



Impermanent identifiers: Enhanced source code comprehension and refactoring[☆]

Eduardo Martins Guerra^{a,*}, André A.S. Ivo^b, Fernando O. Pereira^b, Romain Robbes^a, Andrea Janes^d, Fábio Fagundes Silveira^c

^a Free University of Bozen-Bolzano, Bozen-Bolzano, Italy

^b National Institute for Space Research (INPE), São José dos Campos, São Paulo, Brazil

^c Federal University of São Paulo, São José dos Campos, São Paulo, Brazil

^d Free University of Bozen-Bolzano, Bolzano, Italy

ARTICLE INFO

Keywords:

Source code identifiers
Program comprehension
Software refactoring
Software evolution

ABSTRACT

In response to the prevailing challenges in contemporary software development, this article introduces an innovative approach to code augmentation centered around *Impermanent Identifiers*. The primary goal is to enhance the software development experience by introducing dynamic identifiers that adapt to changing contexts, facilitating more efficient interactions between developers and source code, ultimately advancing comprehension, maintenance, and collaboration in software development. Additionally, this study rigorously evaluates the adoption and acceptance of *Impermanent Identifiers* within the software development landscape. Through a comprehensive empirical examination, we investigate how developers perceive and integrate this approach into their daily programming practices, exploring perceived benefits, potential barriers, and factors influencing its adoption. In summary, this article charts a new course for code augmentation, proposing *Impermanent Identifiers* as its cornerstone while assessing their feasibility and acceptance among developers. This interdisciplinary research seeks to contribute to the continuous improvement of software development practices and the progress of code augmentation technology.

1. Introduction

In the contemporary digital landscape, software development plays a central role in driving innovation and functionality within technological systems. However, this process is not devoid of challenges, especially due to the increasing complexity of modern software systems. As software engineering evolves, the constant need to enhance developer efficiency and understanding becomes increasingly crucial. Critical information for understanding, such as control structures and dependencies, often gets lost in the intricate web of lines of code (McConnell, 2004; Martin and Coplien, 2009). This growing complexity underscores the importance of tools and approaches to alleviate developers' cognitive load. In response to this demand, the concept of "Code Augmentation" emerges as a promising approach to improve the interaction between developers and source code (Beelders and du Plessis, 2016; Sulír et al., 2018).

The "Code Augmentation" approach seeks to enrich the development experience by incorporating visual information directly into the

source code. By adding visual elements such as colors, icons, and labels to source code, developers can gain faster and more accurate insights into the structure, flow, and relationships between pieces of code (Sulír et al., 2018; Beelders and du Plessis, 2016). Code augmentation, therefore, aims to create an environment where understanding the code is facilitated, allowing for a more intuitive and efficient interaction with software development.

The history of visual source code augmentation goes back to the early days of programming, when the first code editors introduced syntax highlighting features to improve readability and identify elements in code, called *identifiers* (Sulír et al., 2018). The 1980s marked a milestone with the LEXX (Cowlshaw, 1987) editor, which introduced real-time syntax highlighting. The evolution continues with advances, such as immediate feedback on syntax errors, as exemplified by the Magpie system (Schwartz et al., 1984). In the 1990s, projects such as ZStep95 (Ungar et al., 1997) demonstrated interactivity by highlighting

[☆] Editor: Dr Gabriele Bavota.

* Corresponding author.

E-mail addresses: eduardo.guerra@unibz.it (E.M. Guerra), andre.ivo@inpe.gov.br (A.A.S. Ivo), fernando.opc@gmail.com (F.O. Pereira), romain.robbes@gmail.com (R. Robbes), andrea.janes@unibz.it (A. Janes), fsilveira@unifesp.br (F.F. Silveira).

URL: <https://www.unibz.it/en/faculties/engineering/academic-staff/person/43879-eduardo-martins-guerra> (E.M. Guerra).

<https://doi.org/10.1016/j.jss.2024.112137>

Received 2 November 2023; Received in revised form 30 May 2024; Accepted 12 June 2024

Available online 17 June 2024

0164-1212/© 2024 Elsevier Inc. All rights reserved, including those for text and data mining, AI training, and similar technologies.

expressions under evaluation and displaying graphical output close to the source code. During the past two decades, with increasing computational power and advances in analysis algorithms, conventional IDEs have incorporated various augmenting features, encouraging research in this field (Sulír et al., 2018; Murphy, 2019).

Identifiers in the source code are ubiquitous: they make up around 70% of a software system's source code size (measured in characters) and approximately a third of source code tokens (Deissenboeck and Pizka, 2006). Multiple studies have shown that descriptive variables, methods, or class names positively impact program comprehension: descriptive identifiers support program comprehension (Lawrie et al., 2006; Schankin et al., 2018), shorter identifiers hamper comprehension (Hofmeister et al., 2017) (although some specific ones might not Beniamini et al., 2017). Moreover, the identifier style and coding conventions can also have an impact on program comprehension (Binkley et al., 2009; Sharafi et al., 2012).

Refactorings (Fowler, 2018) are used to improve source code maintainability, where a large part of them are renamings (Hora et al., 2018). Many approaches have been proposed to improve identifiers, providing suggestions for renaming. These approaches range from normalizing source code identifiers (Guerrouj, 2013) to detecting naming bugs (Høst and Østvold, 2009), bad smells (Abebe et al., 2009), or anti-patterns (Arnaoudova et al., 2016). Other approaches leverage advanced algorithms or machine learning techniques to expand abbreviations (Lawrie et al., 2007; Corazza et al., 2012), provide renaming suggestions (Allamanis et al., 2015), spot inconsistencies (Liu et al., 2019), find bugs related to identifier names (Pradel and Sen, 2018), or even de-obfuscate source code (Vasilescu et al., 2017).

Beyond program comprehension, software identifiers, due to their predominance in software systems, are extensively used in Natural Language Processing (NLP) (Arnaoudova et al., 2015), and Machine Learning for Software Engineering (Allamanis et al., 2018) approaches to support a variety of software engineering tasks, including several of the renaming suggestions mentioned above (thus closing the loop), among many other approaches.

In the face of this scenario, this article aims to present a novel approach to code augmentation described as Impermanent Identifiers, aiming not only to enhance the software development experience but also to comprehensively analyze the adoption and acceptance of this technology in the development process.

The objectives of this study encompass two interrelated and equally significant aspects. First, we strive to make a substantial contribution to the field of code augmentation by proposing an approach centered around Impermanent Identifiers. This approach aims to explore the creation of dynamic identifiers that adapt to evolving contextual needs, enabling a more seamless and accurate interaction between developers and source code. In this regard, our research seeks to expand the concept of code augmentation, moving beyond conventional approaches and establishing the groundwork for an innovative methodology that enhances understanding, maintenance, and collaboration in the software development process.

Furthermore, our second objective is to conduct a rigorous evaluation of the adoption and acceptance of this technology in the software development landscape. Recognizing that the introduction of new technological approaches often encounters acceptance and integration challenges, a comprehensive empirical study was carried out. This involved assessing how developers perceive and adopt Impermanent Identifiers as part of their daily programming practices. We aimed to understand the perceived benefits, potential barriers, and factors influencing the decision to incorporate this innovative approach into their workflows.

In summary, this article seeks to set a new direction in the field of code augmentation by proposing the use of Impermanent Identifiers as a foundation, while simultaneously exploring the feasibility and acceptance of this approach within the developer community. Through this comprehensive and interdisciplinary research, we hope to contribute to

the continuous improvement of software development practices and to the advancement of code augmentation as a whole.

Section 2 provides the theoretical background used in this study, Section 3 introduces the concept of Impermanent Identifiers, Section 4 outlines the methodology employed for the assessment of adoption and acceptance, Section 5 presents the evaluation results, and finally, Section 6 concludes this work.

2. Background

This section provides fundamental background information on Code Identifiers (Section 2.1) and the UTAUT technique (Section 2.2), which was employed to assess technology acceptance.

2.1. Code identifiers

Identifiers, often referred to as Code Identifiers, are fundamental elements within the realm of software development. They encompass variable names, method names, class names, and other symbols that developers use to label and reference various elements of their codebase. These identifiers are dynamic in nature, continually evolving as the software project progresses. They play a pivotal role in enhancing code readability, maintainability, and overall comprehension.

In software engineering, the choice of identifier names can significantly impact how effectively developers can understand and maintain a piece of code. Descriptive and meaningful identifiers improve program comprehension, making it easier for developers to grasp the purpose and functionality of different code components. Conversely, cryptic or overly brief identifiers can hinder comprehension and even introduce errors into the codebase (Arnaoudova et al., 2016). Some programming languages, like C++, provide support for aliasing (Hackett and Aiken, 2006), when a data location in memory can be accessed through different identifiers in the program. While this allows different contexts to use different identifiers, this practice can also confuse developers (Arnaoudova et al., 2016).

Given their paramount importance, identifiers have been the subject of extensive research and tool development (Petrescu et al., 2023). Techniques and approaches have been devised to assist developers in choosing appropriate identifier names, managing their consistency, and even automating aspects of identifier maintenance. These efforts underscore the critical role that identifiers play in software development and the ongoing quest to optimize their usage for better code quality and efficiency.

In summary, we can assert that identifiers play two different roles: they function as a means of referencing a software entity and as a comprehensible description of its purpose for who is reading the code (Arnaoudova et al., 2016). These roles often conflict because the identity must remain constant, as highlighted earlier, while the description presented to the programmer would benefit from being more adaptable, capable of varying over time and adjusting to diverse contexts. Thus, the first role of identifiers requires them to be permanently stored in the source code. However, for the second role identifiers could be impermanent and modified according to the context when being visualized by developers.

This paper advocates for the possibility of identifiers having textual representations different from their identity, which we call "*impermanent identifiers*". Freeing the textual representation from the constraints bound to the identity significantly reduces the cost of changing it. Based on this approach, the identifier can be adapted to different contexts, allowing the incorporation of additional information, omission of non-relevant parts, and its representation in another language. In addition, the temporary nature of impermanent identifiers allows them to leverage Unicode characters in a way that would not be practical for permanent identifiers. Using Unicode allows for concisely conveying context-sensitive information in the impermanent identifiers themselves.

For example, consider a Java method with the “synchronized” keyword, indicating that it can be accessed by only one thread at a time. While this information is relevant when developing or maintaining other classes that invoke this method, it is not explicit only by the method name. By using the concept of impermanent identifiers, the visualization of this method name could be changed (such as by adding an emoji) to make that fact explicit. Since the code is not changed, only the visualization, it does not affect any other code that depends on it.

Since the change affects only how the identifiers are visualized and not directly the code, it solves the issues above related to conflicts and changes in the code repository. The cost of renaming identifiers vanishing opens up new use cases for renaming identifiers with increasing frequency and also allows the surfacing of additional information in the identifiers. This practice also differs from aliasing since the identifier would be visualized consistently through the code, and two identifiers would not refer to the same variable simultaneously. Both can potentially improve program comprehension and evolution, allowing each developer to tailor how the code is visualized based on their needs.

Standard IDEs have altered the way they display source code in several ways. Different fonts and colors have been used for syntax highlighting for decades. Incorrect code is often underlined in red, while calls to deprecated methods can be displayed with a struck-out font. IDEs also display warnings or suggestions as clickable icons in the editor’s left margin. However, the space available in the margin is limited, and the warning concerns the entire line.

IDEs have applied more changes to source code’s appearance in recent years. This includes an inline “rename refactoring” interface that allows for quick (permanent) renaming of an identifier; nevertheless, this does not address the consequential implications of the refactoring process. Another example is source code folding, which selectively hides a section of source code (such as the body of an entire method or hiding exception handling code). Visual Studio Code¹ (Microsoft, 2023) can display a “Code Lens”, a textual line with additional information displayed in a smaller font above a method declaration.

JetBrain’s suite of IDEs (JetBrains, 2023) can show “inlay hints”. Inlay hints are utilized to incorporate concise pieces of information immediately into the editor. This particular functionality offers enhanced code comprehension without inducing any interference with the workflow of the developer. A well-known example in this suite is the “parameter hints” in function or method calls: the IDE displays in smaller font the name of the parameter that will be bound to each call argument. Finally, Visual Studio Code’s decoration API can display textual information on the right side of the code (such as who recently committed a line of code).

Academic works have also tackled related problems. The Version Editor (Atkins et al., 2002) used different fonts to show recently added and deleted code as far back as the 1980s. SubText (Edwards, 2005) is a research tree-based programming language that also challenges the notion of textual identifiers. While Subtext is a new language, we aim for compatibility with existing languages. These examples illustrate how visual and block-based programming languages might not rely on textual identifiers. Instead, we explore these notions in more conventional text-based languages. Murphy-Hill and Murphy (2014) discusses factors that recommendations within IDEs must consider. Some academic IDE extensions surface information in the code editor. Sliwinski et al. (2005) displays the “risk” of a code section (obtained by querying software repositories) in the editor’s margin. Code_call_lens extends Visual Studio’s “Code Lens” with runtime and usage information (Janes et al., 2018).

The work by Sulír et al. (2018) presented a systematic mapping of source code editor augmentation tools and approaches. They reviewed 5553 articles published between 1998 and 2018 in two phases — keyword search and reference search. The result was a list of 103 relevant articles and a taxonomy of source code editor augmentation tools with seven dimensions:

1. **Source:** Where does the data representing the augmentation come from?
2. **Type:** What data type does the augmentation represent?
3. **Visualization:** What does the augmentation look like?
4. **Location:** Where is the visual augmentation displayed?
5. **Target:** To what is the augmentation visually assigned?
6. **Interaction:** How can we interact with the augmentation?
7. **IDE:** For what IDE/editor is the augmentation implemented?

In Fig. 1, a chart illustrates how many reviewed tools satisfy each attribute within a specific dimension.

Most tools obtain their data through static source code analysis. The most common data category is “boolean”, indicating the prevalence of binary information representation. The coloring technique is widely used for visualizing the enhancement. Most tools directly display the enhancement within the source code. The most frequent association for enhancement is with lines of code. Informational pop-ups are a common means of interaction with the enhancement. Most tools are implemented within existing IDEs.

2.2. Unified theory of Acceptance and Use of Technology — UTAUT

This section describes the model we used to assess the adoption of code augmentation technology.

The UTAUT (Unified Theory of Acceptance and Use of Technology) is an outcome of a comprehensive review conducted by Venkatesh et al. (2003) on various prior models that address the acceptance and utilization of technology. Models examined during the review include Theory of Reasoned Action (TRA), Technology Acceptance Model (TAM), Motivational Model (MM), Theory of Planned Behavior (TPB), Combined TPB and TAM (C-TPB-TAM), Model of PC Utilization (MPCU), Innovation Diffusion Theory (IDT), and Social Cognitive Theory (SCT). Through this meticulous review, the researchers integrated and unified the key components of these models to develop the UTAUT, which provides a comprehensive framework for understanding and predicting the behavior of technology adoption.

UTAUT 2 (Fig. 2) is a unified theory devised by Venkatesh et al. (2012), serving as a comprehensive framework for examining individual technology adoption behavior. By extending and refining the original UTAUT model, UTAUT 2 further enhances the understanding of the factors influencing individual technology consumption.

As shown in Fig. 2, UTAUT 2 comprises seven distinct constructs, each playing a crucial role in understanding individual technology adoption:

- CT1 – Performance Expectation:** This construct relates to how individuals perceive the extent to which a particular technology enhances task performance.
- CT2 – Expectation of Effort:** Referring to the level of ease or difficulty experienced in using the technology.
- CT3 – Social Influence:** This construct assesses the significance individuals attach to the technology’s use based on its importance to others.
- CT4 – Facilitating Conditions:** Characterized by the availability of support and resources for technology use.
- CT5 – Hedonic Motivations:** Linked to the pleasure, enjoyment, and interest individuals associate with adopting technological innovations.
- CT6 – Learning value:** This construct involves the individual’s Learning Value, which refers to the value they place on education and their perception of its cost-effectiveness.
- CT7 – Habit:** Ascribing meaning to routine, custom, and behavioral patterns.

¹ <https://code.visualstudio.com/> - 01/10/23.

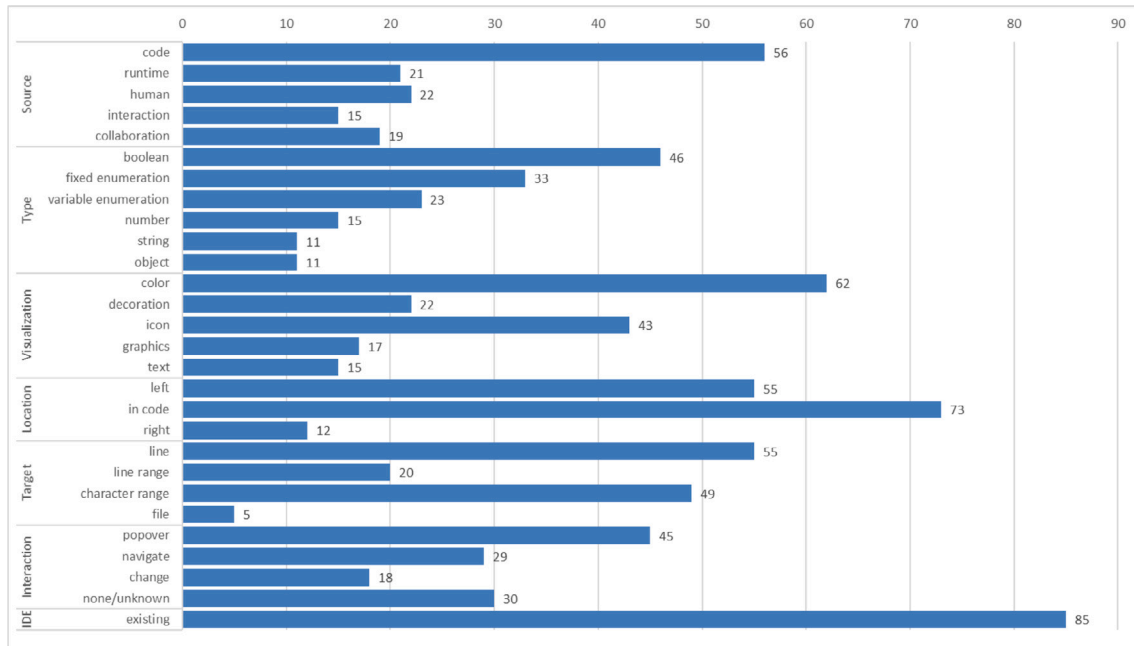


Fig. 1. The numbers of tools fulfilling the given attributes of the dimensions (Sulfr et al., 2018).

