FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

# An Investigation Into User Expectations for Formal Verification

**Mariana Catarina Pereira Soares**

Mestrado em Engenharia Informática e Computação

Supervisor: Alexandra Sofia Ferreira Mendes

Second Supervisor: João Fernando Peixoto Ferreira

August 21, 2023

# An Investigation Into User Expectations for Formal Verification

**Mariana Catarina Pereira Soares**

Mestrado em Engenharia Informática e Computação

August 21, 2023

# Resumo

A verificação formal desempenha cada vez mais um papel essencial em garantir o comportamento correto e seguro dos sistemas de software modernos. Atualmente, existem várias descrições que tentam explicar o que é a verificação formal. Por exemplo, a verificação formal pode ser descrita como o processo de verificar matematicamente se um sistema satisfaz a especificação formal do seu comportamento. No entanto, algumas dessas descrições podem ser demasiado técnicas ou extensas para que alguém não familiarizado com a área entenda completamente o conceito.

Este trabalho tem como objetivo explorar a literatura atual na área de verificação formal e, através de estudos com utilizadores, encontrar uma melhor forma de comunicar a verificação formal a pessoas que desconhecem o conceito.

Rever a literatura existente é um primeiro passo essencial para entender melhor que tipo de descrições são usadas atualmente para verificação formal. Realizamos uma pesquisa sistemática para selecionar e recolher descrições já existentes, explorando diferentes APIs de pesquisa. Em seguida, estudamos diretrizes para uma melhor comunicação e identificamos sugestões que podem ser utilizadas na criação de novas descrições mais adequadas. Por fim, foi realizado um estudo com utilizadores por meio de questionários. Neste estudo, uma seleção de descrições foi apresentada aos participantes de modo a avaliar quais as descrições mais fáceis de entender. Este tipo de estudo com utilizadores deverá fornecer conhecimento sobre como comunicar melhor a verificação formal a pessoas que não conhecem o conceito.

Com esta investigação, percebemos que os utilizadores entendem melhor as descrições de verificação formal associadas à segurança e às propriedades do código. Como muitos participantes sugeriram, adicionar exemplos práticos pode ser uma mais valia para as descrições. Além disso, os resultados obtidos podem ajudar a comunicar outros conceitos técnicos aos utilizadores em geral.

**Palavras-chave:** Verificação Formal; Estudo com utilizadores; Questionário

# Abstract

Formal verification plays an increasingly essential role in guaranteeing modern software systems' correct and safe behavior. There are already many descriptions that attempt to explain what formal verification is. For example, formal verification can be described as the process of mathematically verifying that a system satisfies the formal specification of its behavior. However, some of these descriptions may be too technical or extensive for someone unfamiliar with this subject area to fully understand the concept.

This work aims to explore the current literature on the formal verification area and, through user studies, find a better way to communicate formal verification to non-expert users.

Reviewing the existing literature is an essential first step to understand better what kind of descriptions are currently used for formal verification. We performed a systematic search to select and collect existing descriptions by exploring different search APIs. Following that, we studied guidelines for better communication and identified suggestions that can be used to create new and more adequate descriptions. Finally, a user study using surveys was carried out. In this study, a selection of descriptions was presented to participants to assess which descriptions were easier to understand. This type of user study should give us insights into how to better communicate formal verification to non-expert users.

Regarding this investigation, we have seen that non-experts best understand formal verification descriptions associated with security and code properties. As many participants suggested, adding practical examples could be an asset to the descriptions. Moreover, the results obtained may help communicate other technical concepts to users in general.

**Keywords:** Formal Verification; User study; Survey

# Acknowledgments

I want to thank my parents, grandparents, and sister for their unconditional support. They always believed in me, motivated me, and were always there to support and love me.

I would also like to thank Prof. Alexandra Sofia Ferreira Mendes, Prof. João Fernando Peixoto Ferreira, and Carolina Carreira for all the help they gave me in carrying out this project. They always had alternative solutions when I did not see any. It would not have been possible otherwise.

Finally, I also thank all my friends who always encouraged me to continue and not give up.

Thank you.

Mariana Catarina Pereira Soares

*“Success is stumbling from failure to failure with no loss of enthusiasm.”*

Winston Churchill

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| JSON | JavaScript Object Notation |
| API | Application Programming Interface |
| XML | Extensible Markup Language |
| HTML | Hypertext Markup Language |
| CSV | Comma-separated values |
| RSS | Really Simple Syndication |
| REST | Representational State Transfer |
| E2EE | End-to-end Encryption |
| DP | Differential Privacy |
| PDF | Portable Document Format |
| FV | Formal Verification |
| IEEE | Institute of Electrical and Electronics Engineers |
| ACM | Association for Computing Machinery |

# Chapter 1

# Introduction

In this section, we present the context and motivation of our project. We also describe our main goals by presenting our research questions. We finish this section with a description of the structure of this document.
In section 1.1, we introduce the context and motivation of our study. In section 1.2, we present our main goals. In section 1.3, we describe the document's layout.

## 1.1 Context and Motivation

We are increasingly dependent on software that helps us simplify and automate processes making them more efficient. However, when dealing with software, it is essential to avoid any possibility of system failures occurring.

In non-safety-critical systems, software flaws are generally realized and later corrected through updates. In safety-critical systems, no error must occur. We must be aware of the software's reliability before installing productive systems. We can consider software reliable when it fulfills previously defined specifications. One typically primary method to check this is formal verification which plays an increasingly essential role in guaranteeing modern software systems' correct and reliable behavior [17].

Formal verification seeks to verify all possible program paths using different mathematical algorithms, depending on the software. The algorithm can conclude whether the program fulfills the given specification or not, either because it found a violation or the analysis went wrong [17].

There are several different descriptions for the concept of formal verification, but these can be too complex for users unfamiliar with this subject area. It is essential to ensure that users fully understand the concept of formal verification and know its guarantees and what formal verification does not guarantee to avoid exposing themselves to unnecessary risks.

Human-centered studies on better communicating the concept of formal verification, and the formal guarantees it provides, to non-expert users are lacking. Our goal with this work is to contribute towards filling this gap.

The results of these studies can potentially contribute to ensuring that non-experts fully comprehend formal verification and its benefits to prevent users from taking unnecessarily risky actions when using the software.

## 1.2 Objectives

This study's main objective is to properly inform the community about how to convey the concept of formal verification to non-expert users in a better way in order to improve the understanding of this concept.

Our research might help the general user understand other technical concepts in different subject areas by applying similar approaches to ours or even improving them.

