 2023. Transformers Documentation: LLM Tutorial. https://huggingface.co/docs/transformers/en/llm_tutorial. (2023). Accessed: 2024-11-04.
- [26] Ranim Khojah, Francisco Gomes de Oliveira Neto, and Philipp Leitner. 2024. From Human-to-Human to Human-to-Bot Conversations in Software Engineering. In *Proceedings of the 1st ACM International Conference on AI-Powered Software (Porto de Galinhas, Brazil) (AIWare 2024)*. Association for Computing Machinery, New York, NY, USA, 38–44. doi:10.1145/3664646.3664761
- [27] Ranim Khojah, Mazen Mohamad, Philipp Leitner, and Francisco Gomes de Oliveira Neto. 2024. Beyond code generation: An observational study of chatgpt usage in software engineering practice. *Proceedings of the ACM on Software Engineering* 1, FSE (2024), 1819–1840.
- [28] Pete Ladwig, Kajsa E Dalrymple, Dominique Brossard, Dietram A Scheufele, and Elizabeth A Corley. 2012. Perceived familiarity or factual knowledge? Comparing operationalizations of scientific understanding. *Science and Public Policy* 39, 6 (2012), 761–774.
- [29] Jonathan Lazar, Adam Jones, and Ben Shneiderman. 2006. Workplace user frustration with computers: An exploratory investigation of the causes and severity. *Behaviour & Information Technology* 25, 03 (2006), 239–251.
- [30] Shelby L Levine, Claire J Brabander, Amanda M Moore, Anne C Holding, and Richard Koestner. 2022. Unhappy or unsatisfied: distinguishing the role of negative affect and need frustration in depressive symptoms over the academic year and during the COVID-19 pandemic. *Motivation and Emotion* 46, 1 (2022), 126–136.
- [31] Yue Liu, Thanh Le-Cong, Ratnadira Widyasari, Chakkrit Tantithamthavorn, Li Li, Xuan-Bach D Le, and David Lo. 2024. Refining chatgpt-generated code: Characterizing and mitigating code quality issues. *ACM Transactions on Software Engineering and Methodology* 33, 5 (2024), 26.
- [32] Sebastian Lubos, Alexander Felfernig, Thi Ngoc Trang Tran, Damian Garber, Merfat El Mansi, Seda Polat Erdeniz, and Viet-Man Le. 2024. Leveraging llms for the quality assurance of software requirements. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*. IEEE, 389–397.
- [33] Mika Mäntylä, Bram Adams, Giuseppe Destefanis, Daniel Graziotin, and Marco Ortu. 2016. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?. In *Proceedings of the 13th international conference on mining software repositories*. 247–258.
- [34] Mika V Mäntylä, Kai Petersen, Timo OA Lehtinen, and Casper Lassenius. 2014. Time pressure: a controlled experiment of test case development and requirements review. In *Proceedings of the 36th International Conference on Software Engineering*. 83–94.
- [35] Esperanza Marcos, Roberto Hens, Teresa Puebla, and Juan Manuel Vara. 2020. Applying emotional team coaching to software development. *IEEE Software* 38, 4 (2020), 85–93.
- [36] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2015. Focused, aroused, but so distractible: Temporal perspectives on multitasking and communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 903–916.
- [37] Cristina Martinez Montes and Ranim Khojah. 2025. Replication Package for The Study "Emotional Strain and Frustration in LLM Interactions in Software Engineering". (Jan. 2025). doi:10.5281/zenodo.14766606
- [38] Jesse G Meyer, Ryan J Urbanowicz, Patrick CN Martin, Karen O'Connor, Ruowang Li, Pei-Chen Peng, Tiffani J Bright, Nicholas Tatonetti, Kyoung Jae Won, Graciela Gonzalez-Hernandez, et al. 2023. ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining* 16, 1 (2023), 20.
- [39] Cristina Martinez Montes, Fredrik Sjögren, Adam Klevfors, and Birgit Penzenstadler. 2024. Qualifying and Quantifying the Benefits of Mindfulness Practices for IT Workers. In *2024 10th International Conference on ICT for Sustainability (ICT4S)*. IEEE, 272–281.
- [40] Rebecca Müller, David Schischke, Benedikt Graf, and Conny H Antoni. 2023. How can we avoid information overload and techno-frustration as a virtual team? The effect of shared mental models of information and communication technology on information overload and techno-frustration. *Computers in Human Behavior* 138 (2023), 107438.
- [41] Sebastian C Müller and Thomas Fritz. 2015. Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 1. IEEE, 688–699.
- [42] Megan Anna Neff. 2023. Emotional Identification. (2023). <https://neurodivergentsights.com/blog/emotional-identification> Accessed: 2024-10-07.