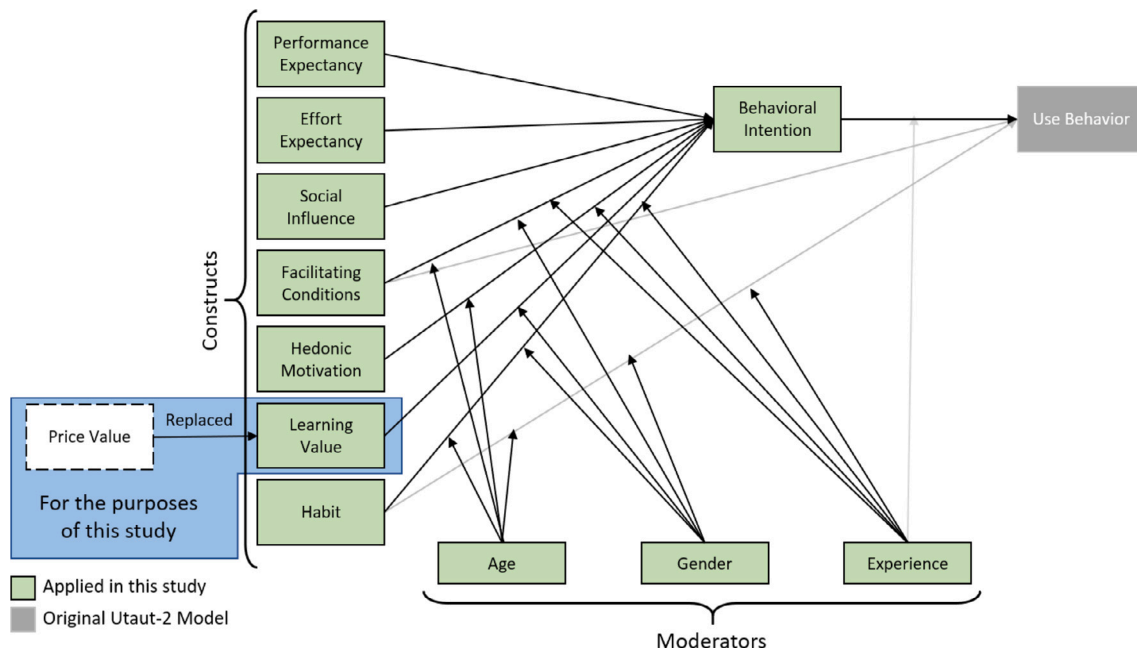


Fig. 2. UTAUT 2 Model.

These constructs serve as independent variables in the UTAUT 2 model, with the dependent variables being Behavioral Intention and, subsequently, Use Behavior. Moreover, gender, age, and experience play a moderating role in this theoretical framework.

The moderators in the UTAUT 2 are additional factors that influence the relationship between the independent variables (constructs) and the dependent variables (Behavioral Intention and Use Behavior). These moderators help to explain how individual differences can impact the technology adoption process. In the UTAUT 2 model, the following moderators are considered:

MT1 – Gender: The effect of gender differences on technology adoption is examined to understand whether

certain technologies are more likely to be adopted by one gender over another.

MT2 – Age: Age-related differences in acceptance and use of technology are analyzed to explore how age influences individuals' willingness to adopt and use technology.

MT3 – Experience: The level of prior experience and familiarity with similar technologies is investigated to determine its influence on individuals' acceptance and use of the new technology.

These moderators provide insights into how various demographic and experiential factors may interact with the core constructs to influence an individual's behavioral intention and actual use of technology.

3. From code augmentation to impermanent identifiers

The evolution of software development tools, such as IDEs (Integrated Development Environments), has brought numerous improvements aimed at making the programming process more efficient and collaborative, especially in the case of code augmentation. However, as IDEs have become increasingly powerful, a common characteristic has persisted: the lack of comprehensive customization for code augmentation functionalities. Many IDEs implement features in a general way, providing little flexibility for developers to tailor the tools to their specific needs. This becomes particularly evident in critical areas, such as code manipulation and team collaboration.

A clear example of this phenomenon is the refactoring of identifiers, such as the renaming operation. Although it is a crucial functionality for improving code readability and maintenance, most IDEs implement this operation in a way that directly affects all team members. Renaming an identifier can have significant implications, and these changes are often uniformly reflected throughout the project, without considering individual preferences or different interpretations of the same code.

This situation raises a fundamental question: is there an alternative approach that allows developers to have more refined customization of their programming experiences while preserving integrity and team collaboration? The answer to this question leads us to explore impermanent identifiers. This approach decouples the textual representation of an identifier from its identity, enabling deeper customization without the complications associated with physical changes to the source code.

3.1. Challenge: Identifiers are hard to change

While an identifier can convey a lot of very useful information to developers, this is not its technical function. The technical function of a software identifier is to *identify* the specific software element it refers to, be it a variable, a function, a method, a class, a package, a module, etc. A computer that executes a program can completely disregard the identifiers themselves, which is why computer-generated code often uses meaningless identifiers.

In text-based languages (the most widely used in practice), the two properties of identifying a software element and describing its purpose to developers are conflated. It makes renaming an identifier more difficult than it should be, as it involves changing all textual occurrences of the identifier to the new name. Moreover, different identifier approaches might be more suitable for developers with diverse backgrounds and roles. For instance, shorter identifiers might make reading more straightforward for those who already know that part of the code, but more descriptive ones might improve the understanding of first-timers.

Even if several IDEs automate the rename refactoring, the execution of manual steps and the presence of naming conflicts can lead to incomplete refactorings and bugs. In some cases, renaming an identifier to an alternative that is more suitable for one developer might make it worse for another,² as in multilingual codebases. In other cases, the refactoring is not accessible to the developer or might affect external code. If the identifier comes from an external library, it cannot be changed by the code that uses it. In this case, an alternative can be to create an adapter that provides access to external services under new names. From the perspective of the API designer, changing a name can potentially break the client's code. Even with mechanisms such as deprecation, evidence shows that they are not very successful, as clients often do not migrate or do it after a very long time (Robbes et al., 2012; Sawant et al., 2016).

Even successful rename refactorings might cause inconvenience, such as causing disorientation in developers used to the previous alternative. This impact can also be noticed in the versioning system,

especially in large-scale text-based renaming: it generates large commits that might tangle and conflict with other changes. This affects tools and empirical studies that evaluate commits since they might overestimate the importance of a commit or lose track of a source code entity (Hora et al., 2018).

All of these factors make renaming an identifier a decision that should not be taken lightly. Thus, developers may avoid renaming opportunities that could benefit code comprehension due to their cost. Arnaudova et al. (2014) surveyed 71 developers on their renaming habits. They found that only 8% of developers said that renaming is straightforward. Regarding cost, most surveyed developers answered that even with the help of automated tools, renaming can have a significant cost (particularly for large-scale renames or ones that break client compatibility).

3.2. Proposal: Impermanent identifiers

This work aims to investigate the feasibility of adapting and adjusting impermanent identifiers according to the specific needs of different software projects or programming contexts. This involves creating a flexible approach that allows developers to customize and shape impermanent identifiers effectively, aiming to enhance code comprehension, facilitate the development process, and promote more efficient and sustainable software maintenance.

In the following subsections, we provide a non-exhaustive list of scenarios in which “Impermanent identifiers” could be useful.

3.2.1. Impermanent identifiers facilitating aliasing

One particularly interesting aspect of impermanent identifiers is that they enable *aliasing*, i.e., to choose a *personal name* for an identifier. In other words, a developer might choose to “rename” identifiers for their own use while not affecting the identifier that everyone else sees. Additionally, the metadata from these personal renaming operations might be shared with other team members, creating name profiles that allow switching between different ways of visualizing the code.

Another important application of aliasing is that it can also be applied to identifiers that come from external code, such as libraries and APIs. This allows the adaptation of the terminology used in external classes and methods to the local context of a particular application. It also allows an API designer to provide alternative names for different user profiles or use cases. For instance, a user that uses the default settings of an API can use a name profile with simple and direct names, while a more advanced user might see more descriptive names that reveal details of the internal behavior.

Automated suggestions. These alias suggestions could come either from automated algorithms or be manually suggested by developers. An instance of automated aliasing can be the change between alternatives for naming conventions, like camelCase, snake_case, or others, such as Hungarian Notation. Automation could also abbreviate and expand names according to certain rules. On the one hand, the abbreviation can be used to reduce the amount of reading and visual overload from the code. One example could be to simplify the name of a `VeryLongJavaLanguageException` to `VLJException` or even just `VLJL` 🛑, with a “stop” Unicode character denoting an exception, and decorated with a “shorts” Unicode character indicating that the name has been shortened.³ On the other hand, expansion can help developers unfamiliar with a certain area of the code in its comprehension. Adding into the name information about parameters and returns can make its functionality more explicit. For instance, a method named `getUsers()` could be automatically expanded to `getUsersByStatus()` by adding information about the arguments.

² A famous quote says (Fowler, 2009): “There are only two hard things in Computer Science: cache invalidation and naming things”.

³ Please note that the precise choice of emojis is not the point of the paper and could even be personalized by each developer or even by domain applications.

Some code annotations can also carry some information that might be useful to be present in the identifier name (Guerra, 2016), such as the ones that can be used to identify the class role or configure additional method behavior (e.g., transactions, access control).

Manual suggestions. Manual aliasing could also be helpful in all the cases mentioned in the previous paragraph. The trade-off is between the possibility of automatically aliasing several identifiers quickly and the capability to add semantic information that could not be inferred through the code. A scenario that could combine both approaches is the translation of the identifiers from an API to a different natural language spoken by some developers. While an initial translation of terms present in the identifiers can be done by an automated tool, manual adjustments can be used to tailor some names to more suitable alternatives related to the application domain.

Supporting change processes. Supporting systematic change processes, such as impact analysis and change propagation, is another application of manually inserted impermanent identifiers. These processes ensure that changes are thoroughly performed by maintaining a list of software entities that need to be checked for changes, marking them with status (e.g., inspection pending, needs to change, no change needed, changes implemented). This list is usually maintained either in a separate tool (e.g., a spreadsheet) or in a dedicated IDE plugin (e.g., JRipples Buckner et al., 2005). Impermanent identifiers might straightforwardly support such a process by simply aliasing the identifiers concerned to show their current status after their name. Since modern IDEs support Unicode, this status can be encoded in a single character (e.g., inspection pending 🟡, needs to change 🔴, no change needed 🟢, changes implemented 🟣). Accordingly, a status change is immediately followed by an action to generate the updated alias for the identifier.

Aliasing vs. renaming. Technically, all these applications are possible with the classical rename refactoring (except when applied to library code and modulo each programming language's support for Unicode). However, in scenarios where the renaming is contextual or personal, they would likely need to be undone before committing code to avoid interference. In practice, this considerable friction makes applying code refactoring for these scenarios unfeasible. Moreover, the possibility to quickly switch between different impermanent identifiers can be used to adapt the code visualization for the task being performed.

3.2.2. Impermanent identifiers surfacing information

Beyond facilitating aliasing, impermanent identifiers make surfacing information more immediate and straightforward by temporarily appending relevant information to the identifier name. The additional information can be textual or, when leveraging Unicode, be significantly more compact.

The information could come from any analysis tool (e.g., source code, software repository). It can be used not only to surface important information while a programmer looks at the identifier definition but also when a programmer encounters a usage of the identifier in the code base. This information may help the user decide if they need further inspecting a software entity's declaration or implementation. Similar mechanisms exist in IDEs, particularly by showing information (warnings, quick fixes, etc.) in the margin of the code. However, while these are limited to information that applies to the entire line, impermanent identifiers can show multiple pieces of information associated with single or multiple identifiers.

In conjunction with aliasing. In the context of aliasing, an analysis could evaluate if an identifier name has quality issues and surface the result as a part of it (e.g., 🏠, 🟡, 🟢, 🟣, or 🏠, depending on the naming issue's severity). The developers would then be prompted to investigate alternative names manually or via a suggestion algorithm. After a recent rename, the refactoring can be shared with the entire development team, and a symbol indicating this status could be appended to the

identifier, such as 🏠 or 🏠^{NEW}. That could mitigate the disorientation associated with renaming by pointing out which identifiers were renamed or aliased, possibly showing the previous name. Since identifiers can be personalized, a developer might even elect to keep the old name as their version.

Surfacing additional properties. However, the approach is more general and can be applied beyond aliasing. An example is presenting static properties of source code entities that are visible only when declared but not when used. This could lead to less source code navigation and fewer context switches, affecting productivity (Meyer et al., 2014; Sanchez et al., 2015). Examples include visibility modifiers (e.g., private 🔒, public 🔓), asynchronous (e.g., async ⏳), and nullability (e.g., null ?). In some JavaScript libraries, asynchronous method or function calls end conventionally with the suffix “Async”, such as `Camera.takePictureAsync()`. An indicator of the presence of some code annotations might also be useful since they can reveal a class architectural role (e.g., persistent entities 💾, controllers 🎮) or method modifiers (e.g., access control 🛂, callback conditions 🔄) (Lima et al., 2022).

The surfaced information can also come from external sources, like analysis tools or the code repository. For instance, according to an analysis of the source code repository, source code entities that have recently changed could display this information (e.g., 🔄). This could help focus code review activities, and small teams might even have an identification of *who* changed it if each team member picks a specific character. Other repository analyses could add information about calls to code that are known to be risky (Sliwerski et al., 2005) (⚠️) or that show signs of decay (🕒). The information can be shown in the identifier declaration and its usage for that application.

Needless to say, static or dynamic analyses (e.g., FindBugs Hovemeyer and Pugh, 2004, etc.) could also be integrated in this fashion. There are a variety of possible indicators of different warnings with several levels of severity (e.g., 🔥, 🚨, 🛑). Advanced static analysis tools such as Infer (Calcagno and Distefano, 2011) could be leveraged to surface information such as a method's estimated complexity or side effects. Also, an analysis of the energy consumption of components (Corral et al., 2015) or APIs (Hasan et al., 2016) could be used to annotate software elements with high energy consumption (e.g., 🔋, ⚡).

System-specific adaptations. Additionally, system-specific rules can also be implemented at the team and individual levels to better reflect project-specific coding conventions. API providers could use impermanent identifiers to highlight usage patterns of their libraries (e.g., to highlight that a call to `open` 📁 should be followed by a call to `close` 🗑️).

Dealing with information overload. Clearly, excessive proliferation of such annotations may have adverse effects. Each identifier annotation should be easily enabled/disabled to customize the environment by each team or user according to each context or activity. For instance, users could selectively enable annotations that are related to code quality issues when they are reviewing code. They may enable annotations related to performance when they are optimizing code, or enable annotations displaying properties such as visibility, architectural roles, when doing code comprehension tasks. Focusing on a specific category of annotations instead of all of them should already reduce the amount of annotations visible at any time, as well as making the annotations more semantically homogeneous. We also envision ordering the surfaced annotations by user or team-defined priority and providing a “slider” to quickly regulate the amount of information that is visible at a glance (e.g., displaying only high-priority annotations by default and adding additional information on demand).

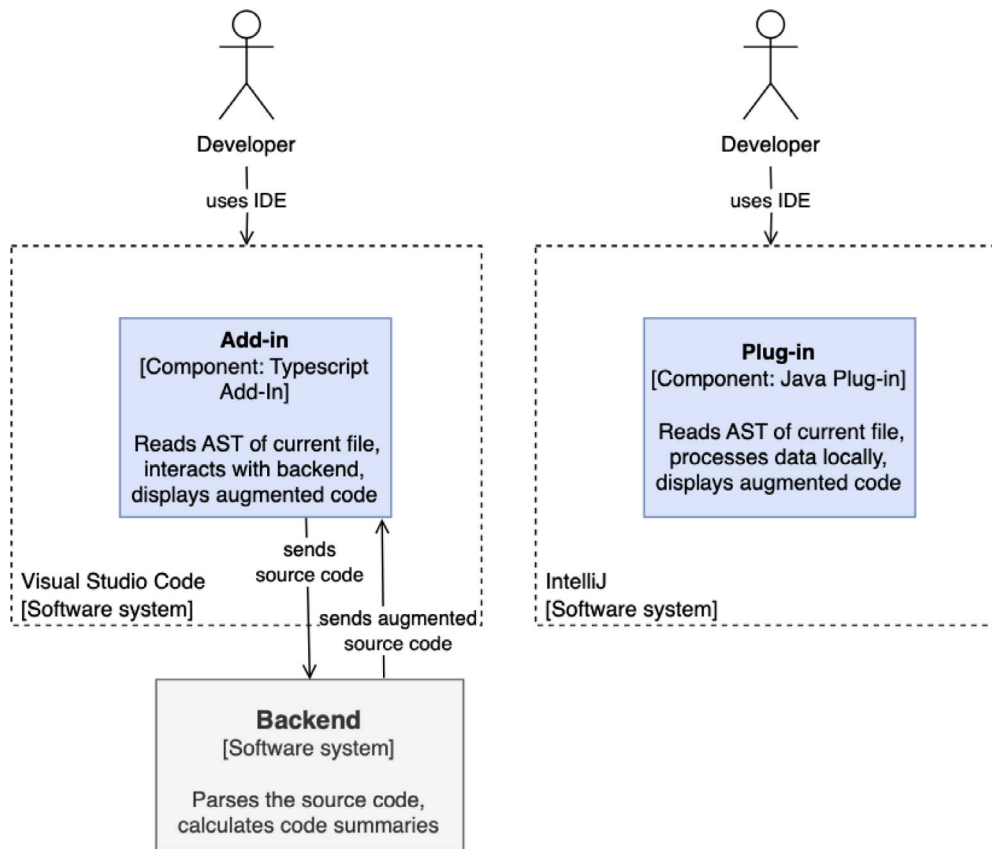


Fig. 3. Component diagram (following the C4 notation) (Brown, 2023) of the first (on the left) and second (on the right) proof-of-concept.

3.2.3. Implementation examples

We created two proof-of-concept implementations in different IDEs to show the feasibility of implementing impermanent identifiers. The first implementation is an IDE extension for Visual Studio Code called “Adverb”⁴ was written in TypeScript and was used to experiment with different options to augment code, e.g., the display of variables in snake_case (while being written in camelCase) or vice versa; also the display of inline parameters was tested (i.e., the visualization of a function `add(2, 5)` as `add(argument1: 2, argument2: 5)` in the editor, obviously without changing the underlying source code. The architecture of this first proof of concept is depicted in Fig. 3 on the left: since code summaries required the use of a large language model, these operations took place on the backend, while the add-in⁵ was responsible for the visualization of the results.

In the future, we do not foresee that there will be a dedicated backend for the plug-in, but there will be interaction with external systems like vulnerability databases, source code quality tools like SonarQube, etc.

The described architecture made it difficult to progress with the project: two parts had to be kept in synch and the use of large language models required to be very attentive not to slow down the developer experience. The source code of the first proof-of-concept is still available at the above mentioned link but was not pursued any further. The second proof-of-concept implementation was developed with a more lightweight architecture in mind, i.e., as a plugin for JetBrains IntelliJ Idea, named “codeEmoji”, which provides support for the Java programming language. Its architecture is depicted in Fig. 3 on the right: a single plug-in reads the abstract syntax tree of the

current source code file, processes the data and displays the augmented code. The codeEmoji plugin is also available on the JetBrains IntelliJ Marketplace⁶ and can be installed directly through the IDE.

The Visual Studio Code extension “Adverb” locates the identifiers to be renamed by traversing the current file’s Abstract Syntax Tree. The extension maintains a list of mappings from impermanent identifiers to their permanent counterparts in a project-level file. If the extension encounters an identifier that should be aliased to an impermanent identifier, it uses Visual Studio Code’s Decoration API (Microsoft, 2022) to visually replace the permanent identifier with the impermanent one.

The extension provides several semi-automated aliasing strategies (such as abbreviating identifiers, changing coding conventions, and using Neural Language Models such as CodeBERT (Feng et al., 2020) for renaming suggestions), as well as allowing fully manual renames. Initial tests show that the proof of concept scales appropriately, even for larger files with a large number of impermanent identifiers. Currently, the proof-of-concept limitations are that impermanent identifiers are disabled when a line is being edited (i.e., the line where the cursor is located) and that we support only Javascript and TypeScript at the moment.