### 1.2.1 Research Questions

We intend to answer the research questions below through this project.

**RQ1:** How is the concept of formal verification currently described?

**RQ2:** How can we better communicate the concept of formal verification?

**RQ3:** Which descriptions better communicate the concept of formal verification to people with no background in this domain?

## 1.3 Document Structure

Besides the introduction, where we already presented the context, the motivation, and our research objectives, this document has three more chapters.

In chapter 2, we describe the state of the art, where we make an overview of formal verification and its techniques. We also present and discuss some existing search APIs. We also discuss some background work examples and analyze approaches to related problems. Finally, we discuss surveys in user studies and qualitative data analysis. Here, we focus on thematic analysis.

In chapter 3, we present the problem statement and methodology used to perform our research study. This chapter also describes the analysis of the results. We also explain the limitations of our approach, threats to validity, and what we conclude from the study.

In chapter 4, we conclude this document with final considerations and propose future work.

# Chapter 2

# State of the Art

This chapter presents the current state of the art.

In section 2.1, we present an overview of the formal verification context and its techniques. In section 2.2, we introduce the core concept of the search API (Application Programming Interface) since we need to use a search API to collect formal verification descriptions. Then we present some existing APIs in subsection 2.2.1 and compare them in subsection 2.2.2. In section 2.3, we have some background descriptions of the formal verification concept, and we discuss them in subsection 2.3.1. The section 2.4 approaches the qualitative data analysis theme, focusing on thematic analysis. In section 2.5, we analyze two related problems to ours, where the intent was to communicate technical concepts better to non-expert users. Finally, the section 2.6 presents an overview of surveys as a method used in user studies.

## 2.1 Formal Verification

According to Per Bjesse, "Formal verification is the use of mathematical techniques to ensure that a design conforms to some precisely expressed notion of functional correctness. Concretely, assume that you have (1) a model of a design, (2) some description of the environment that the design is supposed to operate in, and (3) some properties that the design is intended to fulfill. Given this information, you may want to search for some input patterns that the environment could generate that will violate the properties." [10].

After reading the description above carefully, we can describe FV (Formal Verification) in another way. For example, we can say that formal verification exhaustively confirms a system's correctness using mathematical methods to guarantee that the behavior of a system matches its specifications.

Nowadays, plenty of descriptions try to explain the concept of formal verification.

According to Aijaz Fatima, formal verification catches bugs earlier since a design must undergo formal verification before being functionally validated through simulation and emulation [30]. It is to be noted that "Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and its solution. Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model." [27].

### 2.1.1 Formal Verification Techniques

The techniques used in formal verification are model checking, theorem proving, and equivalence checking.

In **model checking**, we represent the system as a finite state machine, indicating how the system goes from one state to another. The model checker verifies if the model satisfies the property, and if it does not, the model checker generates a counter-example. It is a state-based approach. The verification process is entirely automatic after we have the model and the property. The number of states has to be finite, which is a limitation. In **theorem proving**, we create a mathematical model to represent the system's properties the system has to fulfill. We use a theorem prover to ensure the system complies with the requirements. It is a proof-based approach. Theorem proving can manage systems with high complexity. It necessitates human interaction to complete the verification. In **equivalence checking**, we evaluate whether two designs are functionally equivalent during the design process by confirming that they are functionally equivalent and produce the same results when given the same inputs or that the two systems contain the same combinational logic between registers [30].

Formal verification is complete for each requirement since it can provide exhaustive coverage for a particular requirement. However, we must perform additional actions to state all possible specifications. Furthermore, formal verification outcomes rely on assumptions, and in complex systems, it can be significantly expensive, time-consuming, and complex [23].

## 2.2 Search APIs

Search APIs enable developers to add search functionality to applications or websites quickly. Additionally, a search API lets us perform different data queries and filter the outcomes. We used a search API to collect formal verification descriptions to be a replicable process in similar future studies.

### 2.2.1 Existing search APIs

There are several search APIs available, including the following:

- Wayback Machine [2]
- Google Search APIs, such as Custom Search JSON API [1] or SerpApi [4]
- Bing Web Search API [22]

- Web Search API [5]
- GeoRanker API [3]

Regarding the existing APIs found, we could separate them into three different variations.

Some APIs are free, such as Wayback Machine, which allows free access to all the available features. Other APIs allow a free or paid plan, such as Google Search APIs, Bing Web Search API, and Web Search API. In these cases, the main difference between the free and paid plan is that the first one usually limits the number of queries we can do per day or month. However, we do not see this as a limitation for this work because the number of queries allowed in free plans is at least one hundred per day, which is acceptable for us.

We also have paid APIs, such as the GeoRanker API, which offers a seven days free trial. We discarded all paid APIs.

We explored the Wayback Machine and the Custom Search JSON API more deeply because the first one allows access to content no longer available in the source, and the second because the Google search engine is the most used.

Regarding the Wayback Machine [2], the internet archive is a non-profit digital library with free access to the public [2]. It allows searching for different content and filtering it in various ways, such as title, creator, description, or date. The results can be returned in a user-friendly interface or a JSON (JavaScript Object Notation), XML (Extensible Markup Language), HTML (Hypertext Markup Language), CSV (Comma-separated values), or RSS (Really Simple Syndication) file.

On the other hand, the Custom Search JSON API [1] can display search results from the Programmable Search Engine. With the help of this API, we can make RESTful (Representational State Transfer) queries to obtain JSON-formatted web search or image search results [1]. It also allows filtering the results by different fields.

### 2.2.2 Discussion

We compared the Wayback Machine and the Custom Search JSON API by using both to search formal verification descriptions. We verified that both APIs have the query parameters that might be useful to this study, such as filtering by term, date, file type, or domain. However, as seen above, while the Custom Search JSON API returns the results only in a JSON file, the Wayback Machine returns the results either in a JSON file or in an XML, HTML, CSV, RSS file, or a user-friendly interface.

In addition, we compared the number of results obtained by both APIs by gathering the number of PDF (Portable Document Format) results for the term of formal verification and each of the different terms equivalent to formal verification, such as *formally verified*, *formal methods*, *software proofs*, and *software proof*.

We used the PDF parameter to ensure that the only variable in the search with the two APIs was the different terms of formal verification.