- [43] Megan Anna Neff. 2024. The Feelings Wheel. (2024). <https://neurodivergentsights.com/blog/the-feelings-wheel> Accessed: 2024-10-27.
- [44] Johannes J Norheim, Eric Rebentisch, Dekai Xiao, Lorenz Draeger, Alain Kerbrat, and Olivier L de Weck. 2024. Challenges in applying large language models to requirements engineering tasks. *Design Science* 10 (2024), e16.
- [45] Gloria A Opoku-Boateng. 2015. USER FRUSTRATION IN HIT INTERFACES: EXPLORING PAST HCI RESEARCH FOR A BETTER UNDERSTANDING OF CLINICIANS' EXPERIENCES. In *AMIA Annual Symposium Proceedings*, Vol. 2015. 1008.
- [46] Birgit Penzenstadler, Richard Torkar, and Cristina Martinez Montes. 2022. Take a deep breath: Benefits of neuroplasticity practices for software developers and computer workers in a family of experiments. *Empirical Software Engineering* 27, 4 (2022), 98.
- [47] James Prather, Brent N. Reeves, Paul Denny, Brett A. Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett Powell, James Finnie-Ansley, and Eddie Antonio Santos. 2023. "It's Weird That it Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *ACM Trans. Comput.-Hum. Interact.* 31, 1, Article 4 (Nov. 2023), 31 pages. doi:10.1145/3617367
- [48] James A Russell. 1991. Culture and the categorization of emotions. *Psychological bulletin* 110, 3 (1991), 426.
- [49] Marisa Salanova, Susana Llorens, and Mercedes Ventura. 2014. Technostress: The dark side of technologies. In *The impact of ICT on quality of working life*. Springer, 87–103.
- [50] Mary Sánchez-Gordón and Ricardo Colomo-Palacios. 2019. Taking the emotional pulse of software engineering—A systematic literature review of empirical studies. *Information and Software Technology* 115 (2019), 23–43.
- [51] Shirish C Srivastava, Shalini Chandra, and Anuragini Shirish. 2015. Technostress creators and job outcomes: theorising the moderating influence of personality traits. *Information Systems Journal* 25, 4 (2015), 355–401.
- [52] Robert A Stebbins. 2001. *Exploratory research in the social sciences*. Vol. 48. Sage.
- [53] Klaas-Jan Stol and Brian Fitzgerald. 2020. Guidelines for conducting software engineering research. In *Contemporary Empirical Methods in Software Engineering*. Springer, 27–62.

- [54] Margaret-Anne Storey and Alexey Zagalsky. 2016. Disrupting developer productivity one bot at a time. In *Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering*. 928–931.
- [55] Monideepa Tarafdar, Qiang Tu, Bhanu S Ragu-Nathan, and TS Ragu-Nathan. 2007. The impact of technostress on role stress and productivity. *Journal of management information systems* 24, 1 (2007), 301–328.
- [56] Daniel S Tawfik, Amrita Sinha, Mohsen Bayati, Kathryn C Adair, Tait D Shanafelt, J Bryan Sexton, and Jochen Profit. 2021. Frustration with technology and its relation to emotional exhaustion among health care workers: cross-sectional observational study. *Journal of medical Internet research* 23, 7 (2021), e26817.
- [57] Junjie Wang, Yuchao Huang, Chunyang Chen, Zhe Liu, Song Wang, and Qing Wang. 2024. Software testing with large language models: Survey, landscape, and vision. *IEEE Transactions on Software Engineering* (2024).
- [58] Xuena Wang, Xueting Li, Zi Yin, Yue Wu, and Jia Liu. 2023. Emotional intelligence of Large Language Models. *Journal of Pacific Rim Psychology* 17 (2023), 18344909231213958. doi:10.1177/18344909231213958
- arXiv:https://doi.org/10.1177/18344909231213958
- [59] Justin D Weisz, Michael Muller, Steven I Ross, Fernando Martinez, Stephanie Houde, Mayank Agarwal, Kartik Talamadupula, and John T Richards. 2022. Better together? an evaluation of ai-supported code translation. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 369–391.
- [60] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. “As an AI language model, I cannot”: Investigating LLM Denials of User Requests. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [61] Gloria Willcox. 1982. The feeling wheel: A tool for expanding awareness of emotions and increasing spontaneity and intimacy. *Transactional Analysis Journal* 12, 4 (1982), 274–276.
- [62] Novia Wong, Victoria Jackson, André Van Der Hoek, Iftekhar Ahmed, Stephen M Schueller, and Madhu Reddy. 2023. Mental wellbeing at work: Perspectives of software engineers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.