A screen capture of our VSCode extension “Adverb” is shown in Fig. 4 with several impermanent identifiers: an identifier with ① a naming issue; ② a source code warning; ③ an expanded description; ④ an algorithm-generated abbreviation; ⑤ a code lens description that suggests to shorten a method. Furthermore, in this scenario, the developer opted to display identifiers using a snake_case style, and we show that the impermanent identifiers are active decorating the file name with 🐱.

In the codeEmoji plugin, the emojis are used to represent cases involving anti-patterns, such as violations of naming conventions or bad

⁴ <https://github.com/lalenzos/ar>.

⁵ In Visual Studio Code, extensions are called “add-ins”, in IntelliJ, they are called “plug-ins”. We follow this convention in our description.

⁶ <https://plugins.jetbrains.com/plugin/22416-codeemoji/>.

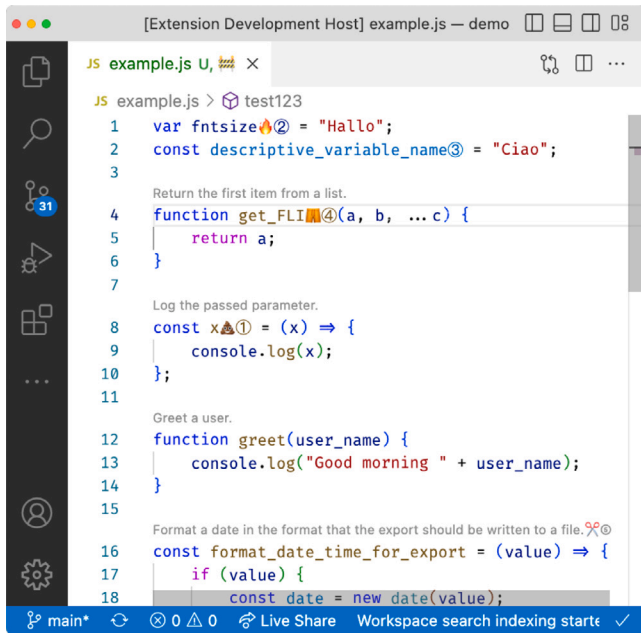


Fig. 4. A screen capture of our VSCode extension.

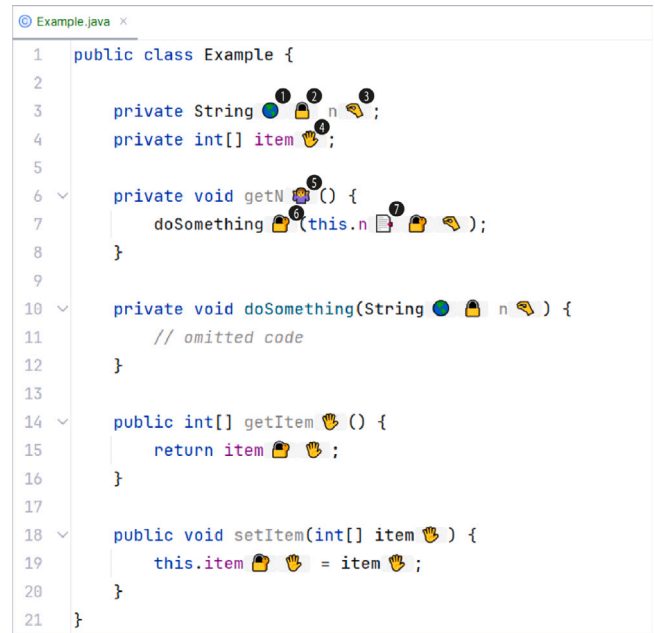


Fig. 5. A screen capture of our IntelliJ Idea plugin (codEEmoji).

programming practices (Arnaoudova et al., 2016). The plugin prototype also displays emojis to represent the use of modifiers in classes, fields, methods, parameters, or local variables. Similarly, the developer has the ability to specify the emoji that will be shown for each of these items based on a predefined rule. For example, emojis can be assigned to a class that is being instantiated and implements a specific interface, a method that is invoked and annotated with a specific annotation, and a local variable of a specific type.

The codEEmoji plugin is under to the terms and conditions of the GNU General Public Licence version 3.0. Using the API for Inlay Hints from the IntelliJ IDE,⁷ the plugin was developed such that it could function as an extensible framework. It is possible for developers to gain more code understanding with the use of inlay hints, which render small amounts of information right into the editor without disruption to the workflow. Emojis are automatically added to the code during coding time when the code is being developed. All details about the implementation can be viewed at the address <https://github.com/codeemoji/codeemoji-plugin>.

A screen capture of the codEEmoji plugin is shown in Fig. 5 with several impermanent identifiers: an identifier with ① a class has public modifier; ② class has final modifier; ③ a naming issue (short descriptive name); ④ a naming issue (says one but contains many); ⑤ bad conventions for method (getter does not return); ⑥ method has private modifier; ⑦ specific emoji for project (fields of String type).

With IDE support, implementing impermanent identifiers does not affect source code parsers. Our current IntelliJ Idea and Visual Studio Code implementations do not change the (permanent) identifiers in the source code. The identifiers are only modified when the code is displayed to the user in the visualization and not directly in the source code; all the IDE tools, such as linters, static analyzers, and code completion, that work on the original code do not need any modification whatsoever. It is important to highlight that this property depends on the specific IDE: for Visual Studio Code, we use the Code's Decoration API; for IntelliJ idea, we use the Inlay Hints API. If an IDE lacks an API that allows us to implement the extension unobtrusively,

an approach that modifies the code would significantly hamper the approach's applicability.

The two presented plugins are currently in the experimental stage. Nevertheless, both serve as evidence of the practical viability of the impermanent identifiers. In both cases, it was possible to show the code with the impermanent identifiers without actually changing the code.

4. Adoption and acceptance evaluation

4.1. Objective

After introducing the concept of “code augmentation”, substantial work remains to be done. The next phase involved evaluating the acceptance and adoption of this new technology. The evaluation focused on understanding how the described scenarios impacted program comprehension and software evolution tasks. To ensure a comprehensive analysis, individual scenario assessments were conducted first, followed by combined scenario assessments to examine the potential effect of information overload.

In this section, we present the evaluation of the adoption and acceptance of the code augmentation technology. We developed a survey based on the UTAUT 2 methodology to conduct this evaluation.

In the context of this study, “Price Value” was replaced by “Learning Value” since technology consumption does not necessarily involve monetary acquisition. However, code augmentation can contribute to the learning process, so “Learning Value” was included as a relevant construct.

4.2. Survey design

A comprehensive survey in the form of a questionnaire was meticulously designed and encompasses both open-ended and closed-ended questions to evaluate the degree of adoption and acceptance of “code augmentation”.

The survey consisted of five distinct modules. Commencing with the initial module, participants were required to confirm their consent to participate and agree to the survey's terms. Subsequently, in the second module, participants were prompted to provide their data, sociodemographic information, and technical profiles.

⁷ <https://plugins.jetbrains.com/docs/intellij/inlay-hints.html>.

Moving forward to the third module, participants were presented with crucial survey instructions and asked to indicate their clarity regarding them. The fourth module introduced 11 application and usage scenarios to assess user acceptance using the UTAUT 2 concept and framework. Finally, the fifth module offered participants the opportunity to express their preferences and perceptions through open-ended questions that sought to elicit overall impressions regarding the code augmentation presented. In this particular case, the participant had the freedom to express their preferred form of feedback, including recommendations, compliments, evaluations, and observations.

The fourth module was purposefully crafted to evaluate 11 distinct code augmentation usage scenarios. These scenarios were designed to gauge varying acceptance levels within each context, recognizing that the benefits of code augmentation technology can differ based on its application context. By assessing the technology's reception across diverse scenarios, we aim to gain insights into its potential impact and effectiveness in different settings. The scenarios considered are described (code examples of each one are available in the survey text, as part of the replication package):

- SC01 – Showing more descriptive names:** Code augmentation can be used to display more descriptive names for methods and classes in their usage, based on the information extracted from their declarations. For example, the methods “addTranslation()” and “translate()” can be expanded to show additional information based on their parameter names.
- SC02 – Showing abbreviated names:** This code augmentation technique can employ rules to shorten method and class names as they appear in the code. In this example, the prefix “get” can be removed from methods, and the names of exception classes can be simplified in the “catch” clause, transforming, for instance, “InvocationTargetException” into “InvTarExp”. Furthermore, the names of well-known method calls in Java, such as “System.out.println()” and “printStackTrace()”, can also be simplified.
- SC03 – Variables with bad naming practices:** This code augmentation showcases variables using improper naming practices that can be enhanced with emojis representing their issue. The emoji 📦 can signify a variable for storing multiple elements, even though it has a singular name. The emoji 📧 can label a variable with just a single letter, lacking meaningful naming.
- SC04 – Recent renames:** This case uses code augmentation to add the emoji 🔄 to a variable name to represent that this variable was recently renamed.
- SC05 – New methods and updated methods:** The code augmentation add the emoji 🆕 to methods that were recently added in the code and the emoji 🔄 in methods that were recently changed. In this scenario, the emojis are added to the method declaration.
- SC06 – Showing the presence of modifiers in method usage:** The code augmentation adds the emoji 🔒 to a method name to indicate that it has the “synchronized” modifier.
- SC07 – Showing the presence of annotation on method and class usage:** The code is augmented to represent the transaction attribute of a method when it is invoked. The emojis 🆕, ➕, and 🚫 represent respectively the configurations “REQUIRES_NEW”, “REQUIRES” and “NOT_SUPPORTED” for the @TransactionAttribute annotation.
- SC08 – Calls to risky code:** This code augmentation to add the emoji 🚨 to a method call from a library that can bring known security risks to the code (identified by an external tool).

SC09 – Problems detected through static analysis tools:

The code augmentation can highlight a bad practice where an exception is printed and not re-thrown, potentially leading to the continuation of execution in a situation where it should be halted. In such cases, an emoji 🚨 is added to the variable name representing the ‘lost’ exception.

SC10 – Visual reminders on API usage: Code augmentation can be used to add reminders, such as providing a visual cue when a PrintWriter instance needs to be closed after usage. In this case, the variable that holds the created instance is annotated with the emoji 📄 to indicate that it is in an open state, and the method used to close it is marked with the emoji 📄.

SC11 – Showing the last person that updated some code: In this scenario, code augmentation can be employed to identify the most recent developer who modified a specific method. Here, the 🧑 and 🧑 emojis are used to symbolize developer avatars.

In this module (fourth) the survey was designed according to the UTAUT 2 methodology, wherein closed-ended questions were specifically formulated. The closed-ended questions are formulated as declarative statements and employ Likert’s unidimensional scale for the response options, prompting the participant to indicate their level of agreement on a continuum ranging from “Strongly Agree” (denoted as 5) to “Strongly disagree” (denoted as 1). For each scenario, 16 questions were applied (Table 1), which were classified according to the constructs predicted by UTAUT 2 (Figs. 2 and 6).

The code snippets, with the respective code augmentation, were presented in the browser so that the participant did not need access to any tool. This is aligned with the study’s goal, which is to evaluate the acceptance and applicability of a wide range of scenarios according to the participants’ perceptions. Moreover, the tools developed aimed to evaluate the feasibility of the approach implementation in current IDEs and do not implement all the scenarios. For each scenario, the participant had access to the scenario description and code snippets that exemplified how it enhanced the code visualization.

The results of the UTAUT 2 survey were then analyzed using descriptive and inferential statistics to provide objective answers to the following research questions (RQ):

- RQ₁ – What is the general acceptance of code augmentation?
- RQ₂ – Which scenarios are more suitable for *Code Review*, *Refactoring*, and *Team Efficiency*?
- RQ₃ – What are the rankings of scenario acceptance?
- RQ₄ – What are the primary findings from each individual scenario?

4.3. Ethical implications

Researchers must carefully consider the ethical implications of any procedures conducted in a survey involving human subjects or sensitive personal information. When risks surpass the minimum threshold, conducting a comprehensive and independent ethical evaluation before beginning any survey activities is essential. The researchers in this study adhered to the guidelines of the participating institutes and universities and concluded that the potential risks were minimal, leading to the study not being evaluated by an Institutional Review Board.

To ensure the protection of participants, the necessary measures were taken to maintain anonymity, safeguard data, and uphold confidentiality. Informed consent was obtained from all participants, focusing on their voluntary and anonymous involvement. Additionally, the study was designed to pose no potential harm to participants.

Table 1

Closed-ended questions were formulated as declarative statements.

#	Declarative statements	UTAUT 2	Technical application
01	This code augmentation would improve my performance in creating new code.	CT1	New Code
02	This code augmentation would improve my performance in understanding and reviewing code.	CT1	Code Review
03	This code augmentation would improve my performance in refactoring and modifying code.	CT1	Code Refactoring
04	This code augmentation would make it easy to create new code.	CT2	New Code
05	This code augmentation would make it easy to understand and review code.	CT2	Code Review
06	This code augmentation would make it easy to refactor and modify code.	CT2	Code refactoring
07	This code augmentation would make my team work better together.	CT3	Team Efficiency
08	If other team members use this code augmentation, that would make me want to use it too.	CT3	Team Efficiency
09	It would facilitate the application of this code augmentation if it is available in my IDE.	CT4	
10	It would facilitate the application of this code augmentation if it is available in web tools to browse the source code, like code repositories and continuous integration	CT4	
11	Using this code augmentation would make me learn new things.	CT6	
12	Using this code augmentation would make me consolidate the learning of things I already know.	CT6	
13	It would be pleasant to work with the code using this code augmentation.	CT5	
14	It would be more fun to work with the code using this code augmentation.	CT5	
15	Using this code augmentation would create good habits when dealing with code.	CT7	
16	After using this code augmentation, it would be hard to work without it.	CT7	
17 ^a	Has a positive attitude towards the functionality of code augmentation and a genuine interest in utilizing it to enhance code comprehension, review, and development.	Behavioral Intention	

^a This observation was derived from the codification process of the qualitative analysis carried out on open questions. The aim is to determine whether there exists a behavioral intention to use, as indicated by positive attitudes expressed during the survey. To determine the degree on Likert's unidimensional scale, the median of the responses from the first 16 questions was calculated, which reflects the extent of their positive attitude towards usage.

4.4. Dissemination

The survey was conducted remotely using Google Forms, with the aim of gathering responses from individuals with professional or academic backgrounds. Participants were invited to participate in the questionnaire and were encouraged to share the link to the survey within their respective academic and professional networks. The survey was distributed to software development experts in private companies and academic institutions.

4.5. Evaluation method

4.5.1. Quantitative evaluation

In order to understand the behavior of system programmers, this study aimed to analyze the main characteristics that influence their intention to use code augmentation technology during the system development process. For this analysis, the UTAUT 2 model was adapted to the use of technology, performing a quantitative evaluation of the survey results using code augmentation. By implementing this methodology, still considering the survey instrument specified in the survey design as described in Section 4.2, it is possible to establish the following hypotheses, according to the factors of the UTAUT 2 model (Fig. 6):

- H_1 – The performance expectation positively influences the intention to use code augmentation.
- H_2 – The expectation of effort positively influences the intention to use code augmentation.
- H_3 – Social influence positively influences the intention to use code augmentation.
- H_4 – The facilitating conditions positively influence the intention to use code augmentation.
- H_4A_1 – The facilitating conditions are moderated by age group.
- H_4G_1 – The facilitating conditions are moderated by gender.
- H_4E_1 – The facilitating conditions are moderated by experience.
- H_5 – Hedonic motivation positively influences the intention to use code augmentation.

H_5A_1 – The Hedonic motivation are moderated by age group.

H_5E_1 – The Hedonic motivation are moderated by experience.

H_6 – Learning positively influences the intention to use code augmentation.

H_6A_1 – The value they place on education is moderated by age group.

H_6G_1 – The value they place on education is moderated by gender.

H_6E_1 – The value they place on education is moderated by experience.

H_7 – Habit positively influences the intention to use code augmentation.

H_7A_1 – The habit are moderated by age group.

H_7G_1 – The habit are moderated by gender.

H_7E_1 – The habit are moderated by experience.

The model was evaluated by checking Cronbach's Alpha (A.C.) (Bland and Altman, 1997), internal consistency using Composite Reliability (C.R.), convergent validity analyzed using Average Variance Extracted (A.V.E.), and discriminant validity estimated using the cross-loading criterion (Henseler et al., 2009).

As suggested by Venkatesh et al. (2012), the UTAUT is employed to identify the pivotal factors affecting an individual's or group's willingness to adopt a new technology. It is not a hypothesis test in the statistical sense; rather, it is a conceptual model crafted to comprehend the underlying factors that influence technology adoption. Consequently, it does not incorporate the traditional null and alternative hypotheses typically associated with statistical tests such as the Student's t-test or the q-square test. Instead, the UTAUT focuses on elucidating the theoretical relationships among the mentioned constructs to explain technology adoption behavior.

In empirical research, null and alternative hypotheses are used to test statistical relationships or differences between variables. This involves formulating a null hypothesis, which typically states that there is no relationship or difference between the variables, and an alternative hypothesis, which asserts that there is a significant relationship or difference. These hypotheses are then tested using statistical methods (Venkatesh et al., 2003, 2012).

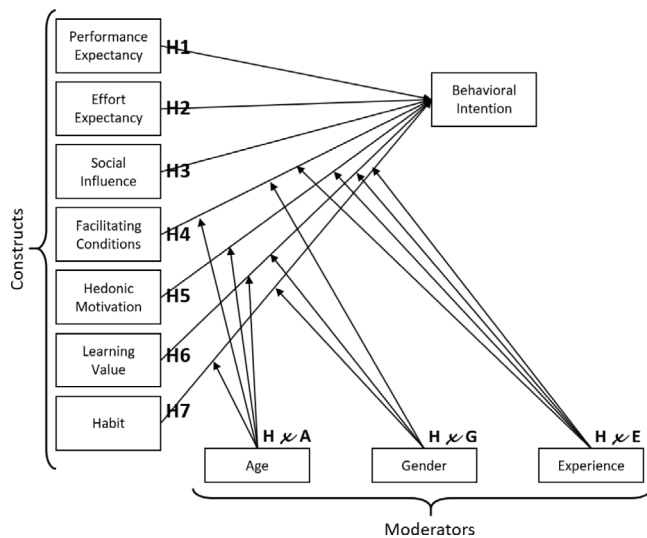


Fig. 6. UTAUT 2 complete model hypothesis.

In the case of UTAUT, it is a theoretical framework that provides a conceptual model for understanding why people adopt or do not adopt technology. UTAUT is not an experiment in itself and does not directly engage in statistical hypothesis testing. Instead, it provides a conceptual framework that researchers can use to formulate their own hypotheses based on the specific variables they are studying and the relationships they wish to investigate (Venkatesh et al., 2003, 2012).

As defined by Venkatesh et al. (2003), in the UTAUT Model, it is important to emphasize that hypotheses are formulated to assess the relationship between constructs, with the goal of formalizing clear expectations regarding the interconnections among variables in the model. Each hypothesis represents a verifiable assumption concerning how these variables are linked, making it easier to conduct specific empirical studies to support or challenge these relationships. This not only improves the testability and efficiency of research but also contributes to the advancement of knowledge in the field of technology acceptance and utilization, as empirical findings either confirm or refute theoretical predictions (Venkatesh et al., 2003, 2012).