The Table 2.1 illustrates the results obtained.

| | **Custom Search JSON API** | **Wayback Machine** |
|---|---|---|
| ***formal verification*** | 3 | 112 901 |
| ***formally verified*** | 7 | 16 676 |
| ***formal methods*** | 7 | 444 118 |
| ***software proofs*** | 9 | 44 863 |
| ***software proof*** | 2760 | 257 974 |

Table 2.1: Number of PDF results for each term for each search API

We verified that the Wayback Machine presented more significant results for the different terms compared to Custom Search JSON API. This difference is because the Custom Search JSON API only returns results from the Google database, while the Wayback Machine searches in several domains. Furthermore, the Custom Search JSON API only makes the patents available, hence the number of results shown in the Table 2.1. We have to pay to have access to the content. For this reason, mainly, we chose to use the Wayback Machine as the search API for this investigation.

We followed some steps to perform the search using the Wayback Machine. The first was to choose the terms for the search query and the restrictions we wanted to add. We explain this in subsection 3.2.1. Then, we performed the search using the Wayback Machine and analyzed each obtained result (591 in total). We present this in detail in subsection 3.2.2.

## 2.3 Background descriptions

After performing the systematic search, we obtained different results for descriptions of the formal verification concept in the current literature. We present some of them below.

1. "In the context of hardware and software systems, formal verification is the act of proving or disproving the correctness of intended algorithms underlying a system with respect to a certain formal specification or property, using formal methods of mathematics. Formal verification can be helpful in proving the correctness of systems such as: cryptographic protocols, combinational circuits, digital circuits with internal memory, and software expressed as source code.

   The verification of systems is done by providing a formal proof on an abstract mathematical model of the system, the correspondence between the mathematical model and the nature of the system being otherwise known by construction." [18].

2. "Formal methods use mathematics or mathematical analysis techniques in the development of a design.

Exhaustive and partial simulations are replaced or augmented with mathematical proofs or a systematic state space search." [24].

3. "Proving system properties by using concepts and techniques from logic and discrete mathematics. Design and the properties: must be expressed in a formally defined language." [21].

4. "An automated decision procedure that can prove or disprove statements in some logical system of reasoning." [9].

### 2.3.1 Discussion

From the systematic search results, we could make some important conclusions about the current existing descriptions of the formal verification concept.

We could see that some descriptions require previous ideas of other technical concepts or even some background knowledge about the formal verification concept they are describing. An example is description 1, which mentions terms such as

> "cryptographic protocols, combinational circuits, digital circuits with internal memory, and software expressed as source code" [18].

that are unknown for non-expert users. Another example is the description 3, which uses concepts such as

> "logic and discrete mathematics [...] formally defined language." [21].

without explaining them. Introducing new technical concepts only makes the description more challenging to understand by ordinary users. Descriptions like these are intended for experts in the field and not for non-expert users.

We also could see that in the description 2, there is no clear definition of the purpose of formal verification methods, which is vital information for users unfamiliar with that subject.

Regarding the extent of the descriptions found, we had examples of descriptions that are too extensive, such as 1, which give more information beyond the necessary, turning it more complex to understand. Others are more concise, such as 3, or excessively simplistic, such as 4, which is missing crucial information.

## 2.4 Qualitative Data Analysis

To analyze qualitative data, we should create a set of coding categories that precisely summarize or describes the underlying relationships or patterns hidden in the data while evaluating text content. It is typically advised to use numerous coders to create high-quality coding for the data. The coders should continuously investigate the data for statements containing important information, inquire about the data, and make comparisons at various levels [20].

One accessible and flexible method for coding the data is **thematic analysis**. A method of conducting research that could otherwise appear hazy, enigmatic, conceptually tricky, and unduly complex is made more understandable by the thematic analysis. It provides an entry point into qualitative research and methodically teaches the mechanics of coding and evaluating qualitative data, which can be connected to more general theoretical or conceptual difficulties [12].

In [12], the authors mentioned a six-phase approach to the thematic analysis method:

**Phase 1-Familiarizing Yourself With the Data:** During this stage, we should actively study and reread textual data critically and analytically to fully understand what the facts mean [12].

**Phase 2-Generating Initial Codes:** In this stage, coding-based systematic data analysis is started. Coding identifies and labels a data feature pertinent to the study issue. Once we have determined which data extract has to be coded, we must record the code and highlight the relevant text. A chunk of the data can be coded using several codes [12].

**Phase 3-Searching for Themes:** The coded data is reviewed to find locations where the codes overlap and are comparable [12].

**Phase 4-Reviewing Potential Themes:** The emerging themes are examined about the coded data and the complete data set during this iterative phase. The main focus of this stage is quality assurance [12].

**Phase 5-Defining and Naming Themes:** In this stage, extracts are chosen to exhibit and examine before the story of each theme is laid out around or in conjunction with them [12].

**Phase 6-Producing the Report:** Thematic analysis' final phase is to write a report, such as a journal article or a dissertation. We should keep writing throughout the process [12].

Throughout the coding process, reliability control metrics should be computed and assessed. One of these metrics is **Cohen's Kappa**.

A statistical metric that evaluates the degree of agreement between two raters or observers who give categorical evaluations to a group of objects or subjects. It examines the agreement beyond chance while considering the agreement that could happen by chance [28].

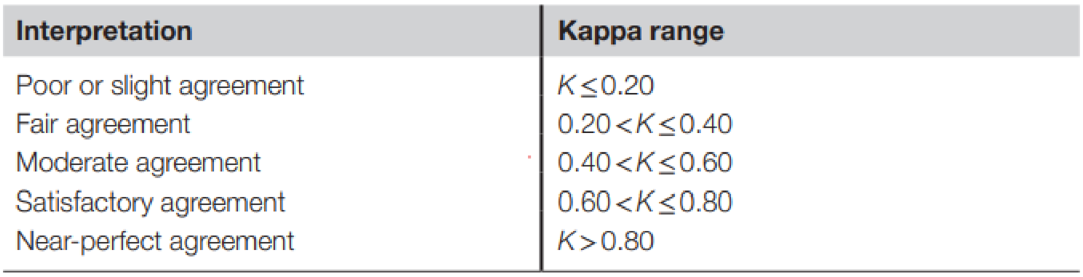
Cohen's kappa formula is presented in Figure 2.1. Pr($\alpha$) is the observed agreement, Pr(e) is the expected agreement, and K is the Cohen's Kappa value.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Figure 2.1: Cohen's kappa formula [28]

Cohen's kappa ranges from -1 to 1, where a score of 1 denotes complete agreement among the raters. The 0 represents a random agreement, and a number below 0 denotes an agreement worse than the random.