For each of the eleven (11) constructed scenarios, the UTAUT 2 model was applied and evaluated in isolation to respond to the defined hypotheses. The collected data was organized and descriptively analyzed using the SmartPLS⁸ software. In order to analyze the most relevant characteristics of the intention to use code augmentation through the studied variables, the study was carried out through the analysis of structural equations.

4.5.2. Qualitative evaluation

In this research, a coding methodology was employed to conduct a comprehensive qualitative assessment, specifically focusing on analyzing the open-ended and free-response segments of the survey. This methodology entails a systematic process of assigning codes to individual responses, which involves a sequential examination of each line of gathered data (Thomas and Harden, 2008).

The inductive approach allows for the emergence of codes and themes organically from the collected responses (as depicted in Fig. 7). In the context of software engineering, it is advisable to execute a series of steps, including data extraction, coding of data, and transformation of codes into thematic patterns (Cresswell, 2012). The systematic process improves our understanding of qualitative insights and facilitates the identification of underlying patterns and recurring themes that may have not yet been predefined.

4.6. Threats to validity

According to Wohlin et al. (2012), this subsection discusses potential threats to the survey in terms of internal validity, external validity, construct validity, and conclusion validity.

1. **Internal Validity:** Threats to internal validity concern the observed causal relationship between treatment and outcome that is not a result of the influence of another factor and cannot be controlled or measured (Wohlin et al., 2012). To mitigate the threat, an approach was adopted that included the joint application of the UTAUT2 model and Structural Equation Modeling (SEM), along with the use of standardized instruments for measuring UTAUT2 variables, the implementation of appropriate statistical analyses considering the UTAUT2 model structure, and continuous monitoring of the model throughout the process.
2. **External Validity:** The threats to external validity concern the generalization of results to industrial practice (Wohlin et al., 2012). This survey was conducted within the context of evaluating the acceptance of identifiers and represents the initial phase of a broader, real-world industrial experiment. Ideally, future studies should involve the construction of practical experiments and replication on a larger scale within an industrial environment. However, all participants involved have extensive experience and are industry professionals, which may somewhat mitigate this threat in the context of this research.
3. **Construct Validity:** Analyses that consider the relationships between theory and observation, that is, whether the treatment (method) accurately reflects the cause and whether the outcome accurately reflects the effect, represent threats to construct validity (Wohlin et al., 2012). To mitigate this threat, the UTAUT2 model was used in conjunction with SEM and descriptive and inferential statistical analyses, such as bootstrapping, which ensure good reliability of the analyses.
4. **Conclusion Validity:** Threats to conclusion validity are related to issues that may affect our analysis of the experiment's results, such as sample size selection, statistical test selection, and caution in implementing and analyzing an experiment (Wohlin et al., 2012). The number of participants is too low to detect a significant effect in this experiment. The study's validity is enhanced by the use of highly qualified participants; however, the data should be viewed as indicative, not conclusive. However, to mitigate this threat and add greater reliability to the survey analysis, the bootstrapping technique was employed. By applying bootstrapping, more robust estimates of parameters and statistics of interest were obtained, thus reducing the risk of erroneous conclusions based on limited or biased samples. The benefits of bootstrapping include the ability to construct confidence intervals and estimate variance and standard error more accurately, as well as to perform robust hypothesis tests that do not rely on specific assumptions about the data distribution.

The UTAUT 2 model is a widely recognized framework for understanding technology adoption and use, especially in the context of information systems and technology. However, when applying UTAUT 2 in the context of Structural Equation Modeling (SEM) (Jun et al., 2019), several threats can emerge that affect the validity and reliability of the results (Venkatesh et al., 2012). Some of the main threats include the following:

1. **Context Specificity:** UTAUT 2 was developed based on empirical studies across different contexts. Specific context characteristics can influence the model's relationships between constructs and variables. Therefore, generalizing the results to a specific context may be a threat if the constructs do not behave similarly.

⁸ <https://www.smartpls.com>.

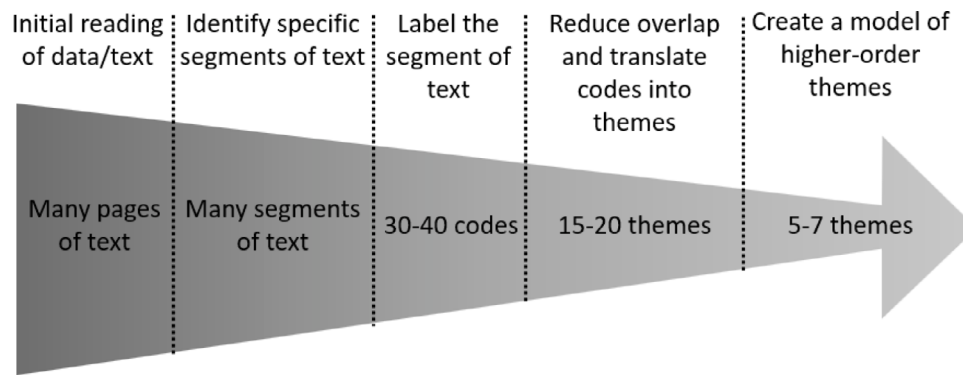


Fig. 7. Coding process for qualitative analysis.
Source: Adapted from Cresswell (2012).

- **Mitigation Action:** To mitigate this threat, it is essential to conduct a contextual validation of the model. In response to this concern, several trials were conducted with a generic dataset on the proposed model to ensure its suitability before being applied to the specific case study data examined in this work.
 - 2. **Construct Measurement:** Proper operationalization and measurement of constructs are crucial in SEM. Any imprecision in measuring latent variables can affect the validity of the results. If measures are not reliable or valid, the model's outcomes may not accurately reflect the expected theoretical relationships.

 - **Mitigation Action:** To address this threat, meticulous attention must be given to construct measurement, utilizing established measurement scales, validation procedures, and sensitivity analyses to ensure the reliability and validity of measures in SEM analysis. Model evaluation involved examining Cronbach's Alpha (CA) for internal consistency, followed by Composite Reliability (CR). Convergent validity was assessed using Average Variance Extracted (AVE), and discriminant validity was determined using the cross-loadings criterion.
 - 3. **Sampling:** The sample quality used for SEM analysis is pivotal. A non-representative or inadequately sized sample can lead to distorted or invalid results. Sampling should be carefully planned to avoid biases and ensure the representativeness of the target population.

 - **Mitigation Action:** To mitigate this threat, it is essential to pay careful attention to the quality of sampling. Using a representative and adequately sized sample is crucial for SEM analysis. Meticulously planning the sampling process helps to avoid biases and ensures the representativeness of the target population. In this study, the bootstrapping technique was employed, which can be used to assess the robustness of the results. By utilizing bootstrapping in SEM, we can generate multiple data samples from the original sample, allowing us to assess the variability of the estimated parameters and calculate confidence intervals for these parameters (Hair, 2014). This helps to mitigate potential biases and increase the reliability of the conclusions drawn in SEM analysis.
 - 4. **Endogeneity:** Causal relationships between constructs can be affected by endogeneity, where variables are interrelated in ways not considered in the model. This can result in spurious effects and distort parameter estimates (Henseler et al., 2009; Bland and Altman, 1997).

 - **Mitigation Action:** To mitigate endogeneity, the discriminant validity technique was used in this study. Discriminant validity helps verify if the model constructs are distinct from each other, minimizing the overlap of latent variables and thus contributing to a clearer interpretation of the specific effects of each construct. By ensuring that the constructs are truly different, discriminant validity reduces the likelihood of unconsidered interrelationships in the model, thereby mitigating endogeneity.
 - 5. **Collinearity:** The presence of collinearity among latent variables can complicate result interpretation. Collinearity may arise when variables are highly correlated, making it difficult to identify which variables contribute independently to construct an explanation (Henseler et al., 2009; Bland and Altman, 1997).

 - **Mitigation Action:** To mitigate the threat of collinearity, this study conducted a diagnostic analysis of multicollinearity to identify highly correlated variables and their impact on the model by calculating variance inflation factors (VIFs) to assess the severity of collinearity.
 - 6. **Respondent Bias:** How participants respond to questionnaires can introduce bias into collected data. This can occur due to socially desirable responses, lack of question comprehension, or even memory errors.

 - **Mitigation Action:** To mitigate the threat of respondent bias, this study implemented a careful questionnaire design to minimize ambiguity and ensure clarity of the questions. Additionally, instructions were developed clearly, and standardized scales were used to help improve participant understanding and reduce response errors.
- Hence, when applying the UTAUT 2 model within SEM, addressing these threats is important through careful research planning, rigorous data collection, meticulous analysis, and appropriate interpretation of findings. Additionally, acknowledging limitations and potential threats to validity when reporting results is crucial for robust and reliable analysis.
- To mitigate threats when applying the UTAUT 2 model in SEM, it is essential to establish strong theoretical validation of proposed relationships, employ appropriate sampling, and ensure data cleanliness.
- Testing SEM assumptions, ensuring measures' convergent and discriminant validity, and addressing endogeneity are crucial steps. Additionally, selecting an appropriate model, conducting cross-validation, sensitivity analyses, and seeking input from expert reviewers contribute to more reliable outcomes. Interpretation of results should be cautious, accounting for limitations and potential biases. Replicating the analysis across different datasets enhances the robustness of the result.

5. Survey results

This section provides a context for the in-depth analysis of the empirical findings of this study. All the data used in this research

can be obtained through the replication package available at the Zenodo.org⁹ repository. The analysis and interpretation of empirical evidence were carried out with a focus on the validity and reliability of the investigated constructs. The objective was to gain a comprehensive understanding of the underlying relationships and their theoretical and practical implications. Additional information regarding the process of analyzing reliability (Cronbach's Alpha, Composite Reliability, and Average Variance Extracted) and the survey results can be found in [Appendix](#) for a more comprehensive understanding.

The assessment of measurement reliability is a fundamental aspect of result interpretation. A reliability test in scientific research is a measure used to assess the consistency and stability of responses or measurements obtained from a data collection instrument (such as a questionnaire, scale, or test) over time or in different situations. Reliability is an important indicator of data quality and measurement consistency that allows for a degree of confidence in extrapolating and generalizing results beyond the tested population group.

The overall observation of Cronbach's Alpha values consistently equal to or higher than 0.6 suggests that each construct's items are internally cohesive, denoting a reliable measurement of the constructs in question.

Composite Reliability scores, employed to assess the internal consistency of constructs, consistently exceed the recommended threshold of 0.7. This trend reinforces the reliability of the measurement, indicating that the constructs are solidly capturing the underlying variability.

Average Variance Extracted (AVE) values provided insights into the extent to which the variance captured by construct items surpasses variance due to measurement error. In many scenarios, AVE values surpass the acceptable threshold of 0.5, signaling satisfactory convergent validity. This suggests that the constructs adequately represent the variability they aim to capture. The coherence and reliability observed in Composite Reliability and AVE scores across various scenarios enhance the robustness of the measurement model. The uniformity of reliability measures across constructs and scenarios suggests that questionnaire items consistently measure intended constructs with high precision and consistency.

In the end, these reliability metrics strengthen the robustness of the measurement model and engender confidence in the data collection procedure, thereby facilitating subsequent analyses and interpretations. In addition, a comprehensive examination of discriminant validity was performed, analyzing the interrelationships between constructs. The investigation of discriminant validity contributes to a greater understanding of inter-concept relationships and provides a comprehensive view of emerging implications for theory and practice.

It is essential to emphasize that, as predicted by [Venkatesh et al. \(2003\)](#), hypotheses are created to assess the relationship between constructs, meaning that one hypothesis is formulated for each predicted relationship. The analysis of the proposed hypotheses yielded significant insights into the relationships between the investigated constructs and their influence on behavioral intention. Eleven distinct scenarios were examined using the bootstrapping method in conjunction with the Student's *t*-test in order to examine these relationships, and the hypotheses were evaluated as follows:

- **Hypothesis H_1** , which posited a positive relationship between **Performance Expectancy** and **Behavioral Intention**, was confirmed in scenario 11 (**SC11**-Person updated). The result was supported by a *t*-value of 3.467, a *p*-value of 0.001, and an R-Square (R^2) of 0.652, indicating a substantial explanatory power of 65.2%.
- **Hypothesis H_2** , suggesting an association between **Effort Expectancy** and **Behavioral Intention**, was corroborated in scenarios 01 (**SC01**-Descriptive names), 02 (**SC02**-Abbreviated names), 06 (**SC06**-Showing modifiers), 07 (**SC07**-Showing annotation), 09

(**SC09**-Problems detected), 10 (**SC10**-Visual reminders), and 11 (**SC11**-Person updated). *t*-values ranged between 2.01 and 2.707, *p*-values ranged from 0.007 to 0.045, and R-Squares (R^2) ranged from 0.547 to 0.652. These findings consistently demonstrate that a favorable perception of effort contributes to a higher behavioral intention.

- **Hypothesis H_3** , exploring the influence of **Social Influence** on **Behavioral Intention**, was validated in scenario 05 (**SC05**-New update). With a *t*-value of 2.009, a *p*-value of 0.045, and an R-Square (R^2) of 0.622, a moderate effect of **Social Influence** on **Behavioral Intention** was evident.
- Conversely, **Hypothesis H_4** , which sought a relationship between **Facilitating Conditions** and **Behavioral Intention**, found no support in any of the analyzed scenarios.
- Regarding **Hypothesis H_5** , exploring the relationship between **Hedonic Motivation** and **Behavioral Intention**, none of the scenarios provided support for this association.
- **Hypothesis H_6** , concerning the association between **Learning Value** and **Behavioral Intention**, was corroborated in scenarios 03 (**SC03**-Bad naming) and 04 (**SC04**-Recent renames). *t*-values ranged from 2.318 to 2.723, *p*-values ranged from 0.006 to 0.02, and R-Squares (R^2) ranged from 0.616 to 0.738. This indicates that learning value positively influences behavioral intention.
- **Hypothesis H_7** , addressing the association between **Habit** and **Behavioral Intention**, was confirmed in scenario 06 (**SC06**-Showing modifiers), with a *t*-value of 2.43, a *p*-value of 0.015, and an R-Square (R^2) of 0.6, underscoring the relevant role of habit in predicting behavioral intention.

The investigated moderators did not demonstrate a consistent moderating effect on the relationships between constructs and behavioral intention in the examined scenarios. In the context of code augmentation, these analyses highlight the significance of specific constructs such as **Performance Expectancy**, **Effort Expectancy**, **Learning Value**, and **Social Influence**. A nuanced comprehension of these relationships contributes to a broader understanding of the determinants of user behavior in this particular context, thereby providing valuable insights for theory and practice.

The analysis of various scenarios and hypotheses sheds light on the factors influencing the intention to use code augmentation. The findings reveal distinct patterns of influence, allowing a comprehensive understanding of the interplay between different constructs and their impact on user behavior.

5.1. Statistical results

Based on a combination of the results obtained from statistical analyses conducted for each hypothesis across different scenarios, it is possible to determine the ranking and classification of the influence of scenarios on the use of code augmentation. *t*-values indicate the size of the statistical effect, while *p*-values indicate statistical significance. A higher magnitude *t*-value suggests a stronger influence and a lower *p*-value indicates greater statistical confidence in the relationship. R-Square (R^2) measures the proportion of variability in the response variable explained by the model.

Combining these indices for each hypothesis and scenario makes it possible to classify and rank the influence of variables on the intention to use code augmentation. Hypotheses and scenarios with higher *t*-values, lower *p*-values, and higher R-Squares (R^2) tend to have a stronger and more significant influence on usage intention. This enables the identification of the most important variables and their corresponding scenarios of greater relevance in explaining the adoption behavior of code augmentation. [Table 2](#) presents the results of hypothesis analysis in different scenarios along with relevant statistical metrics.

Although the hypotheses H_1 and H_2 have a lower R-Square (R^2) compared to H_6 , they are ranked higher on the list due to consideration

⁹ <https://zenodo.org/records/10888469>.

Table 2
Ranking of hypotheses and scenarios based on statistical metrics.

Ranking	Hypothesis	Scenario	t-value	p-value	R^2	Rating
01	H_1 -Performance Expectancy	(SC11) Person updated	3.467	0.001	0.652	High
02	H_2 -Effort Expectancy	(SC11) Person updated	2.111	0.035	0.652	High
03	H_6 -Learning Value	(SC03) Bad naming	2.723	0.006	0.738	Moderate
04	H_3 -Social Influence	(SC05) New update	2.009	0.045	0.622	Moderate
05	H_6 -Learning Value	(SC04) Recent renames	2.318	0.020	0.616	Moderate
06	H_7 -Habit	(SC06) Showing modifiers	2.430	0.015	0.600	Moderate
07	H_2 -Effort Expectancy	(SC01) Descriptive names	2.705	0.007	0.251	Low
08	H_2 -Effort Expectancy	(SC07) Showing annotation	2.090	0.037	0.546	Low
09	H_2 -Effort Expectancy	(SC02) Abbreviated names	2.442	0.015	0.547	Low
10	H_2 -Effort Expectancy	(SC09) Problems detected	2.052	0.040	0.584	Low
11	H_2 -Effort Expectancy	(SC10) Visual reminders	2.707	0.007	0.612	Low
–	H_4 -Facilitating Conditions	–	–	–	–	Not Confirmed
–	H_5 -Hedonic Motivation	–	–	–	–	Not Confirmed
–	Moderators	–	–	–	–	Not Confirmed
					R^2 Mean	0.585
					R^2 Median	0.606
					R^2 StDev	0.117

of factors beyond R-Square (R^2) in determining the ranking. The key determinant for this classification was that hypotheses H_1 and H_2 could be validated by the same scenario.

The classification of hypotheses regarding the use of code augmentation is conducted based on R-Square (R^2), which indicates the proportion of variability in usage intention explained by relationships between constructs. The higher the R-Square (R^2), the stronger the influence of the independent construct (predictor variable) on the dependent construct (usage intention). The classification is divided into three categories:

1. High: Hypotheses with a significantly high R-Square (R^2), indicating that the relationship between constructs has a substantial influence on the intention of use. This suggests that variations in the independent constructs explain a large portion of the variation in usage intention.
2. Moderate: Hypotheses with a moderate R-Square (R^2), indicating that the relationship between constructs has a reasonable influence on usage intention. While not as strong as in the “High” category, there is still a significant influence.
3. Low: Hypotheses with a relatively low R-Square (R^2), indicating that the relationship between constructs has a limited influence on usage intention. In this case, variations in independent constructs contribute less to explaining the variation in usage intention.

Unconfirmed hypotheses are marked as “Not Confirmed” and are not classified, as they did not provide sufficient statistical evidence to demonstrate a significant relationship between constructs.

Moderators were also not confirmed and are classified as “Not Confirmed”, indicating that there is not enough evidence to show that they influence the relationship between constructs in the tested hypotheses.

In summary, the ranking provides insight into the relative influence of each hypothesis and scenario on code augmentation usage, based on a statistical assessment of relationships between constructs. This helps identify factors with a greater or lesser impact on the intention to use this type of technology. The main findings and insights are:

1. High Influence Scenarios

- (a) **Scenario 11 (SC11)**: Showing the last person that updated some code;
 - i. **Findings**: In this scenario, both performance expectancy and effort expectancy significantly contribute to users’ intention to use code augmentation. Users perceive that the technology enhances their performance and is easy to use, thereby increasing their willingness to adopt it.

2. Moderate Influence Scenarios

- (a) **Scenario 03 (SC03)**: Variables with bad naming practices;
 - i. **Findings**: Learning value has a moderate effect on users’ intention to use code augmentation. Users perceive that the technology offers valuable learning opportunities, leading to increased adoption.
- (b) **Scenario 05 (SC05)**: New methods and updated methods;
 - i. **Findings**: Social influence moderately impacts users’ intention to use code augmentation. The opinions and behaviors of peers or colleagues influence users’ decisions to adopt the technology.
- (c) **Scenario 04 (SC04)**: Recent renames;
 - i. **Findings**: Similar to Scenario 3, learning value moderately influences users’ intention to adopt code augmentation. The perception of meaningful learning outcomes drives users’ willingness to use the technology.
- (d) **Scenario 06 (SC06)**: Showing the presence of modifiers in method usage;
 - i. **Findings**: In this scenario, both ease of use (effort expectancy) and habitual behavior significantly impact users’ intention to use code augmentation. Habitual usage and perceived ease of use drive adoption.

3. Low Influence Scenarios:

- (a) **Scenario 01 (SC01)**: Showing more descriptive names;
 - i. **Findings**: While ease of use is a contributing factor, it holds a lower influence compared to other scenarios. Users’ intention to use code augmentation is shaped by the perceived effort required.
- (b) **Scenario 07 (SC07)**: Showing the presence of annotation on method and class usage;
 - i. **Findings**: Ease of use plays a minor role in influencing users’ intention to use code augmentation in this scenario.
- (c) **Scenario 02 (SC02)**: Showing abbreviated names;
 - i. **Findings**: Similar to Scenario 1, ease of use has a relatively lower impact on users’ intention to adopt code augmentation.
- (d) **Scenario 09 (SC09)**: Problems detected through static analysis tools;

- i. **Findings:** Similar to previous scenarios, perceived ease of use has a limited influence on users' intention to adopt code augmentation.

(e) **Scenario 10 (SC10):** Visual reminders on API usage;

- i. **Findings:** In this scenario, as in others, ease of use moderately contributes to users' intention to use code augmentation.

4. **Not confirmed Hypotheses:**

(a) **Hypothesis H_4** (Facilitating Conditions \rightarrow BI): Not confirmed.

- i. **Findings:** Facilitating conditions do not play a significant role in shaping users' intention to use code augmentation in this context.

(b) **Hypothesis H_5** (Hedonic Motivation \rightarrow BI): Not confirmed.

- i. **Findings:** Hedonic motivation has negligible influence on users' intention to adopt code augmentation.

In conclusion, scenarios and hypotheses have varying degrees of influence on the intention of users to utilize code augmentation. **Performance expectancy, effort expectancy, learning value, and social influence** emerge as key motivators, whereas **facilitating conditions** and **hedonic motivation** have limited or no influence.

The R-Square (R^2) Mean of 0.585 indicates that the relationships between constructs explain approximately 58.5% of the variation in usage intention. The higher R-Square (R^2) Median (0.606) suggests that many scenarios exhibit a strong relationship between the constructs and the Behavioral Intention. Remember that the R-Square (R^2) indicates the proportion of variability in the response variable that is accounted for by the model, and higher R-Squares (R^2) generally correspond to a more pronounced and substantial impact on usage intention. In other words, a higher R-Square (R^2) value means that there is a stronger relationship between the constructs of the model and Behavioral Intention.

Combining these indices for each hypothesis and scenario makes it possible to classify and rank the influence of variables on the intention to use code augmentation. Hypotheses and scenarios with higher t-values, lower p-values, and higher R-Squares (R^2) tend to have a stronger and more significant influence on usage intention. This enables the identification of the most important variables and their corresponding scenarios of greater relevance in explaining the adoption behavior of code augmentation. Table 2 presents the results of hypothesis analysis in different scenarios along with relevant statistical metrics.

In conclusion, the R-Square (R^2) statistics collectively suggest that the constructed model adequately captures the variance in the intention to use variables across various scenarios. Although individual scenarios may exhibit varying levels of influence, the model's moderate average R-Square (R^2), higher median, and low dispersion indicate a reasonable and consistent level of explanatory ability. This bolsters the confidence in the model's reliability and its potential to provide meaningful insights into the factors influencing the use of code augmentation.

5.2. Thematic analysis

Thematic analysis is a commonly used technique in social, psychological, educational, and other humanities areas to analyze qualitative data. This approach is particularly useful for identifying, analyzing, and interpreting patterns of meaning (themes) within a dataset.

In the process of thematic analysis, the content of questionnaire responses was carefully examined to identify recurring patterns, common topics, concepts, or emerging ideas in the data.

Following the steps outlined by Fig. 7, the process proceeded objectively as follows:

1. **Familiarization with the data:** Data readings were conducted multiple times to become familiar with the content and gain a general understanding;
2. **Generation of codes:** Codes were assigned to relevant segments of the data reflecting key concepts, ideas, or themes related to transient identifiers and programming practices;
3. **Identification of themes:** Codes were grouped into broader themes based on similarities and conceptual relationships.
4. **Review and refinement of themes:** Themes were reviewed, refined, and adjusted as the analysis progressed, ensuring they adequately captured the essence of the data.

It is important to note that at least two authors were more intensively involved in the process of thematic analysis and code generation; however, all authors reviewed the process and provided critiques until it was in accordance with all authors.

One of the most crucial processes in thematic analysis is the generation of codes, which serves as the foundation for the entire research. While there are several techniques available, in this study, we opted for the method of double coding, where two or more independent researchers coded the same datasets to identify emerging themes. To address discrepancies and ensure the reliability of the results, the following strategy was employed:

1. **Open and collaborative discussion:** The authors engaged in open and collaborative discussions to understand the different perspectives and approaches regarding the data. This involved meetings aimed at reviewing and debating the emerging themes, as well as resolving any conflicts or divergences.
2. **Data review and triangulation:** All authors reviewed and triangulated data from various sources to verify the consistency of emerging patterns and codes.

In cases of disagreement, authors who disagreed on the interpretation of a particular dataset described their differing perspectives and sought to reach a consensus or compromise. In our study, there was only one discrepancy related to the Target Audience, where one author classified it as "Experience". Following a discussion, consensus was reached to use the term "Target Audience". The remaining generated codes were consensually refined and grouped, particularly when one author excessively detailed a code that, in subsequent revisions, was grouped into a broader theme.

It is important to highlight that, overall, the questionnaire was constructed based on the UTAUT2 model and its responses were on the Likert scale. Additionally, two open-ended questions were included at the end for participants to provide their impressions on the use of impermanent identifiers. Thus, thematic analysis was conducted in addition to UTAUT2 in the open-ended questions, making the application of Inter-Rater Reliability (IRR) not suitable.

5.2.1. Data extraction

The initial phase of thematic analysis involved the execution of the data extraction procedure (Cresswell, 2012). The study received responses from 114 different individuals, all anonymously labeled P1 to P114, of which 108 were considered valid and included in the analysis.

It is noteworthy that the study comprises two open-ended questions:

1. What is your opinion about using this code augmentation in general? Is it useful? Is there any specific public that would benefit from it? Do you see any drawbacks or problems in its usage?
2. Do you have an idea of any additional scenarios for code augmentation not covered by one of the scenarios? Or maybe any variation of these scenarios?

The survey gathered feedback from 68 respondents regarding Question 1 and 45 respondents regarding Question 2. It is noteworthy that there is a degree of overlap, as the total number of respondents who answered at least one of the two questions was 68.

Table 3
Themes obtained from the thematic analysis.

	Codes	Total
Theme: Developer experience		
Target audience	Beginners	21
	Experienced	09
	Senior	09
Strengths	Syntax and Design Innovation	07
	Impact on Code Visualization and Reading	28
Weaknesses	Excessive Sensory Stimulation	20
	Ambiguity	35
Theme: Architecture and design		
Strengths	Enhanced Clarity and Readability	28
	Improved Code Quality	24
	Problem Identification Agility	15
	Enhanced Refactoring	11
	Vulnerabilities Identification	09
Weaknesses	(No. applicable)	–
Theme: Learning curve		
Strengths	Enhanced Learning	11
Weaknesses	Blocks learning	05

It is important to highlight that the entire process of thematic analysis was conducted and reviewed by at least two of the researchers involved in this publication.

5.2.2. Coding data

The responses were entered into a spreadsheet, where columns were allocated for questions Q1 through Q2 based on their sequential order in the survey. Rows were designated for participants, identified as P1 through P108. Subsequently, the data encoding procedure was initiated, encompassing a per-question approach across all respondents. In this process, one or more codes were assigned to each of the 108 responses.

To illustrate the process, see some remarkable examples of response text: from participant P10 — “Ones which show possible problems with code are especially useful” adds the code “**Problem Identification Agility**”. From the participant P13 — “If it’s not too much can be very useful to make code review way faster and allows you to focus more on coding”. In this fragment, we can directly infer the code “**Enhanced Refactoring**”. And from the participant P68 — “I see it as applicable. But without the use of emojis, emojis bring a different meaning to each one who reads, this causes greater difficulty than traditional textual writing”, adds the code “**Ambiguity**”.

5.2.3. Translate codes into themes

The process of translating involved categorizing identified codes as either *strengths* or *weaknesses* of code augmentation use and grouping them under relevant top-level themes. Table 3 presents the outcome of the aforementioned translation and the reduction of overlap.

The table groups and identifies strengths and weaknesses associated with different aspects of this approach. The numbers correspond to the total number of individuals expressing each specific perception.

In the strengths and weaknesses Table 3, it is important to recognize that classifying the target audience into Beginners, Experienced, and Senior may initially seem inappropriate for this analysis. After all, this categorization refers to the audience’s level of experience with the product, service, or project in question, rather than directly to its strengths and weaknesses.

However, classifying the Target Audience is crucial for understanding the context in which strengths and weaknesses are assessed. Each segment of the audience has different needs, expectations, and abilities, which directly influence how they perceive and interact with what is being offered.

By understanding the level of experience of the target audience, we can tailor our strategies, resources, and approaches according to their specific needs. For example, the characteristics that are considered strengths for beginners may not be perceived in the same way by experienced or senior users.

Therefore, although the classification of the Target Audience may not be a strength or weakness in itself, it plays a fundamental role in identifying and understanding the variables that shape the strengths and weaknesses of a product, service, or project. It is a key element for developing effective, user-centered strategies.

Thematic analysis reveals that the perception of code augmentation is predominantly positive, with many participants highlighting its benefits in terms of clarity, readability, quality, and the agility of the development process. The applicability of code augmentation to different levels of developer experience is also notable, underscoring its adaptability. However, concerns about potential downsides, such as ambiguity and sensory overload, are also raised. These insights provide a comprehensive view of the pros and cons of code augmentation and can guide decision-making regarding its adoption and implementation in software engineering practices.

5.2.4. Thematic discussion

Through a qualitative examination of the responses, a comprehensive understanding of the strengths, weaknesses, and overall sentiments surrounding code augmentation has been attained.

5.2.4.1. Strengths. One of the primary insights derived from the analysis is the acknowledgment of the usefulness and benefits of code augmentation across different audience. The results of the thematic analysis reveal a positive sentiment, indicating that developers at various audience, ranging from beginners to seniors, recognize the value of implementing code augmentation. Indeed, it was the “**beginners**” code who received the highest number of recommendations, with citations such as:

- “No problems at all, it would greatly benefit a team of beginners in programming”
- “I think there are some interesting use cases that can help specially new developers. ...”

The feedback received indicates that developers recognize the potential for code augmentation to impact the comprehension of code and its visual presentation directly. This insight reflects the significance of clean and intelligible code as a fundamental aspect of software development and highlights the role that code augmentation can play in advancing these qualities. The “**Impact on Code Visualization and Reading**” code who received the highest number of recommendations, with citations such as:

- “...code augmentation proves generally beneficial, as it enhances code visualization and provides additional information ...”
- “I think it’s really useful and would help a better reading of the code”.

These comments underscore the potential usefulness and adaptability of code augmentation techniques, which can cater to developers with diverse levels of experience and expertise. Furthermore, the analysis underscores the importance of enhancing code clarity, readability, and quality through augmentation. In this instance, codes “**Enhanced Clarity and Readability**” and “**Improved Code Quality**” received the most recommendations, with mentions such as:

- Enhanced Clarity and Readability:
 - “...code augmentation in Scenario 1 and Scenario 8 could be valuable for creating self-explanatory code, eliminating the need for line-by-line comments. ...”
 - “Including emojis instead of errors simplifies understanding and reinforces communication between teams. ...”

- Improved Code Quality:
 - “Generating code variations aids bug fixing, enhances performance, and explores features, valuable for large, complex codebases with time-consuming manual reviews”.
 - “Utilizing emojis for code comprehension can be a strategic choice for programmers, accelerating issue resolution; however, excessive emoji use may result in confusion. ...”

5.2.4.2. *Weaknesses.* However, the analysis also raises concerns regarding potential drawbacks. The concept of “**excessive sensory stimulation**” and the risk of “**ambiguity**” in augmented code underscore the importance of a balanced implementation, as indicated by mentions such as:

- Excessive Sensory Stimulation:
 - “...if too many features are used at once, I imagine it can become a bit messy or too much”.
 - “I feel like having too many emoji in the code might hinder its readability (information overload). ...”
- Ambiguity:
 - “...emojis in code augmentations may require an adjustment period due to varying individual interpretations”.
 - “Drawbacks may result from misunderstandings or emoji interpretation lapses, primarily the responsibility of the programmer”.

This finding serves as a cautionary reminder that while augmentation can enhance certain aspects, care must be taken to ensure that it does not compromise the code’s understandability or impede the learning process, such as dealing with information overload. We highlight some possible remedies for this in Section 3.2.2. Regarding ambiguity, some of the respondents propose some solutions, such as adding a textual description of the augmentation when hovering on an emoji.

In addition, some respondents were not receptive to the general idea of using emojis as a mechanism to augment code, e.g., “I don’t believe it’s really useful. Maybe it’s because I’m an ‘old school’ programmer”, or “codes with emoji are the most ridiculous thing I ever saw.”.

5.2.4.3. *Proposed scenarios.* Besides the thematic analysis based on open-ended questions, participants were queried regarding potential scenarios in which code augmentation could prove significantly beneficial. The responses indicate a strong endorsement of the technique, and the mentions provided illustrate that participants recognize its applicability in a variety of other contexts. Such as:

- Visualizing and Enhancing Code Readability
 - “Maybe one augmentation showing features like a deprecated method being used for instance ...”
 - “Using external and internal parameters in Swift method declarations, like ‘serveCoffee(for person: Person)’, enhances code readability”.
- Parameters
 - “...but I really like when you can read the declaration of the method as a sentence - including the parameters”
 - “...maybe emojis for the parameters to understand what type they are ...”
- Syntax and Semantics Enhancements
 - “Show when a variable named with a singular term is used as a type that allows multiple values (Array, List, ...), and the other way around”

- “For the scenario 1, I would prefer if the name of the parameter would be added beside the value rather than in the name: addTranslation(word: ‘buon’, translation: ‘good’) with ‘word’: and ‘translation’: in a different color ...”

• Code Analysis and Program Structures

- “... Yes, thread-safe code in Java ...”
- “... Emojis that indicate polymorphism points ...”
- “... Emojis that indicate composition, injection, and hotspots ...”
- “For RFC (remote calls), could have a remote control/As we have a lot of infrastructure as code and several applications running in the cloud, there could be a scenario that assists in this regard ...”

• Testing and Code Coverage Analysis

- “How about unit test coverage analysis? Alerts for uncovered or potentially buggy code snippets would be beneficial. Also, indicate test coverage for methods. ...”

• Documentation and Coding Conventions

- “An Emoji for all the methods that have not a comment ...”
- “An emoji that alerts the programmer if a standard coding convention is not being respected ...”
- “Could include the emoji in @author for the methods and classes Javadoc ...”

• Code Tracking and History

- “Maybe indicating how long was the last change in each method (e.g., long ago, some time ago, recently). Or which code has been refactored ...”

• Improved Error Messages

- “Using code augmentation only in error messages (to better communicate the error) or warnings. See the Rust error messages as an example - they use heavy code decoration to ...”

The prominence of positive sentiment surrounding the concept of code augmentation indicates its general acceptance and recognition of its usefulness. Developers seem to appreciate the potential of augmentation to expedite refactoring, identify problems, and improve the overall quality of the code. This enthusiasm suggests a positive receptiveness toward incorporating augmentation practices into software development workflows. Developers recognize that augmentation may be more suitable for certain types of projects, audiences, or development stages. This insight underscores the flexibility and potential customizability of code augmentation methodologies to cater to diverse requirements.

5.3. Overall survey discussion

This section presents the discussions and general findings obtained from the analysis of the responses collected in the study on code augmentation through descriptive and inferential statistics, including Structural Equation Modeling (SEM).

Considering the descriptive statistics of [Appendix A.1](#), the scenarios analyzed reveal a general trend of agreement among the participants. The answers indicate that the majority of participants demonstrated agreement or neutrality regarding the proposed scenarios. Response averages consistently ranged between 3.1 and 3.6, suggesting that the opinions remained closely aligned and positively assessed.

The distribution of responses, with a standard deviation generally ranging from 1.0 to 1.2, demonstrates a certain level of variability in opinions, as is typical in surveys of this nature. However, it is worth noting that the responses did not exhibit excessive dispersion relative

to the means, which further underscores the consistency of the evaluations. Consistently observed medians, primarily at 3 (“Don’t agree or disagree”) or 4 (“Agree”), underscore that a majority of participants leaned toward selecting these values. This pattern implies a certain level of consensus in participants’ opinions, enhancing confidence in the outcomes.

Regarding the distribution between response categories, the prominence of the “Agree” is notable and reinforces the positive sentiment of the assessments. Moreover, the relatively lower proportions in extreme categories like “Strongly Agree” and “Strongly disagree” imply a sense of equilibrium and significance in the conveyed opinions.

Considering an overview of responses across all scenarios, the “Agree” category continues to emerge as the most prevalent, further underscoring the positive tendency in participants’ feedback. Initial descriptive analysis serves as a strong indicator to address the research question *RQ₁*.

RQ₁ - What is the general acceptance of code augmentation?

Answer: Considering the analysis of Appendix and an overview of responses across all scenarios, the ‘Agree’ category stands out, suggesting a trend of positive acceptance with some cohesion in opinions and moderate variability. Thus, it can be concluded that the results indicate a favorable reception of the technology.

Building upon the analysis of descriptive statistics, we can evaluate in which scenarios the practice of code augmentation is more suitable, thus addressing the research question *RQ₂*.

RQ₂ - Which scenarios are more suitable for Code Review, Refactoring, and Team Efficiency?

Answer: Among the analyzed scenarios, the following scenarios are more favorably indicated for each of the technical applications:

1. **Code Review: (SC01)** Descriptive names: Code Review has the highest average rating (4.1) among all four applications. This scenario seems to be particularly suitable for code review activities, as it received the highest average evaluation in this aspect.
2. **Code Refactoring: (SC01)** Descriptive names, (SC05) New update, (SC06) Showing modifiers, (SC08) Risky code, (SC09) Problems detected, (SC10) Visual reminders receive relatively higher average ratings (ranging from 3.5 to 3.9) across these scenarios. This suggests that participants found these scenarios conducive to code refactoring activities.
3. **Team Efficiency: (SC01)** Descriptive names, (SC08) Risky code, (SC09) Problems detected, (SC10) Visual reminders, (SC11) Person updated receives relatively higher average ratings (ranging from 3.4 to 3.5) across these scenarios. These scenarios are more conducive to activities related to team collaboration and efficiency.