In Figure 2.2, we can see the range of values and their interpretation.



| Interpretation | Kappa range |
|---|---|
| Poor or slight agreement | $K \leq 0.20$ |
| Fair agreement | $0.20 < K \leq 0.40$ |
| Moderate agreement | $0.40 < K \leq 0.60$ |
| Satisfactory agreement | $0.60 < K \leq 0.80$ |
| Near-perfect agreement | $K > 0.80$ |

Figure 2.2: Interpretation of Cohen's Kappa [20]

Cohen's Kappa must be at or above 0.60 in order to be considered adequate for intercoder reliability [20].

## 2.5 Approaches to related problems

There are different approaches to testing users' understanding of a technical concept beyond their knowledge.

One of them is by doing interviews, structured or non-structured, which can provide qualitative and precise information that can serve as the foundation for comprehending the aims, expectations, vocabulary, and perceptions of users [13].

Another approach is using questionnaires, a powerful quantitative analysis tool that we can use with other methods, such as usability tests and interviews. According to the study's intent, we must adapt these methodologies, choose the more appropriate sampling method for the study, and consider all possible biases that will influence the results [13, 25].

In [8], we found an approach to communicate a technical concept to ordinary users. In that case, regarding E2EE (End-to-end encryption), which restricts access to online messages for service providers and unaffiliated third parties. Researchers discovered that non-expert users do not comprehend the advantages and constraints of the E2EE concept. So they conducted a qualitative study to explain that concept to general users.

The study started with a quiz regarding participants' thoughts on E2EE. Then a brief E2EE tutorial was presented to each participant, and a quiz about participants' opinions of the E2EE concept. In an interview, they discussed participants' responses. Then, the participants criticized some current E2EE descriptions from existing apps, offered feedback on the tutorial, and created a brief message explaining E2EE in their own words [8].

Despite the small sample and the fact that the participants were more educated and younger than the overall public, they could take some essential conclusions. Participants' changes to their quiz answers after the tutorial and their interview interventions supported that the participants' understanding of E2EE improved due to the tutorial presented. However, some participants kept existing misunderstandings or developed new ones [8].

In [15], we found a different approach with the same intent. In this work, the authors introduce the concept of Differential Privacy (DP), which makes it simpler to compute aggregate statistics about a dataset and formally limits the availability of data that these statistics may disclose about single data points within the dataset.

In that research, the authors looked into DP from the user's point of view, highlighting how DP and users' privacy expectations connect as it is likely to be used in practice. To determine the expected behavior of respondents, they conducted two vignette-based surveys [15].

Vignettes are concise, skillfully crafted descriptions of a character, thing, or circumstance. In vignette studies, vignettes are presented to participants to elicit the participants' ideas, attitudes, judgments, understanding, or planned behavior [6].

The methodology used in the first survey starts with five cognitive interviews. Then the participants also had the chance to list any other information disclosures vital to them [15].

On the other hand, the methodology used in the second survey started by gathering and analyzing DP descriptions used in practice and filtered six of them that were representative. In order to remove the description from its context and make it uniform in its structure, they also generated new descriptions [15].

This user study had some limitations, mainly possible biases associated with that type of study, such as sampling bias, where they used MTurk to select the sample, or reporting biases through the design of the questions. Despite that, they conclude that the interaction between users' innate privacy concerns and how explanations of DP set user expectations affects users' willingness to provide their information under different levels of privacy assurances [15].

The authors concluded that descriptions directly influence user expectations.

## 2.6 User studies using surveys

In [20], the authors defined a survey as

> "a well-defined and well-written set of questions to which an individual is asked to respond" [20].

Commonly, surveys are responded to by an individual without the presence of a researcher. As a result, the data collected could be more thorough with additional research techniques, for example, focus groups [20].

Surveys can gather numerous replies from a geographically distributed user group. When designed with random sampling, we can also produce statistically precise calculations. A significant drawback of surveys is that they may produce biased results when they ask about usage patterns rather than prominent factual events. The significant difficulty is coming up with well-written and impartial questions [20].

Survey questions must be pilot tested to guarantee their unbiased, straightforwardness, and clarity to assure validity and reliability. The general design should make it simple for respondents to comprehend. Even if they are non-probabilistic, we must use appropriate sampling techniques to provide adequate results to address the study's objectives [20].

Surveys can be used with other research techniques like focus groups or interviews [20].

# Chapter 3

# Study of users' expectations for formal verification

In this chapter, we outline the primary issue and explain our approach to the problem, including the steps we made. We analyze the obtained results from the performed study and conclude.

We present the problem statement in section 3.1. Then, in section 3.2, we explain our methodology for performing the users study. In section 3.3, we make a results analysis regarding the participants' demographic information and the questionnaire' answers. The section 3.4 mentions the limitations of our study; in section 3.5, we have the threats to validity. Ultimately, we make our conclusions regarding the study results in section 3.6.

## 3.1 Problem Statement

After analyzing the current literature regarding formal verification, we found multiple descriptions trying to explain the concept. However, we noticed that these descriptions could be too complex for non-expert users to understand the concept completely.

We did not find user studies that show whether the non-expert users fully understand the existing descriptions or not. We consider these user studies essential to guarantee that the users fully understand the concept of formal verification and comprehend what their guarantees are and are not.

As there are no human-centered studies on better communicating formal verification to users unfamiliar with this subject area, we performed a user study to work towards filling this gap. In this way, we aim to inform the community how to explain formal verification to non-expert users and improve their understanding of this concept.

## 3.2 Methodology

We intend to understand how to communicate the formal verification concept to users more effectively. To achieve this, we proposed developing a human-centered study using a survey to comprehend better if users understand the concept and which descriptions of formal verification are more effective in conveying this concept to users.

All user studies are prone to biases that can emerge from the respondents' or researchers' sides. From our research side, we ensured that our questions are valid, the data we received is truthful, and our interpretation of the results was accurate.

We were cautious with the survey designs to reduce all the possible biases and have more truthful conclusions.

This investigation required some crucial steps described below to achieve the desired goals.

### 3.2.1 Finding the search query

We searched for descriptions of formal verification in reliable sources accessible to anyone. We used sources such as Forbes, Quanta Magazine, IEEE (Institute of Electrical and Electronics Engineers), ACM (Association for Computing Machinery), Google Scholar, and Science Direct. From these sources, we obtained a total of 24 formal verification descriptions. We present some of these descriptions below. The descriptions' quality was not considered at this stage of the investigation.

1. "The researchers established the reliability of their file system through a process known as **formal verification**. From Wikipedia, **formal verification** is "…the act of proving or disproving the correctness of intended algorithms underlying a system with respect to a certain formal specification or property, using formal methods of mathematics"." [14].

2. "The technology that repelled the hackers was a style of software programming known as **formal verification**. Unlike most computer code, which is written informally and evaluated based mainly on whether it works, **formally verified** software reads like a mathematical proof: Each statement follows logically from the preceding one. An entire program can be tested with the same certainty that mathematicians prove theorems." [19].

3. "Hales set out to use a technique called **formal proof verification** in which a computer program uses logic and the axioms to assess each baby step of a proof." [29].

4. "**Formal verification** is a powerful technique used to mathematically prove that an appropriately scaled model of a system does or does not exhibit desirable properties." [11].

5. "In principle, this is no different from verifying mathematical claims; for the purposes of **formal verification**, hardware and software systems must be described in mathematical terms, and the statement that such a system meets a certain specification is a theorem to be proved." [7].

6. "**Formal Verification** is a promising method to provide security guarantees by mathematically ascertaining the correctness of designs using a diverse set of mathematical and logical methods. These methods are particularly useful in order to get quantitative statements about safety and security properties of digital systems" [**?**].

7. "**Formal Verification** (FV): the use of tools that mathematically analyze the space of possible behaviors of a design, rather than computing results for particular values." [26].

As we can see in the examples above, we concluded that the keywords mainly used in the descriptions are *formal verification* and *formally verified*.

Furthermore, in description 3, we noticed that sometimes other terms occur between the words *formal* and *verification*.

Taking these aspects into account, we built the string below.

**"formal * verification" OR "formally verified"**

We considered the date interval for the search between the first day of the first month of 2001 and the day before we searched in the Wayback Machine. We only collected descriptions from 2001 onwards since much older information may be outdated. Thus, we used the query below to perform the search in the Wayback Machine.

**("formal * verification" OR "formally verified") AND publicdate:[2001-01-01 TO 2022-10-11]**

The process is explained in the following subsection.

### 3.2.2 Collection and selection of descriptions

We performed the created query on the Wayback Machine.

As a result, we got a JSON file with 591 formal verification descriptions. Of these 591, we discarded descriptions whose media type was audio, software, or movies. Table 3.1 shows the number of descriptions discarded by each media type.

| mediatype | number of results |
|---|---|
| audio | 19 |
| software | 31 |
| movies | 30 |

Table 3.1: Number of discarded results per media type

So in total, we had 511 valid results to analyze. We analyzed these results and extracted each textual description, if any, into an Excel file. Analyzing each of these descriptions, we marked them as valid or not, if the text effectively described formal verification or not, respectively. Cases that raised doubts were analyzed and classified by more than one person separately to conclude whether it was an accurate description.

At the end of this analysis, we obtained 79 valid descriptions extracted from the Wayback machine.

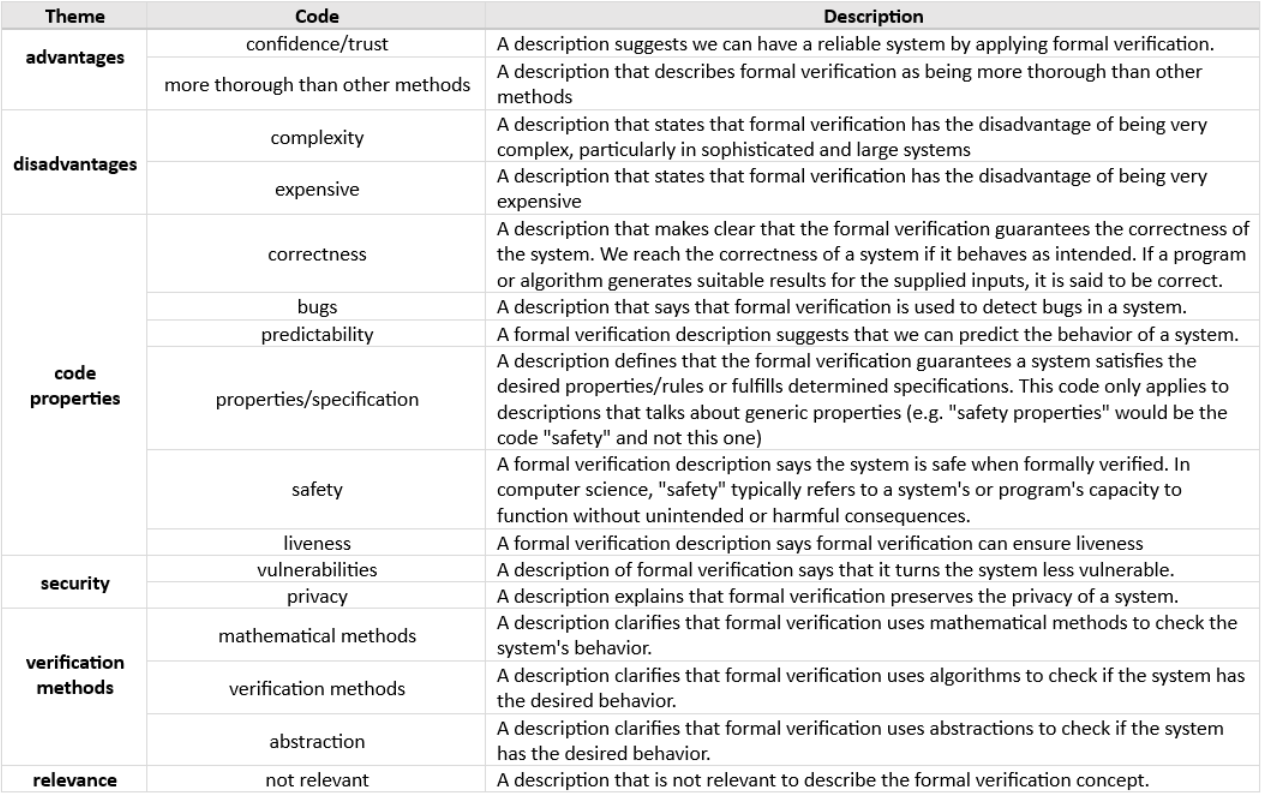
In addition to the descriptions collected using the Wayback Machine, we also collected descriptions by doing a Google search, as this is a standard method for users to search for information. For this search, we restricted the year of the results between 2018-2023 to have relatively recent results. We collected 15 descriptions from Google Search.

In total, we collected 94 accurate descriptions of formal verification. At this stage, the duplicates have not yet been filtered.

### 3.2.3 Thematic analysis

In [8], the authors investigated communicating better the concept of end-to-end encryption to non-experts. Their study evaluated participants' knowledge about E2EE before they presented a developed tutorial and after. To analyze the qualitative data, in that case, the participant's responses, two researchers created a codebook that encapsulates relevant responses' codes. This process was made independently.