Note that these indications are based on average evaluations provided by participants in the study. The appropriateness of specific scenarios for each application may vary depending on the context and specific objectives of the software development process. Additionally, other factors such as project requirements, team dynamics, and individual preferences should also be taken into account when determining the most suitable scenarios for review, coding, and refactoring activities. To address these questions, inferential statistics were employed using Structural Equation Modeling (SEM), which directly tackles the research question *RQ₃*.

RQ₃ - What are the rankings of scenario acceptance?

Answer: By combining outcomes from various statistical analyses for each hypothesis across different scenarios, we can establish the sequence and categorization of the impact of the scenarios on the utilization of code augmentation. Techniques including Bootstrap and Student’s *t*-test were employed to compute statistical parameters like

t-value, *p*-value, and R-Square (R^2), thus facilitating the derivation of rankings and insights, as demonstrated in Table 2, ranked as follows:

1. **Scenario 11 (SC11)** - Showing the last person that updated some code;
 2. **Scenario 03 (SC03)** - Variables with bad naming practices;
 3. **Scenario 05 (SC05)** - New methods and updated methods;
 4. **Scenario 04 (SC04)** - Recent renames;
 5. **Scenario 06 (SC06)** - Showing the presence of modifiers in method usage;
 6. **Scenario 01 (SC01)** - Showing more descriptive names;
 7. **Scenario 07 (SC07)** - Showing the presence of annotation on method and class usage;
 8. **Scenario 02 (SC02)** - Showing abbreviated names;
 9. **Scenario 09 (SC09)** - Problems detected through static analysis tools;
 10. **Scenario 10 (SC10)** - Visual reminders on API usage.
- Scenario 08 (SC08)** - Calls to risky code) has not undergone validation, thus resulting in its non-classification.

Still based on inferential statistics (SEM), the analysis of various scenarios through statistical methods provides insights into the factors influencing users’ intentions to utilize code augmentation. These scenarios are categorized (Section 4.2) based on the degree of influence they exert on users’ perceptions and decisions. The following explanations elucidate the nature of each scenario and help answer the research question *RQ₄*.

RQ₄ - What are the key points from each individual scenario?

Answer:

1. **Scenario 01 (SC01):** Showing more descriptive names;

In Scenario 01 (SC01), Hypothesis H_2 (Effort Expectancy → Behavioral Intention) is confirmed. However, the insight drawn is that ease of use plays a relatively lower role in influencing users’ intentions compared to other scenarios. The willingness of users to use code augmentation is more shaped by their perceived effort required.

2. **Scenario 02 (SC02):** Showing abbreviated names;

Similar to Scenario 01 (SC01), Hypothesis H_2 (Effort Expectancy → Behavioral Intention) is confirmed. Ease of use has a relatively lower influence on users’ intentions to adopt code augmentation.

3. **Scenario 03 (SC03):** Variables with bad naming practices;

The confirmation of Hypothesis H_6 (Learning Value → Behavioral Intention) in this scenario indicates that participants perceive code augmentation as a valuable learning tool. The moderate influence observed suggests that users are inclined to adopt the technology due to its potential to improve their learning experiences.

4. **Scenario 04 (SC04):** Recent renames;

Hypothesis H_6 (Learning Value → Behavioral Intention) is confirmed again in this scenario, reiterating the moderate influence of learning value on users’ intentions. Users perceive meaningful learning outcomes as a motivating factor for adopting code augmentation.

5. **Scenario 05 (SC05):** New methods and updated methods;

In this scenario, Hypothesis H_3 (Social Influence \rightarrow Behavioral Intention) is confirmed, indicating that social factors play a moderate role in shaping users' intentions to use code augmentation. The influence of peers and colleagues on users' decisions to adopt the technology is highlighted, showing the significance of social influence.

6. **Scenario 06 (SC06):** Showing the presence of modifiers in method usage;

Two hypotheses, H_2 (Effort Expectancy \rightarrow Behavioral Intention) and H_7 (Habit \rightarrow Behavioral Intention), are confirmed in this scenario. The insight derived suggests that habitual usage, combined with the ease of use, significantly contributes to users' intentions to use code augmentation.

7. **Scenario 07 (SC07):** Showing the presence of annotation on method and class usage;

Hypothesis H_2 (Effort Expectancy \rightarrow Behavioral Intention) is again confirmed in Scenario 07 (SC07). The insight is that ease of use has a minor impact on users' intentions to use code augmentation in this scenario.

8. **Scenario 08 (SC08):** Calls to risky code;

(No hypothesis confirmed)

9. **Scenario 09 (SC09):** Problems detected through static analysis tools;

In this scenario, Hypothesis H_2 (Effort Expectancy \rightarrow Behavioral Intention) is confirmed, with a similar insight that perceived ease of use has limited influence on users' intentions to adopt code augmentation.

10. **Scenario 10 (SC10):** Visual reminders on API usage;

Hypothesis H_2 (Effort Expectancy \rightarrow Behavioral Intention) is once again confirmed, revealing that ease of use moderately contributes to users' intentions to use code augmentation.

11. **Scenario 11 (SC11):** Showing the last person that updated some code;

In this scenario, two key hypotheses, H_1 (Performance Expectancy \rightarrow Behavioral Intention) and H_2 (Effort Expectancy \rightarrow Behavioral Intention), were confirmed. The insight drawn is that participants highly value the performance enhancement and ease of use provided by code augmentation, resulting in a stronger intention to adopt this technology. The positive relationship between performance expectancy and intention reflects users' belief that the technology positively impacts their performance. Likewise, the positive influence of effort expectancy underscores the perceived simplicity and convenience of using code augmentation.

In summary, the analysis of the results reveals that high-influence scenarios, such as Scenario 11 (SC11), play a crucial role in users' intention to adopt code augmentation, driven by positive expectations of performance and ease of use. Additionally, moderate-influence scenarios like Scenarios 03 (SC03), 04 (SC04), and 05 (SC05) also have a significant impact due to the perception of valuable learning and social influence. On the

other hand, low-influence scenarios like Scenarios 01 (SC01), 02 (SC02), 07 (SC07), 09 (SC08), and 10 (SC10) underscore less prominent factors such as ease of use. The unconfirmed hypotheses of **Facilitating Conditions** and **Hedonic Motivation** suggest that other elements carry more weight in adoption decisions. Consequently, Scenario 11 (SC11) emerges as the most relevant in driving code augmentation adoption based on users' intention to use.

5.4. Practical implications

Based on the results of the study presented in this paper, we can raise some practical implications of it for different audiences:

- **Educators:** The results of this particular research, especially those evaluated by Scenarios 03 (SC03) and 04 (SC04) confirming hypothesis H_6 , with t-values indicating statistical significance and R-Squares demonstrating a positive influence of learning value on behavioral intention. Thus, impermanent identifiers can be utilized as an educational resource to augment source code with additional information. This information may draw students' attention to important issues such as code quality, security, and structure. Even demonstrating basic properties of code elements, such as field visibility or whether a method is synchronized, can also prove beneficial when teaching programming.
- **API Designers:** The results of this research, particularly through the confirmation of hypotheses H_1 , H_2 , and H_6 , as explicitly expressed in the thematic analysis in statements such as “*help specially new developers*”, suggest that impermanent identifiers can be employed to enhance the usability of APIs by providing additional information when developers are using or maintaining code that utilizes the API. As this application has yet to be explored, various new API design practices and standards may emerge from this opportunity.
- **Tool Builders:** This work highlighted the potential and the implementation feasibility of using impermanent identifiers. However, existing tools are not yet mature in this regard, as indicated in the results of the thematic analysis of this study. The thematic analysis identified concerns such as “*Excessive Sensory Stimulation*” and issues with “*Ambiguity*”. Additionally, the API for implementing inlay hints in the IntelliJ IDE, used in codeEmoji, is currently classified as unstable. Our results point to a promising direction for tools that adopt this approach, and tool creators may consider it as an opportunity beyond an alternative or complement to dashboards, other visualizations, and code editor warnings.
- **Developers:** Software developers are the main public for this work, which can directly benefit from an enriched view of the code. As highlighted by the answer to RQ_2 , some scenarios might be more interesting to some activities, like code review or refactoring. When teams start adopting this approach in real projects, some new patterns might be identified in its usage.
- **Researchers:** This work also presents numerous opportunities for researchers interested in exploring the concept of impermanent identifiers in various directions. The following are examples of fields where the proposed approach could be further investigated: Developer eXperience (DX) to address the issues of “*Excessive Sensory Stimulation*” and “*Ambiguity*” identified in thematic analysis, as well as other areas such as code analysis tools, software design, and software engineering education.

6. Conclusion

This article presents an innovative approach to code augmentation using Impermanent identifiers. The goal goes beyond enhancing the software development experience to encompass a comprehensive

analysis of the adoption and acceptance of this technology in the development process. For this purpose, this study employed the techniques of the UTAUT model and inferential statistics to assess the acceptance of this approach.

This study's statistical analysis has yielded a thorough comprehension of code augmentation application scenarios and their influence on the software development process. Through evaluating various scenarios and hypotheses, insights have emerged that influence how developers perceive and adopt code augmentation.

The results indicate that various factors with different degrees of influence impact the adoption of code augmentation. In scenarios with a high degree of influence, such as Scenario 11, performance and effort expectations are crucial to the decision to adopt code augmentation. The significance of these factors highlights the necessity of enhancing the functionality and efficacy of code augmentation tools to promote their widespread adoption.

In scenarios with moderate influence, such as Scenarios 03, 04, and 05, factors such as learning value and social influence play a significant role in the intention of developers to employ code augmentation. This emphasizes the significance of fostering collaborative learning and encouraging peer collaboration in order to increase the adoption of code augmentation.

In scenarios with limited influence, such as Scenarios 01, 02, 07, 09, and 10, the impact of usability on the decision to adopt code augmentation is minimal. Although these factors may not be decisive, they still influence the decisions of developers. This underscores the importance of ensuring that code augmentation tools are user-friendly and seamlessly incorporated into the workflow, even when usability is not the determining factor.

The unconfirmed hypotheses regarding facilitating conditions and hedonic motivation suggest that these variables have little effect on adopting code augmentation. This indicates that developers place a greater emphasis on effectiveness, performance, and collaboration than on auxiliary support factors.

Performance expectation, effort expectation, learning value, and social influence emerge as important motivators, whereas facilitating conditions and hedonic motivation exert little or no influence.

The thematic analysis of code augmentation feedback provides insights into its prospective impact on software engineering. The findings emphasize the importance of mitigating risks associated with ambiguity and sensory overload while highlighting the benefits of improved code clarity, legibility, and quality. The recognition of code augmentation's applicability across experience levels and its potential to expedite problem-solving and refactoring bodes well for its incorporation into software development practices. However, this analysis also highlights the need for a deliberate and well-balanced approach that protects the integrity of code comprehension and learning.

Certainly, based on the analysis conducted, the following findings were considered for each research question:

- **Research Question (RQ₁):** *What is the overall acceptance of code augmentation tools?*

Findings: The study revealed a positive trend in accepting code augmentation technology, with participants showing consensus and favorable reception towards its use. This indicates code augmentation tools' potential value and relevance in software development practices.

- **Research Question (RQ₂):** *Which scenarios are most suitable for Code Review, Refactoring, and Team Efficiency?*

Findings: Specific scenarios were identified as particularly conducive to various technical applications within code augmentation. These findings provide insights for optimizing code review, refactoring, and team efficiency processes, helping developers make informed decisions to improve their development workflows and productivity.

- **Research Question (RQ₃):** *What are the acceptance ratings of the scenarios?*

Findings: Through rigorous statistical analysis, acceptance ratings of the scenarios were established, highlighting their relative impact on the utilization of code augmentation. These ratings offer guidance for prioritizing scenarios for further exploration and development, contributing to the advancement of code augmentation technologies.

- **Research Question (RQ₄):** *What are the key points of each individual scenario?*

Findings: Detailed examination of individual scenarios revealed insights into the factors influencing users' intentions to adopt code augmentation. These insights provide information for designing and implementing code augmentation tools that align with users' needs and preferences, paving the way for future research and development efforts aimed at improving the usability and effectiveness of these technologies.

With a strategic approach, code augmentation can play an important role in the software development process. These scenarios provide fundamental guidelines for designing code augmentation tools that meet the requirements of developers, optimize performance, and integrate seamlessly into the workflow.

As software engineering continues to evolve, code augmentation emerges as a modern approach that has the potential to improve the development process, provided that its implementation is meticulously tailored to each project's specific requirements.

Given that the primary focus of this research has been on evaluating the technological acceptance of the tool, and the results indicate a positive acceptance, future work should consider the implementation and practical validation of the approach. Taking into account the initial version of the tool, which encompasses the most fundamental cases of code augmentation, the next stage involves expanding the tool's functionality to encompass more complex scenarios, along with conducting experimental studies aimed at assessing the practical effectiveness of the technique. The analysis of the tangible impact of the technique in various settings emerges as a crucial element in this next phase of the project.

Acknowledgment of funding

The authors would like to extend their profound gratitude to the Free University of Bozen-Bolzano, Italy for the generous funding and support provided for the ADVERB project. The collaboration with the institution and the financial support were instrumental in the completion of this study.

This work was partially funded by the ADVERB project, with support from the Free University of Bozen-Bolzano, Italy, and we acknowledge the significance of this contribution to the success of this research project. We appreciate the Free University of Bozen-Bolzano for their belief in our research and their ongoing commitment to advancing knowledge and innovation.

CRediT authorship contribution statement

Eduardo Martins Guerra: Writing – review & editing, Writing – original draft, Validation, Supervision. **André A.S. Ivo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Formal analysis. **Fernando O. Pereira:** Writing – review & editing, Writing – original draft, Validation. **Romain Robbes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision. **Andrea Janes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision. **Fábio Fagundes Silveira:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared.

Appendix. Survey analysis details

The survey was distributed through email by providing a hyperlink to the questionnaire and was accessible for six (6) months to collect responses. Upon the designated period's conclusion, 114 individuals had responded to the survey.

Surveys like this are very susceptible to motivational bias, which refers to how dedicated the participant is to answering and how well the answers capture the individual's true feelings. To mitigate this occurrence, participants with a standard deviation smaller than 0.5 who did not answer the open questions were deleted from the data set. Six (6) participants were excluded because they met the criteria.

The survey profile reveals a significant predominance of male participants, accounting for 89.8% (97) of the total sample. On the contrary, the presence of women in the survey is relatively lower, comprising only 8.3% (9) of the participants. Furthermore, a tiny portion of the sample, equivalent to 0.18% (2), chose not to respond to the gender-related question. Table 4 provides an overview of the subjects' key demographic and experiential characteristics, while Fig. 8 shows a participants' Age and Experience Box plot.

The average age of the participants was 28.3 years, with a standard deviation of 8.4. The age range varied from a minimum of 19 years to a maximum of 51 years. The median age, representing the middle value of the data set, was 25 years. This data distribution indicates a relatively diverse age range among the participants.

The average years of experience among the participants was 8.1 years, accompanied by a standard deviation of 7.1. Experience ranged from a minimum of one year to a maximum of 31 years. The median experience was five years. This range highlights a notable variability in the levels of experience within the group of participants.

Table 4
Statistical overview of participants' age and experience.

	Mean	StDev	Min	Median	Max
Age	28.3	8.4	19	25	51
Experience	08.1	7.1	01	05	31

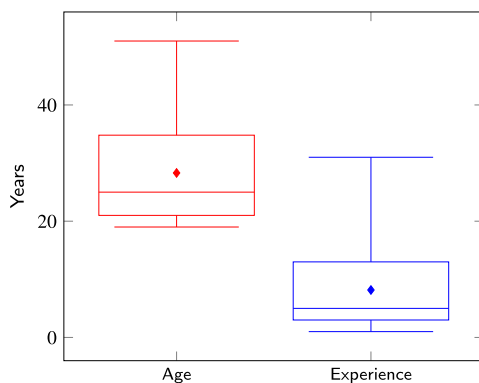


Fig. 8. Participants' age and experience box plot.

Table 5

Descriptive statistics of participant responses across scenarios.

Scenario	Mean	StDev	Min	Median	Max
(SC01) Descriptive names	3.6	1.0	1	4	5
(SC02) Abbreviated names	3.1	1.2	1	3	5
(SC03) Bad naming	3.2	1.2	1	3	5
(SC04) Recent renames	3.1	1.1	1	3	5
(SC05) New update	3.3	1.2	1	3	5
(SC06) Showing modifiers	3.4	1.1	1	4	5
(SC07) Showing annotation	3.3	1.1	1	4	5
(SC08) Risky code	3.5	1.1	1	4	5
(SC09) Problems detected	3.5	1.1	1	4	5
(SC10) Visual reminders	3.5	1.1	1	4	5
(SC11) Person updated	3.1	1.2	1	3	5

A.1. Overview

Based on the study of the provided situations, a general pattern of consistent scores throughout the numerous scenarios may be detected when Likert's unidimensional scale is used. As already mentioned, participants were asked to rate their level of agreement on a scale ranging from "Strongly Agree" (denoted as 5) to "Strongly Disagree" (denoted as 1).

According to Table 5, the average scores in all situations ranged from 3.1 to 3.6, with a standard deviation between 1.1 and 1.2. This shows that, on average, the situations were evaluated closely, and individual differences in the scores were not excessively large. In terms of score centrality, the median remained stable at around 3 or 4 in all scenarios. This suggests that, while individual scores varied, the majority of judgments centered on these mean values.

Fig. 9 shows a scenarios Box plot statistics.

The statistical analysis of the contingency Table 6 reveals significant patterns regarding the utilization of code augmentation.

There is a clear tendency toward high levels of agreement across all scenarios. For example, in Scenario 01, the majority of responses are divided between "Strongly Agree" (23.6%) and "Agree" (36.5%), showing a high level of agreement among participants. Similarly, in other scenarios (02–10), the "Strongly Agree" and "Agree" categories predominate, indicating an overall tendency toward agreement. Fig. 10 depicts a Graphical Scenarios Contingency table.

However, this trend is not consistent. In Scenario 11, for example, the responses are more evenly distributed, spanning all categories in equivalent quantities. This indicates a more significant difference of opinion in this specific circumstance.

In some circumstances, the "Don't Agree nor Disagree" category stands out, reflecting nuances in participants' perspectives and maybe a lack of clarity in their positions. Similarly, the inclusion of the "Strongly Disagree" category in variable proportions across different scenarios emphasizes the existence of opposing opinions, particularly in specific contexts.

The constancy in median categories and generally consistent standard deviations across scenarios suggest that the opinions of the participants are stable. On the other hand, fluctuation in the percentage distribution of each category across multiple scenarios demonstrates the complexities of individual perceptions in the face of diverse given contexts.

In summary, despite the occurrence of disagreement in some instances, the data show a consistent pattern of moderate to strong agreement. This underscores the importance of considering nuances in interpretations while acknowledging the overall degree of agreement among participants.

According to Table 7, the "New Code" application has a moderate average score of 3.3, a relatively constant standard deviation of 1.1, and a median of 3. The "Code Review" application, on the other hand, has a slightly higher average score of 3.5, as well as a moderate dispersion indicated by a standard deviation of 1.2 and a higher median of 4.