Similar to the study presented in [8], we began by carefully reviewing and becoming acquainted with the qualitative data we gathered. Then, we started a coding-based systematic analysis to generate the first codes. We had two coders for this process. After we had the first codes, the two coders met online to discuss the codes found. The two coders considered some of the descriptions irrelevant, either because they did not effectively describe formal verification or because they were duplicate descriptions. The coders removed these descriptions from the analysis. We are left with 86 descriptions. Since there were differences in the remaining descriptions, the two coders analyzed all the data again, determining whether more codes existed and including the agreed codes. After this second review, the two coders met online again to discuss the codes. We studied these codes and searched for patterns or similarities among them. We found five themes: *advantages*, *disadvantages*, *code properties*, *security*, and *verification methods*. We grouped the codes into themes and created a codebook. We created precise definitions that explain what each code in the data stands for. We provide examples for each code. These examples act as a guide to help us comprehend and effectively use the codes. The partial codebook with the themes, codes, and the respective description is presented in Figure 3.1. The complete codebook can be found in Appendix B.



| Theme | Code | Description |
|---|---|---|
| **advantages** | confidence/trust | A description suggests we can have a reliable system by applying formal verification. |
| | more thorough than other methods | A description that describes formal verification as being more thorough than other methods |
| **disadvantages** | complexity | A description that states that formal verification has the disadvantage of being very complex, particularly in sophisticated and large systems |
| | expensive | A description that states that formal verification has the disadvantage of being very expensive |
| **code properties** | correctness | A description that makes clear that the formal verification guarantees the correctness of the system. We reach the correctness of a system if it behaves as intended. If a program or algorithm generates suitable results for the supplied inputs, it is said to be correct. |
| | bugs | A description that says that formal verification is used to detect bugs in a system. |
| | predictability | A formal verification description suggests that we can predict the behavior of a system. |
| | properties/specification | A description defines that the formal verification guarantees a system satisfies the desired properties/rules or fulfills determined specifications. This code only applies to descriptions that talks about generic properties (e.g. "safety properties" would be the code "safety" and not this one) |
| | safety | A formal verification description says the system is safe when formally verified. In computer science, "safety" typically refers to a system's or program's capacity to function without unintended or harmful consequences. |
| | liveness | A formal verification description says formal verification can ensure liveness |
| **security** | vulnerabilities | A description of formal verification says that it turns the system less vulnerable. |
| | privacy | A description explains that formal verification preserves the privacy of a system. |
| **verification methods** | mathematical methods | A description clarifies that formal verification uses mathematical methods to check the system's behavior. |
| | verification methods | A description clarifies that formal verification uses algorithms to check if the system has the desired behavior. |
| | abstraction | A description clarifies that formal verification uses abstractions to check if the system has the desired behavior. |
| **relevance** | not relevant | A description that is not relevant to describe the formal verification concept. |

Figure 3.1: Descriptions' partial codebook

The two coders started coding all the descriptions separately based on the created codebook. Any inconsistencies were then discussed to fix coding problems and guarantee intercoder dependability.

We ended this process using each code at least once. We created the Table 3.2 to analyze the percentage of use of each code.

| Codes | Number of times the code was used | Percentage |
|---|---|---|
| confidence/trust | 8 | 8,89% |
| more thorough than other methods | 9 | 10,00% |
| correctness | 34 | 37,78% |
| bugs | 17 | 18,89% |
| predictability | 1 | 1,11% |
| properties/specification | 56 | 62,22% |
| safety | 11 | 12,22% |
| liveness | 5 | 5,56% |
| complexity | 11 | 12,22% |
| expensive | 5 | 5,56% |
| vulnerabilities | 5 | 5,56% |
| privacy | 2 | 2,22% |
| mathematical methods | 39 | 43,33% |
| verification methods | 33 | 36,67% |
| abstraction | 12 | 13,33% |

Table 3.2: Percentage of use of each code

Observing the table above, we could conclude that the most used code was **properties/specifications**. It was used in more than half of the formal verification descriptions.

The codes mostly used beyond properties/specification code were mathematical methods, correctness, and verification methods.

### 3.2.4 Cohen's kappa

To obtain Cohen's kappa, we calculated the observed agreement, Pr($\alpha$), the percentage of times the ratings the two raters have agreed [28]. We calculated this by dividing the number of agreements (70) by the total ratings (86). We obtained the value 0,8139534884.

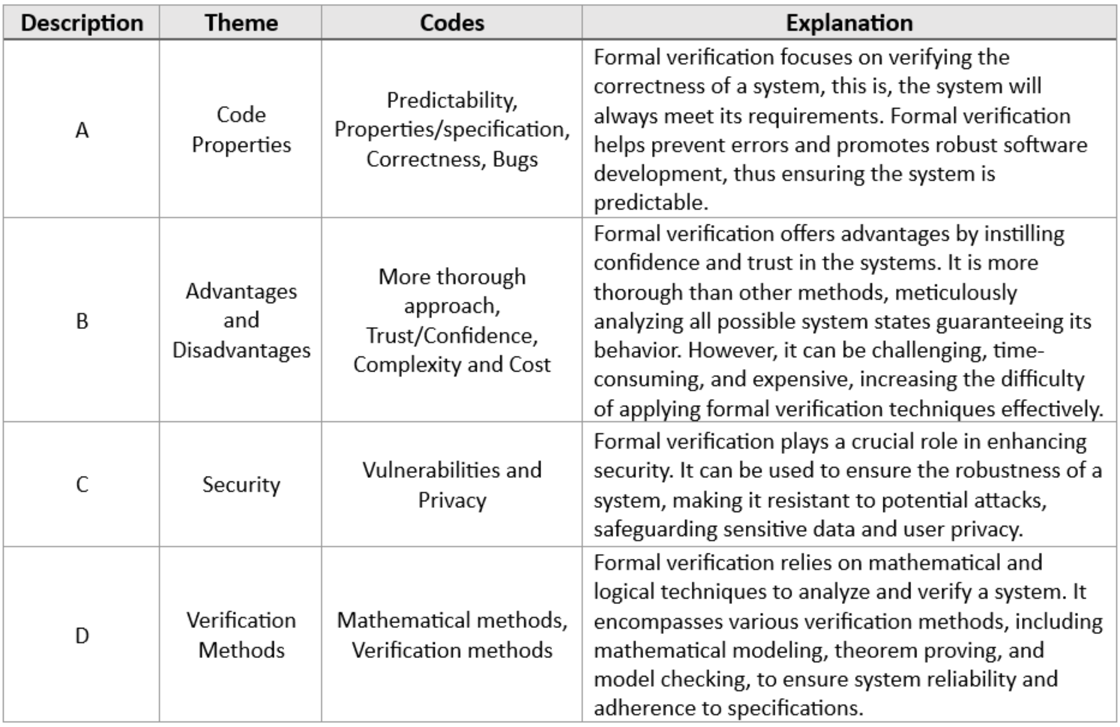
Then we find the expected agreement, Pr(e), which demonstrates the agreement that would be expected based solely on chance. To accomplish this, we determined the proportion of each rater's ratings for each category. The expected agreement is then computed by multiplying the percentages for each category. [28]