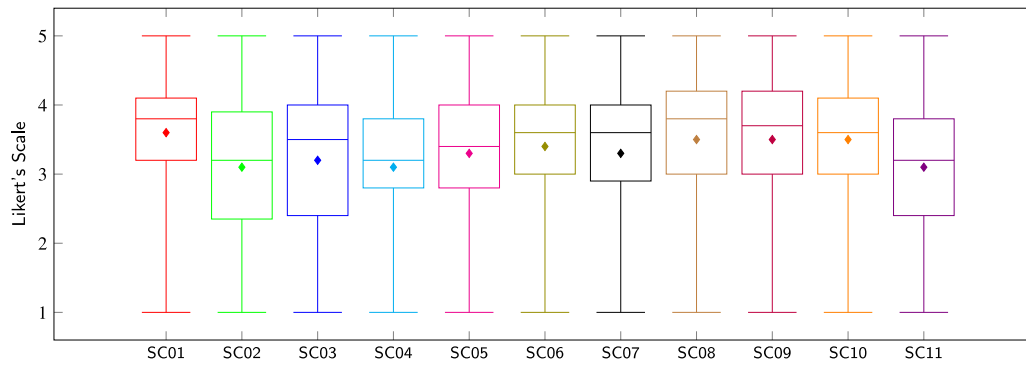


Fig. 9. Participants' scenarios box plot.

Table 6
Scenarios contingency table.

Scenario	Strongly Agree (5)	Agree (4)	Don't agree or disagree (3)	Disagree (2)	Strongly disagree (1)
(SC01) Descriptive names	408 (23.6%)	632 (36.5%)	424 (24.5%)	207 (11.9%)	057 (03.2%)
(SC02) Abbreviated names	267 (15.4%)	434 (25.1%)	459 (26.5%)	350 (20.2%)	218 (12.6%)
(SC03) Bad naming	294 (17.0%)	549 (31.7%)	366 (21.1%)	349 (20.1%)	170 (09.8%)
(SC04) Recent renames	246 (14.2%)	430 (24.8%)	605 (35.0%)	305 (17.6%)	142 (08.2%)
(SC05) New update	342 (19.7%)	471 (27.2%)	469 (27.1%)	304 (17.5%)	142 (08.2%)
(SC06) Showing modifiers	299 (17.3%)	618 (35.7%)	459 (26.5%)	198 (11.4%)	154 (08.9%)
(SC07) Showing annotation	284 (16.4%)	595 (34.4%)	490 (28.3%)	234 (13.5%)	125 (07.2%)
(SC08) Risky code	426 (24.6%)	552 (31.9%)	432 (25.0%)	180 (10.4%)	138 (07.9%)
(SC09) Problems detected	377 (21.8%)	603 (34.8%)	442 (25.5%)	170 (09.8%)	136 (07.8%)
(SC10) Visual reminders	370 (21.4%)	565 (32.6%)	505 (29.2%)	147 (08.5%)	141 (08.1%)
(SC11) Person updated	263 (15.2%)	461 (26.6%)	453 (26.2%)	330 (19.0%)	221 (12.7%)
Total	3576 (18.8%)	5910 (31.0%)	5104 (26.8%)	2774 (14.5%)	1644 (8.6%)

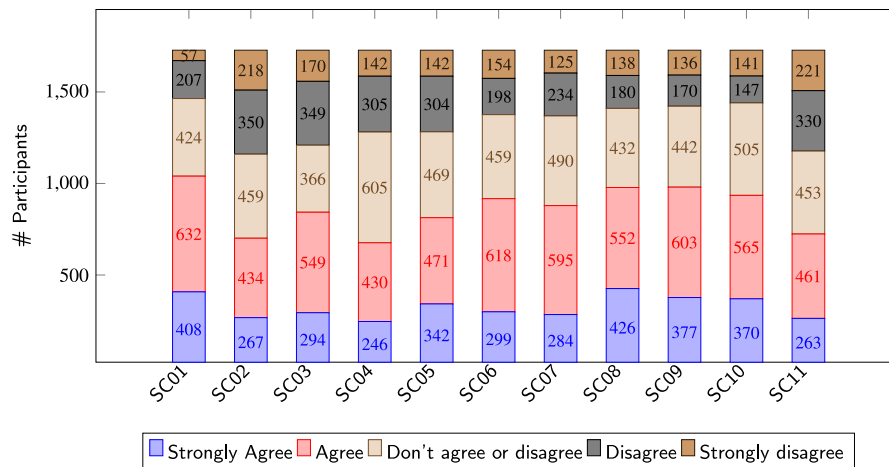


Fig. 10. Graphical scenarios contingency table.

Table 7
Descriptive statistics of participant responses for different technical applications.

	Mean	StDev	Min	Median	Max
(NC) New Code	3.3	1.1	1	3	5
(CR) Code Review	3.5	1.2	1	4	5
(CF) Code Refactoring	3.4	1.2	1	4	5
(TE) Team Efficiency	3.3	1.1	1	3	5

Similarly, the “Code Refactoring” program scores 3.4 on average, with a corresponding standard deviation of 1.2 and a median of 4. Finally, with an average score of 3.3, a steady standard deviation of 1.1, and a median of 3, the “Team Efficiency” application follows the pattern.

Fig. 11 shows a Technical applications Box plot statistics.

These results collectively indicate a moderate level of agreement across the board, with the “Code Review” and “Code Refactoring” applications exhibiting slightly higher mean scores and medians, potentially suggesting a stronger positive perception among participants compared to the “New Code” and “Team Efficiency” applications. The consistent standard deviations imply relatively stable responses within each application. This analysis emphasizes the importance of considering these nuances while interpreting the participants’ perspectives on various technical applications.

Table 8 provides a summary of the average evaluations of the technical applications in each of the 11 scenarios analyzed. Each scenario reflects the averages assigned to the four applications: “New Code”, “Code Review”, “Code Refactoring”, and “Team Efficiency”. The averages were calculated based on participants’ responses, where

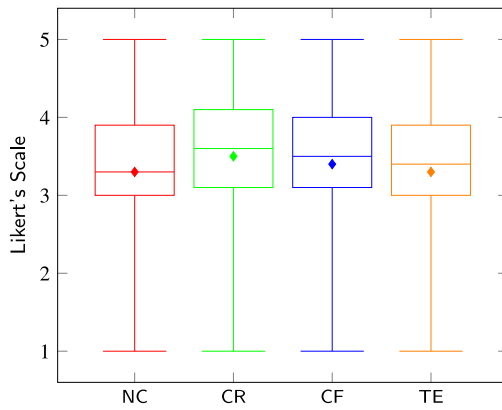


Fig. 11. Technical applications box plot.

Table 8

Average evaluations of technical applications across analyzed scenarios.

Scenario	New Code	Code Review	Code Refactoring	Team Efficiency
(SC01) Descriptive names	3.4	4.1	3.9	3.5
(SC02) Abbreviated names	3.2	3.1	3.1	3.0
(SC03) Bad naming	3.1	3.3	3.3	3.1
(SC04) Recent renames	3.0	3.2	3.3	3.1
(SC05) New update	3.2	3.5	3.5	3.3
(SC06) Showing modifiers	3.3	3.6	3.5	3.3
(SC07) Showing annotation	3.3	3.5	3.4	3.3
(SC08) Risky code	3.5	3.5	3.5	3.5
(SC09) Problems detected	3.5	3.6	3.6	3.4
(SC10) Visual reminders	3.5	3.6	3.5	3.4
(SC11) Person updated	2.9	3.2	3.0	3.4

they indicated their perceptions of the different applications in each scenario.

The analysis reveals variations in the averages in scenarios and applications. For example, in Scenario 1, the “New Code” application received an average of 3.4, while “Code Review”, “Code Refactoring”, and “Team Efficiency” received averages of 4.1, 3.9, and 3.5, respectively. However, in Scenario 11, the averages were generally lower, with “New Code” averaging 2.9, “Code Review” at 3.2, “Code Refactoring” at 3.0, and “Team Efficiency” at 3.4.

The Fig. 12 shows a technical applications across analyzed scenarios.

To better understand individual perceptions and the diversity of opinions in each scenario, and with the aim of analyzing the most relevant features in the intention to use code augmentation, the study was conducted through SEM.

A.2. Structural equation modeling - SEM

Structural Equation Modeling (SEM) is an advanced statistical approach used to investigate complex interactions between observed and unobserved variables within a model (English, 2006). It is frequently used in research domains such as social sciences, psychology, education, and economics to test and validate complex theories or hypotheses (English, 2006).

SEM permits the evaluation of both direct and indirect causal relationships between variables and the fit of the model to the observed data. It is useful for comprehending the interrelationships between variables, identifying latent factors (unobserved) that can influence observed variables, and evaluating the magnitude and direction of their associations (Kaplan, 2009).

Based on the UTAUT 2 model presented in Section 2.2, the analysis was constructed for each of the 11 scenarios. The gender variable has been removed from the model because the research data show that

89.8% of the participants are male. In other words, there is not enough data to apply the Gender moderator. The statistical rationale behind not using the Gender moderator is based on statistical power and sample size concepts. For a moderator variable to be properly examined and have a meaningful impact on the model, the sample size for each moderator category must be sufficiently broad and diverse (Preacher et al., 2007).

The model evaluation was carried out by examining Cronbach’s Alpha (CA) for internal consistency, followed by Composite Reliability (CR) as a measure of internal consistency. Convergent validity was assessed through Average Variance Extracted (AVE), and discriminant validity was estimated using the cross-loadings criterion (Henseler et al., 2009).

This approach facilitated an examination of individual reliability, delving into the correlations among the questionnaire items. The extent of their correlation reflects the explanatory strength of a particular variable. Expanding on this, it is worth noting that a higher correlation signifies a more robust relationship between variables, indicating that changes in one variable are more likely to result in corresponding changes in another. This insight provides a deeper understanding of how interconnected the variables within the model are and underscores the importance of considering the interplay between them when interpreting the results. Furthermore, a strong correlation suggests that the measured constructs are consistent and aligned, enhancing the overall reliability and validity of the analysis.

According to Table 9, it is evident that the AVE tests with values ≥ 0.5 (Fornell and Larcker, 1981) and Cronbach’s Alpha ≥ 0.6 (Hill and Hill, 2002) yielded values within the appropriate standards. Similar to Cronbach’s Alpha, Composite Reliability is a measure of reliability. This coefficient also achieved results that exceeded the acceptable threshold of ≥ 0.7 (Chin et al., 1998) for all constructs.

The consistency and high reliability observed in the CR and AVE values across various scenarios strengthen the confidence in the study’s measurements. The consistently strong reliability measures across constructs and scenarios indicate that the survey items measure the intended constructs with high consistency and accuracy.

Discriminant validity is a pivotal concept in the analysis of constructs within research. By adhering to the Fornell and Larcker (1981), a common standard for discriminant validity, the research ensures that the measures employed to gauge various constructs are indeed separate and not highly correlated. This reinforces the reliability and validity of the interpretations drawn from the conducted analyses.

In addition, the discriminant validity test assesses indicators to verify whether they differ from each other. Both the estimated correlation matrix and the square root of the average extracted variance from constructs are evaluated. As mentioned, the discriminant validity data obtained in this study adhere to Fornell and Larcker (1981), in which the value in each column is greater than those below it, representing the correlation between the constructs (Hair, 2014; Ringle et al., 2014).

Table 10 presents the Estimated Correlation Matrix for Scenario 01. Within this context, discriminant validity is assessed through two primary approaches: the estimated correlation matrix and the square root of the average variance extracted (AVE) from constructs.

In the case of the Estimated Correlation Matrix, correlation coefficients among constructs are examined. Discriminant validity is confirmed when the correlation coefficients among constructs are less than 1, indicating that they are not perfectly correlated. This suggests that the measures are distinct enough to represent different characteristics. Similarly, for the AVE, discriminant validity is met when the square root of the AVE for each construct is greater than the correlations between that construct and the others. This means that the shared variance among constructs is smaller than the variance captured by each construct individually, validating their distinctiveness. By verifying whether the data meet the Fornell and Larcker (1981) criterion, a common standard for discriminant validity, the research ensures that the measures used to assess different constructs are genuinely separate

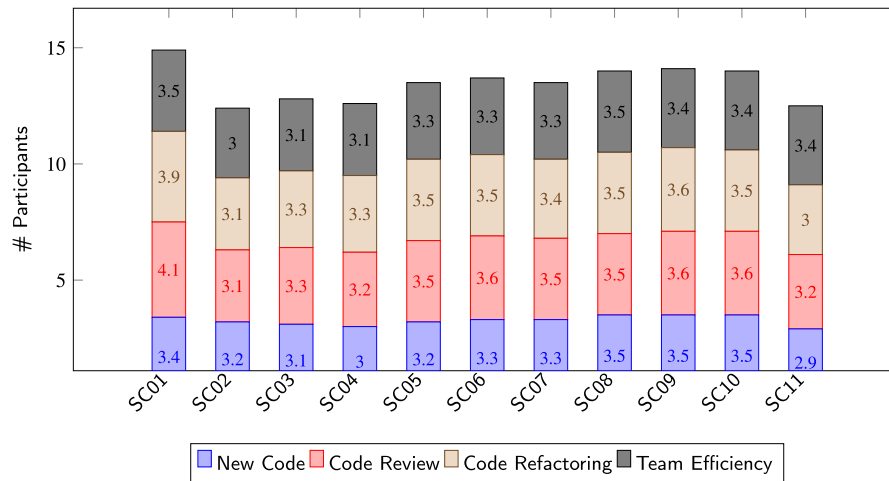


Fig. 12. Graphical scenarios contingency table.

Table 9
Reliability measures for various constructs across scenarios.

Scenario		Performance Expectancy (CT1)	Effort Expectancy (CT2)	Social Influence (CT3)	Facilitating Conditions (CT4)	Hedonic Motivation (CT5)	Learning Value (CT6)	Habit (CT7)
(SC01) Descriptive names	CA	0.801	0.692	0.683	0.769	0.888	0.727	0.570
	CR	0.909	0.866	0.860	0.896	0.946	0.869	0.823
	AVE	0.834	0.764	0.755	0.811	0.898	0.770	0.699
(SC02) Abbreviated names	CA	0.906	N/D	0.819	0.893	0.932	0.885	0.695
	CR	0.942	N/D	0.917	0.949	0.967	0.946	0.866
	AVE	0.843	N/D	0.846	0.904	0.936	0.897	0.764
(SC03) Bad naming	CA	0.868	0.817	0.863	0.941	0.875	0.928	0.842
	CR	0.938	0.916	0.936	0.971	0.941	0.965	0.927
	AVE	0.883	0.845	0.880	0.944	0.888	0.933	0.864
(SC04) Recent renames	CA	0.856	0.862	0.835	0.891	0.937	0.952	0.895
	CR	0.931	0.935	0.923	0.948	0.970	0.976	0.950
	AVE	0.872	0.879	0.858	0.901	0.941	0.954	0.905
(SC05) New update	CA	0.936	0.893	0.903	0.914	0.939	0.969	0.921
	CR	0.969	0.949	0.953	0.959	0.970	0.985	0.962
	AVE	0.940	0.903	0.911	0.920	0.942	0.970	0.927
(SC06) Showing modifiers	CA	0.909	0.943	0.868	0.956	0.940	0.905	0.831
	CR	0.956	0.964	0.938	0.978	0.971	0.954	0.922
	AVE	0.916	0.898	0.883	0.957	0.944	0.913	0.855
(SC07) Showing annotation	CA	N/D	0.940	0.880	0.957	0.911	0.938	0.888
	CR	N/D	0.962	0.943	0.979	0.958	0.970	0.946
	AVE	N/D	0.894	0.893	0.959	0.919	0.942	0.898
(SC08) Risky code	CA	0.943	0.951	0.917	0.922	0.943	0.905	0.916
	CR	0.964	0.968	0.960	0.962	0.972	0.954	0.960
	AVE	0.898	0.910	0.923	0.927	0.946	0.913	0.922
(SC09) Problems detected	CA	0.963	0.963	0.904	0.909	0.940	0.914	0.891
	CR	0.982	0.976	0.954	0.956	0.971	0.958	0.948
	AVE	0.964	0.932	0.913	0.916	0.943	0.920	0.901
(SC10) Visual reminders	CA	0.928	0.917	0.909	0.913	0.933	0.917	0.809
	CR	0.965	0.960	0.956	0.958	0.967	0.960	0.912
	AVE	0.933	0.923	0.917	0.919	0.937	0.923	0.839
(SC11) Person updated	CA	0.929	0.936	0.891	0.947	0.929	0.959	0.857
	CR	0.955	0.959	0.948	0.974	0.966	0.980	0.933
	AVE	0.876	0.887	0.902	0.950	0.934	0.960	0.875

and not highly correlated. This reinforces the reliability and validity of interpretations drawn from the conducted analyses.

For each of the 11 scenarios, we conducted a discriminant validity analysis. To meet the criteria suggested by [Fornell and Larcker \(1981\)](#), which requires that each value in a column be greater than the values below it, we performed cross-load analyses to eliminate predictor variables if necessary.

An analysis was performed using the bootstrapping technique to investigate the values of the Student's *t*-test ([STUDENT, 1908; Field,](#)

[2013](#)) for each of the 11 scenarios. This investigation sought to determine the correlations between the constructs reflecting the hypotheses explored in this study. Bootstrapping is a statistical approach that involves creating several samples from the original data to perform a more accurate and reliable analysis ([Hair, 2014](#)). According to [Hair \(2014\)](#), this adaptable method helps to calculate crucial statistics, build confidence intervals, and rigorously test hypotheses without relying on tight assumptions about data distribution. In this paper, reference values for hypotheses accepted using the Student's *t*-test were *t*-value \geq

Table 10
Scenario 1: Discriminant validity metrics.

	BI ^a	CT2	CT4	CT7	CT5	CT6	CT1	CT3
(BI ^a) Behavioral Intention	1.000							
(CT2) Effort Expectancy	0.484	0.874						
(CT4) Facilitating Conditions	0.334	0.612	0.901					
(CT7) Habit	0.334	0.505	0.266	0.836				
(CT5) Hedonic Motivation	0.334	0.580	0.538	0.691	0.948			
(CT6) Learning Value	0.270	0.477	0.492	0.503	0.574	0.877		
(CT1) Performance Expectancy	0.402	0.812	0.561	0.582	0.638	0.476	0.913	
(CT3) Social Influence	0.370	0.623	0.498	0.547	0.629	0.494	0.595	0.869

^a Behavioral Intention.

Table 11
Results of bootstrapping analysis for hypotheses across scenarios.