We created a table to obtain this value by calculating the probability that the coders would randomly agree. For each row, we calculated the number of responses of that category given by the first coder divided by the total responses multiplied by the number of responses given by the second coder divided by the total responses. We obtained the value 0,030.

In [20], the authors refer that a value of Cohen's Kappa above 0.60 is often interpreted as indicating satisfactory reliability [20]. By applying the formula, we obtained 0,8081160229 as the value of Cohen's kappa, which is a near-perfect agreement. It should be noted that the thematic analysis was carried out by two coders separately, and an agreement was reached on the codes and themes that would belong to the codebook together. This value demonstrates that the two coders understood where to use each code well.

### 3.2.5 Study design

In the thematic analysis, we identified four themes (*code properties*, *advantages and disadvantages*, *security*, and *verification methods*). The research team developed a formal verification description to represent each theme. In Figure 3.2, we can see the created explanations of formal verification. Designing these descriptions was an iterative process, where constant feedback from the research team allowed improvements in the explanations.



| Description | Theme | Codes | Explanation |
|---|---|---|---|
| A | Code Properties | Predictability, Properties/specification, Correctness, Bugs | Formal verification focuses on verifying the correctness of a system, this is, the system will always meet its requirements. Formal verification helps prevent errors and promotes robust software development, thus ensuring the system is predictable. |
| B | Advantages and Disadvantages | More thorough approach, Trust/Confidence, Complexity and Cost | Formal verification offers advantages by instilling confidence and trust in the systems. It is more thorough than other methods, meticulously analyzing all possible system states guaranteeing its behavior. However, it can be challenging, time-consuming, and expensive, increasing the difficulty of applying formal verification techniques effectively. |
| C | Security | Vulnerabilities and Privacy | Formal verification plays a crucial role in enhancing security. It can be used to ensure the robustness of a system, making it resistant to potential attacks, safeguarding sensitive data and user privacy. |
| D | Verification Methods | Mathematical methods, Verification methods | Formal verification relies on mathematical and logical techniques to analyze and verify a system. It encompasses various verification methods, including mathematical modeling, theorem proving, and model checking, to ensure system reliability and adherence to specifications. |

Figure 3.2: Created formal verification descriptions

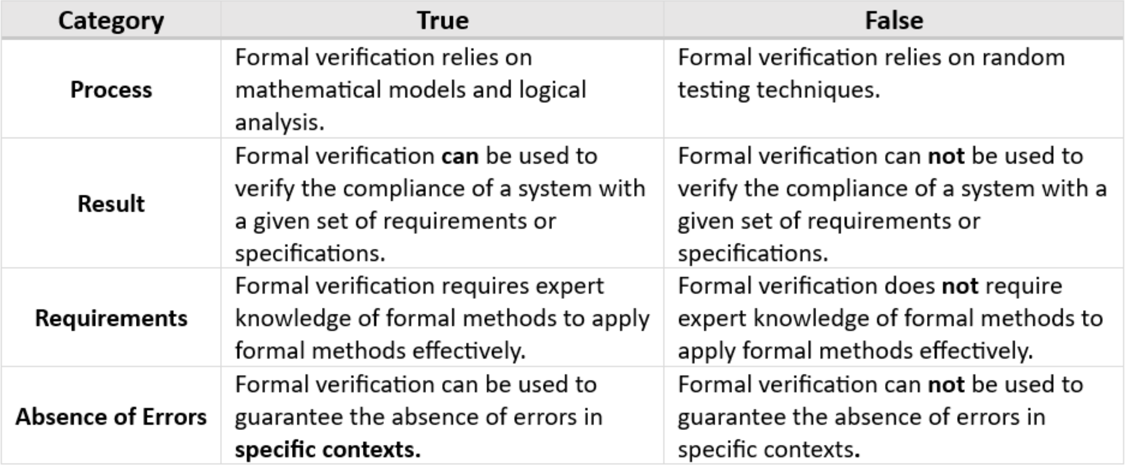
In [16], the authors conducted a study to evaluate how effectively their notifications communicate the privacy risk information and the participants' comfort level in understanding the concept. They presented one notification to the participants, and then they assessed the subjective comprehension of the participants using a Likert scale where the participants had to evaluate a set of statements. Then, the authors of this study performed an objective understanding using true or false questions. They randomized the options to reduce bias [16].

The research team developed, for each description, one equal questionnaire. Similar to the study presented in [16], we presented the description to the participant, and then we assessed the participants' subjective knowledge by asking them to rate the four sentences below. We used a 5-point Likert Scale and randomized the order of these statements.

- *I have understood the explanation.*

- *After reading the explanation, I can make an informed decision about using formally verified software.*
- *The explanation provided all the information I wanted to know about formal verification.*
- *After reading the explanation, I would prefer to use a formally verified software.*

We then assessed participants' objective knowledge with four true or false questions (Figure 3.3). We have added the "Don't know" option since it is considered unethical not to have that option [25]. All sentences have two versions, one true and one false. The participants only evaluate one statement of each category. We randomized the presentation of these versions.



| Category | True | False |
|---|---|---|
| **Process** | Formal verification relies on mathematical models and logical analysis. | Formal verification relies on random testing techniques. |
| **Result** | Formal verification **can** be used to verify the compliance of a system with a given set of requirements or specifications. | Formal verification can **not** be used to verify the compliance of a system with a given set of requirements or specifications. |
| **Requirements** | Formal verification requires expert knowledge of formal methods to apply formal methods effectively. | Formal verification does **not** require expert knowledge of formal methods to apply formal methods effectively. |
| **Absence of Errors** | Formal verification can be used to guarantee the absence of errors in **specific contexts.** | Formal verification can **not** be used to guarantee the absence of errors in specific contexts. |

Figure 3.3: True/False statements

We presented the description again, and through three open questions, we asked the participants to elaborate on the positive and negative aspects of the description they saw, if any. Then, the question for further feedback is optional. We asked the participants demographic questions at the end of the survey.