	Hypothesis	Confirmed by scenario	t-values	p-values (Position)	R ²
Performance Expectancy → BI ^a	H ₁	(SC11) Person updated	3.467	0.001 (01)	0.652
		(SC01) Descriptive names	2.705	0.007 (04)	0.251
		(SC02) Abbreviated names	2.442	0.015 (05)	0.547
		(SC06) Showing modifiers	2.010	0.045 (11)	0.600
Effort Expectancy → BI ^a	H ₂	(SC07) Showing annotation	2.090	0.037 (09)	0.546
		(SC09) Problems detected	2.052	0.040 (10)	0.584
		(SC10) Visual reminders	2.707	0.007 (03)	0.612
		(SC11) Person updated	2.111	0.035 (08)	0.652
Social Influence → BI ^a	H ₃	(SC05) New update	2.009	0.045 (12)	0.622
Facilitating Conditions → BI ^a	H ₄	Not confirmed			
Hedonic Motivation → BI ^a	H ₅	Not confirmed			
Learning Value → BI ^a	H ₆	(SC03) Bad naming	2.723	0.006 (02)	0.738
		(SC04) Recent renames	2.318	0.020 (07)	0.616
Habit → BI ^a	H ₇	(SC06) Showing modifiers	2.430	0.015 (06)	0.600
Moderators					
Age × Facilitating Conditions → BI ^a	H ₄ A ₁	Not confirmed			
Age × Learning Value → BI ^a	H ₅ A ₁	Not confirmed			
Age × Hedonic Motivation → BI ^a	H ₆ A ₁	Not confirmed			
Age × Habit → BI ^a	H ₇ A ₁	Not confirmed			
Experience × Facilitating Conditions → BI ^a	H ₄ E ₁	Not confirmed			
Experience × Learning Value → BI ^a	H ₅ E ₁	Not confirmed			
Experience × Hedonic Motivation → BI ^a	H ₆ E ₁	Not confirmed			
Experience × Habit → BI ^a	H ₇ E ₁	Not confirmed			
				R ² Mean	0.585
				R ² Median	0.606
				R ² StDev	0.117

^a Behavioral Intention.

1.96 and p -value < 0.05, respectively (Hair, 2014). The bootstrapping analysis is presented in Table 11, which highlights the scenarios that were able to confirm the defined hypotheses.

References

- Abebe, S.L., Haiduc, S., Tonella, P., Marcus, A., 2009. Lexicon bad smells in software. In: 2009 16th Working Conference on Reverse Engineering. IEEE, pp. 95–99.
- Allamanis, M., Barr, E.T., Bird, C., Sutton, C., 2015. Suggesting accurate method and class names. In: Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. pp. 38–49.
- Allamanis, M., Barr, E.T., Devanbu, P., Sutton, C., 2018. A survey of machine learning for big code and naturalness. ACM Comput. Surv. 51 (4).
- Arnaoudova, V., Di Penta, M., Antoniol, G., 2016. Linguistic antipatterns: What they are and how developers perceive them. Empir. Softw. Eng. 21 (1).
- Arnaoudova, V., Eshkevari, L.M., Penta, M.D., Oliveto, R., Antoniol, G., Guéhéneuc, Y.-G., 2014. REPENT: Analyzing the nature of identifier renamings. IEEE Trans. Softw. Eng. 40 (5), 502–532. <http://dx.doi.org/10.1109/TSE.2014.2312942>.
- Arnaoudova, V., Haiduc, S., Marcus, A., Antoniol, G., 2015. The use of text retrieval and natural language processing in software engineering. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering. ICSE, IEEE Computer Society, pp. 949–950.
- Atkins, D.L., Ball, T., Graves, T.L., Mockus, A., 2002. Using version control data to evaluate the impact of software tools: A case study of the version editor. IEEE Trans. Softw. Eng. 28 (7).
- Beelders, T.R., du Plessis, J.-P.L., 2016. Syntax highlighting as an influencing factor when reading and comprehending source code. J. Eye Move. Res. 9 (1).
- Beniamini, G., Gingichashvili, S., Orbach, A.K., Feitelson, D.G., 2017. Meaningful identifier names: The case of single-letter variables. In: 2017 IEEE/ACM 25th International Conference on Program Comprehension. ICPC, IEEE, pp. 45–54.
- Binkley, D., Davis, M., Lawrie, D., Morrell, C., 2009. To camelcase or under_score. In: 2009 IEEE 17th International Conference on Program Comprehension. IEEE, pp. 158–167.
- Bland, J.M., Altman, D.G., 1997. Statistics notes: Cronbach's alpha. BMJ 314 (7080), 572. <http://dx.doi.org/10.1136/bmj.314.7080.572>, URL: <https://www.bmj.com/content/314/7080/572>. arXiv:<https://www.bmj.com/content/314/7080/572.full.pdf>.
- Brown, S., 2023. The C4 Model for Visualising Software Architecture. Leanpub, URL: <https://leanpub.com/visualising-software-architecture>. (Accessed 29 May 2024).
- Buckner, J., Buchta, J., Petrenko, M., Rajlich, V., 2005. Jripples: A tool for program comprehension during incremental change. In: 13th International Workshop on Program Comprehension. IWPC'05, IEEE, pp. 149–152.
- Calcagno, C., Distefano, D., 2011. Infer: An automatic program verifier for memory safety of c programs. In: NASA Formal Methods Symposium. Springer, pp. 459–465.
- Chin, W.W., et al., 1998. The partial least squares approach to structural equation modeling. Mod. Methods Bus. Res. 295 (2), 295–336.
- Corazza, A., Di Martino, S., Maggio, V., 2012. LINSSEN: An efficient approach to split identifiers and expand abbreviations. In: 2012 28th IEEE International Conference on Software Maintenance. ICSM, IEEE, pp. 233–242.
- Corral, L., Georgiev, A.B., Janes, A., Kofler, S., 2015. Energy-aware performance evaluation of android custom kernels. In: 2015 IEEE/ACM 4th International Workshop on Green and Sustainable Software. pp. 1–7. <http://dx.doi.org/10.1109/GREENS.2015.8>.
- Cowlshaw, M.F., 1987. LEXX—A programmable structured editor. IBM J. Res. Dev. 31 (1), 73–80. <http://dx.doi.org/10.1147/rd.311.0073>.

- Cresswell, J.W., 2012. Planning, conducting, and evaluating quantitative and qualitative research. *Educ. Res.* 10, 1–12.
- Deissenboeck, F., Pizka, M., 2006. Concise and consistent naming. *Softw. Qual. J.* 14 (3).
- Edwards, J., 2005. Subtext: uncovering the simplicity of programming. In: *Proceedings of the 20th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications*. pp. 505–518.
- English, F., 2006. *Encyclopedia of Educational Leadership and Administration*. SAGE Publications, Inc., Thousand Oaks, pp. 970–971. <http://dx.doi.org/10.4135/9781412939584>, (Chapter Structural Equation Modeling).
- Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., Shou, L., Qin, B., Liu, T., Jiang, D., et al., 2020. Codebert: A pre-trained model for programming and natural languages. *arXiv preprint arXiv:2002.08155*.
- Field, A., 2013. *Discovering Statistics using IBM SPSS Statistics*. Sage.
- Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18 (1), 39–50.
- Fowler, M., 2009. TwoHardThings. <https://martinfowler.com/bliki/TwoHardThings.html>.
- Fowler, M., 2018. *Refactoring: Improving the Design of Existing Code*. Addison-Wesley Professional.
- Guerra, E., 2016. Design patterns for annotation-based APIs. In: *Proceedings of the 11th Latin-American Conference on Pattern Languages of Programming*. SugarLoafPLOP '16, The Hillside Group, USA, pp. 1–14.
- Gueroirou, L., 2013. Normalizing source code vocabulary to support program comprehension and software quality. In: *2013 35th International Conference on Software Engineering*. ICSE, IEEE, pp. 1385–1388.
- Hackett, B., Aiken, A., 2006. How is aliasing used in systems software? In: *Proceedings of the 14th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. pp. 69–80.
- Hair, J., 2014. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, URL: <https://books.google.com.br/books?id=IFiaryXE1PoC>.
- Hasan, S., King, Z., Hafiz, M., Sayagh, M., Adams, B., Hindle, A., 2016. Energy profiles of java collections classes. In: *Proceedings of the 38th International Conference on Software Engineering*. pp. 225–236.
- Henseler, J., Ringle, C.M., Sinkovics, R.R., 2009. *Advances in International Marketing*, vol. 20, Emerald Group Publishing Limited, pp. 277–319. [http://dx.doi.org/10.1108/S1474-7979\(2009\)0000020014](http://dx.doi.org/10.1108/S1474-7979(2009)0000020014), (Chapter the Use of Partial Least Squares Path Modeling in International Marketing).
- Hill, M., Hill, A., 2002. Investigação por Questionário. Sílabo, URL: <https://books.google.com.br/books?id=6neOwaAACAAJ>.
- Hofmeister, J., Siegmund, J., Holt, D.V., 2017. Shorter identifier names take longer to comprehend. In: *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering*. SANER, IEEE, pp. 417–443.
- Hora, A., Silva, D., Valente, M.T., Robbes, R., 2018. Assessing the threat of untracked changes in software evolution. In: *Proceedings of the 40th International Conference on Software Engineering*. pp. 1102–1113.
- Høst, E.W., Østfold, B.M., 2009. Debugging method names. In: *European Conference on Object-Oriented Programming*. Springer, pp. 294–317.
- Hovemeyer, D., Pugh, W., 2004. Finding bugs is easy. *ACM SIGPLAN Not.* 39 (12).
- Janes, A., Mairegger, M., Russo, B., 2018. *code.call.lens: raising the developer awareness of critical code*. In: *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. pp. 876–879.
- JetBrains, 2023. Inlay hints | IntelliJ platform plugin SDK. <https://plugins.jetbrains.com/docs/intellij/inlay-hints.html>.
- Jun, J., Park, H., Cho, I., 2019. Study on initial adoption of advanced driver assistance system: integrated model of PMT and UTAUT 2. *Total Qual. Manage. Bus. Excell.* 30 (sup1), S83–S97. <http://dx.doi.org/10.1080/14783363.2019.1665820>.
- Kaplan, D., 2009. *Structural Equation Modeling: Foundations and Extensions*, second ed. SAGE Los Angeles, Los Angeles.
- Lawrie, D., Feild, H., Binkley, D., 2007. Extracting meaning from abbreviated identifiers. In: *Seventh IEEE International Working Conference on Source Code Analysis and Manipulation*. SCAM 2007, IEEE, pp. 213–222.
- Lawrie, D., Morrell, C., Feild, H., Binkley, D., 2006. What's in a name? A study of identifiers. In: *14th IEEE International Conference on Program Comprehension*. ICPC'06, IEEE, pp. 3–12.
- Lima, P., Melegati, J., Gomes, E., Pereira, N.S., Guerra, E., Meirelles, P., 2022. CADV: A software visualization approach for code annotations distribution. *Inf. Softw. Technol.* <http://dx.doi.org/10.1016/j.infsof.2022.107089>.
- Liu, K., Kim, D., Bissyandé, T.F., Kim, T., Kim, K., Koyuncu, A., Kim, S., Le Traon, Y., 2019. Learning to spot and refactor inconsistent method names. In: *2019 IEEE/ACM 41st International Conference on Software Engineering*. ICSE, IEEE, pp. 1–12.
- Martin, R.C., Coplien, J.O., 2009. *Clean Code: a Handbook of Agile Software Craftsmanship*. Prentice Hall, Upper Saddle River, NJ [etc.], URL: https://www.amazon.de/gp/product/0132350882/ref=oh_details_o0o_s00_i00.
- McConnell, S., 2004. *Code Complete: A Practical Handbook of Software Construction*, second ed. In: *Best Practices for Developers*, Microsoft Press, Redmond, WA, URL: <https://www.safaribooksonline.com/library/view/code-complete-second/0735619670/>.
- Meyer, A.N., Fritz, T., Murphy, G.C., Zimmermann, T., 2014. Software developers' perceptions of productivity. In: *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. pp. 19–29.
- Microsoft, 2022. VS code API. <https://code.visualstudio.com/api/references/vscode-api>.
- Microsoft, 2023. Visual studio code. <https://code.visualstudio.com>.
- Murphy, G.C., 2019. Beyond integrated development environments: Adding context to software development. In: *2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results*. ICSE-NIER, pp. 73–76. <http://dx.doi.org/10.1109/ICSE-NIER.2019.00027>.
- Murphy-Hill, E., Murphy, G.C., 2014. Recommendation delivery. In: *Recommendation Systems in Software Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 223–242. http://dx.doi.org/10.1007/978-3-642-45135-5_9, (Chapter Recommendation Systems in Software Engineering).
- Petrescu, C.C., Smith, S., Giavrimis, R., Dash, S.K., 2023. Do names echo semantics? A large-scale study of identifiers used in C++'s named casts. *J. Syst. Softw.* 202, 111693. <http://dx.doi.org/10.1016/j.jss.2023.111693>, URL: <https://www.sciencedirect.com/science/article/pii/S0164121223000882>.
- Pradel, M., Sen, K., 2018. Deepbugs: A learning approach to name-based bug detection. In: *Proceedings of the ACM on Programming Languages*. Vol. 2, ACM New York, NY, USA.
- Preacher, K.J., Rucker, D.D., Hayes, A.F., 2007. Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivar. Behav. Res.* 42 (1), 185–227. <http://dx.doi.org/10.1080/00273170701341316>, PMID: 26821081.
- Ringle, C., Silva, D., Bido, D., 2014. Modelagem de equações estruturais com utilização do smartpls. *REMark: Rev. Bras. Mark.* 13, <http://dx.doi.org/10.5585/bjm.v13i2.2717>.
- Robbes, R., Lungu, M., Röthlisberger, D., 2012. How do developers react to API deprecation? The case of a smalltalk ecosystem. In: *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*. pp. 1–11.
- Sanchez, H., Robbes, R., Gonzalez, V.M., 2015. An empirical study of work fragmentation in software evolution tasks. In: *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering*. SANER, IEEE, pp. 251–260.
- Savant, A.A., Robbes, R., Bacchelli, A., 2016. On the reaction to deprecation of 25,357 clients of 4+ 1 popular java APIs. In: *2016 IEEE International Conference on Software Maintenance and Evolution*. ICSME, IEEE, pp. 400–410.
- Schankin, A., Berger, A., Holt, D.V., Hofmeister, J.C., Riedel, T., Beigl, M., 2018. Descriptive compound identifier names improve source code comprehension. In: *2018 IEEE/ACM 26th International Conference on Program Comprehension*. ICPC, IEEE, pp. 31–3109.
- Schwartz, M.D., Delisle, N.M., Begwani, V.S., 1984. Incremental compilation in magpie. In: *Proceedings of the 1984 SIGPLAN Symposium on Compiler Construction*. SIGPLAN '84, Association for Computing Machinery, New York, NY, USA, pp. 122–131. <http://dx.doi.org/10.1145/502874.502887>.
- Sharafi, Z., Soh, Z., Guéhéneuc, Y.-G., Antoniol, G., 2012. Women and men—different but equal: On the impact of identifier style on source code reading. In: *2012 20th IEEE International Conference on Program Comprehension*. ICPC, IEEE, pp. 27–36.
- Slivski, J., Zimmermann, T., Zeller, A., 2005. HATARI: raising risk awareness. In: *ESEC/SIGSOFT FSE*. pp. 107–110.
- STUDENT, 1908. The probable error of a mean. *Biometrika* 6 (1), 1–25. <http://dx.doi.org/10.1093/biomet/6.1.1>, arXiv:https://academic.oup.com/biomet/article-pdf/6/1/1/49729532/biomet_6_1_1.pdf.
- Sulir, M., Bačiková, M., Chodarev, S., Porubán, J., 2018. Visual augmentation of source code editors: A systematic mapping study. *J. Vis. Lang. Comput.* 49, 46–59. <http://dx.doi.org/10.1016/j.jvlc.2018.10.001>, URL: <https://www.sciencedirect.com/science/article/pii/S1045926X18301861>.
- Thomas, J., Harden, A., 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med. Res. Methodol.* 8 (1), 1–10.
- Ungar, D., Lieberman, H., Fry, C., 1997. Debugging and the experience of immediacy. *Commun. ACM* 40 (4), 38–43. <http://dx.doi.org/10.1145/248448.248457>.
- Vasilescu, B., Casalnuovo, C., Devanbu, P., 2017. Recovering clear, natural identifiers from obfuscated JS names. In: *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*. pp. 683–693.
- Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: Toward a unified view. *MIS Q.* 27 (3), 425–478, URL: <http://www.jstor.org/stable/30036540>.
- Venkatesh, V., Thong, J.Y.L., Xu, X., 2012. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Q.* 36 (1), 157–178, URL: <http://www.jstor.org/stable/41410412>.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M., Regnell, B., Wesslén, A., 2012. *Experimentation in Software Engineering*. In: *Computer Science*, Springer Berlin Heidelberg, URL: https://books.google.com.br/books?id=QPVsM1_U8nKc.



Eduardo Guerra is currently a researcher at the Free University of Bozen-Bolzano, worked as a researcher at the National Institute for Space Research in Brazil, and as a teacher at the Aeronautics Institute of Technology. His research interests include agile methods, software patterns, framework development, software analytics, and dynamic architectures. Guerra received a Ph.D. in computer engineering from the Aeronautics Institute of Technology and has practical experience in architecture and framework design.



André A.S. Ivo holds a Ph.D. in Space Engineering and Technologies (2023) and an M.Sc. in the same field (2013) from the National Institute for Space Research (INPE). He also earned an MBA in Project Management from Fundação Getúlio Vargas in 2010 and a degree in Computer Engineering from Vale do Paraíba University in 2006. He has extensive experience in the private sector, working as a Software Engineer and later as a Systems Manager, Coordinator, and Systems Engineer. His expertise encompasses system engineering, automation, databases, quality and maturity models, and service models. Since 2016, André has served as a full technology specialist federal civil servant at CEMADEN, where he is responsible for managing the observational network and conducting research and development focused on forecasting and monitoring natural disasters.



Fernando O. Pereira is currently a Information Technologist at the National Centre for Monitoring and Early Warning of Natural Disasters (CEMADEN). Fernando obtained a Bachelor's degree in Computer Science from the Federal University of Ouro Preto (UFOP) in 2008, followed by a Master's degree in Applied Computing from the National Institute of Space Research (INPE) in 2013. At present, he is studying for a doctoral degree in Applied Computing at the National Institute of Space Research (INPE). Fernando has expertise in the domain of software engineering, particularly in Java development, mobile device development, database management, and systems architecture.



Romain Robbes is a Senior Scientist (Directeur de Recherche) as the CNRS, working at the LaBRI and hosted by the University of Bordeaux, since February 2023. Before that, he was: an Associate Professor at the Free University of Bozen-Bolzano, (2017-2023); an Assistant, then Associate Professor at the University of Chile (2019s Computer Science Department (2010-2013, 2017); a Ph.D. student, then post-doc, at the University of Lugano (2004-2013, 2009). His research interests include Empirical Software Engineering, Mining Software Repositories, Software Maintenance and Evolution, and the intersection of Machine Learning with Programming Languages and Software Engineering.



Andrea Janes is a senior lecturer and researcher at the FHV Vorarlberg University of Applied Sciences in Dornbirn, Austria. Previously, he was a researcher at the Free University of Bolzano at the Faculty of Computer Science and a member of the Smart Data Factory, a technology transfer group within the faculty, where he acquired and led several research projects and collaborations with industry. He received the master's degree in computer science from the Technical University of Vienna, Austria and the doctorate in computer science (with distinction) from the University of Klagenfurt (Austria). He is particularly interested in Lean and Agile approaches to software engineering, Value-based Software Engineering, Empirical Software Engineering, and Software Testing.



Fábio F. Silveira is a Full Professor at the Science and Technology Institute of the Federal University of São Paulo (UNIFESP). His research interests encompass Software Testing, Experimental Software Engineering, Agile Methods, Software Metrics, and Software Architecture. He earned his M.Sc. from the Federal University of Rio Grande do Sul in 2001 and his Ph.D. in Computer Science from the Technological Institute of Aeronautics (ITA) in 2007. In 2009, he was a visiting researcher at the Technische Universität Berlin.