The questionnaire can be found in Appendix A.

### 3.2.6 Cognitive interviews

More minor measurement mistakes can be found with cognitive interviews, which entail asking respondents questions about each survey question and having them think aloud while completing it. We should understand how the participants feel while answering the questions or their interpretation of the questions. In the survey literature, cognitive interviewing is frequently advised as a crucial pre-testing strategy [25].

We performed a first cognitive interview, where the participant detected a typographical error in the questionnaire and told us that the "Don't know" option was missing. We corrected the word that was wrong written, and we added the option to the true or false answers. Then, we conducted another cognitive interview with a different person. In that case, all went well.

### 3.2.7 Performed study

We shared the questionnaires with the personal network. Participation was voluntary, and we did not pay the participants.

In order to ensure randomness and reduce possible bias as much as possible, we randomly generated four groups with the study's participants. Then, we assigned a number to each questionnaire, randomized the order of the numbers, and assigned the questionnaire to each group of participants with the equivalent number. In this way, we guarantee randomness in the distribution of the questionnaires among the participants.

## 3.3 Results analysis

### 3.3.1 Participants' demographic information

We had 130 participants in this study. Of these participants, 69 fully completed the questionnaire, so we only considered these 69 questionnaires in the analysis of results.

All participants in this study currently reside in Portugal.

Regarding the age of the participants, 57.9% are younger than 35 years. The other 42.1% comprise participants aged between 35 and 64 years old. The most frequent age group was between 25 and 34 (Figure 3.4). The sample was represented mainly by female participants, with a percentage of 56.5%. We had 39.1% male participants and 1,4% non-binary (Figure 3.5). Concerning the level of education, 58% are graduated or have a professional degree (Figure 3.6). With the question, *"Have you worked professionally in a field related to computers?"* we could conclude that 55.1% of the participants already worked with computers, and 44,9% did not (Figure 3.7).

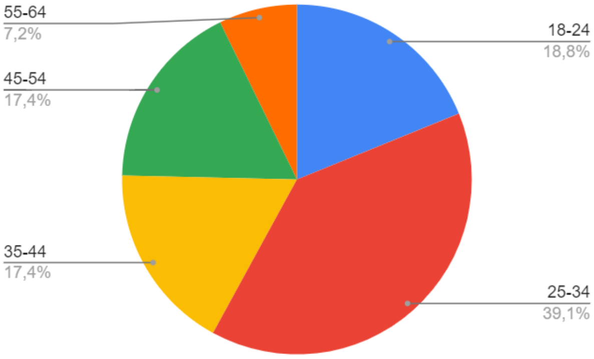


Figure 3.4: Participants' age range

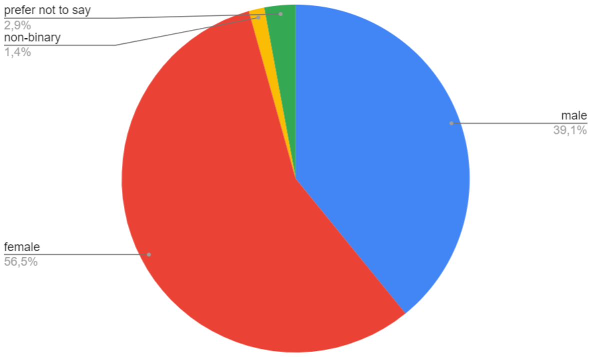


Figure 3.5: Participants' gender

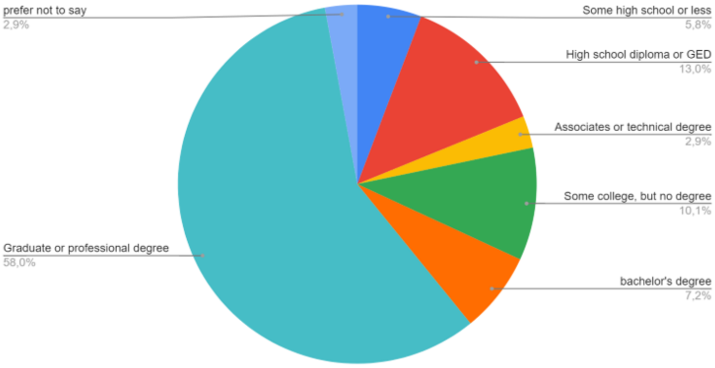


Figure 3.6: Participants' education level

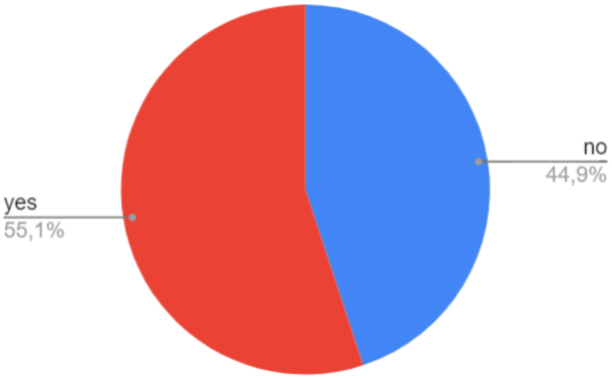


Figure 3.7: Participants who already worked with computers

Most of the participants work in industry, mainly in health and engineering. About 22 participants are computer engineers.

We asked participants whether the formal verification concept was familiar before undertaking the study. More than half answered no. It was 58%. In figure 3.8, we can see the distribution of percentages around familiarity with the concept of formal verification.

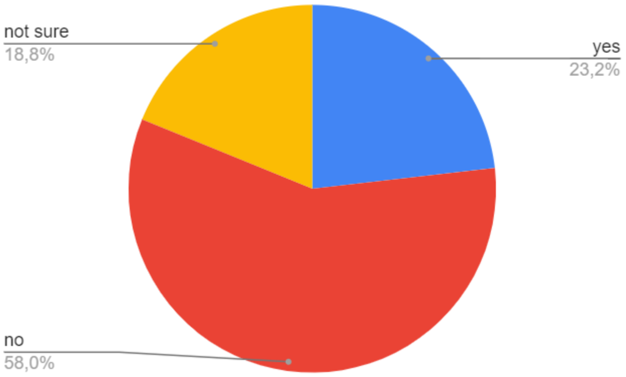


Figure 3.8: Participants' familiarity with the concept

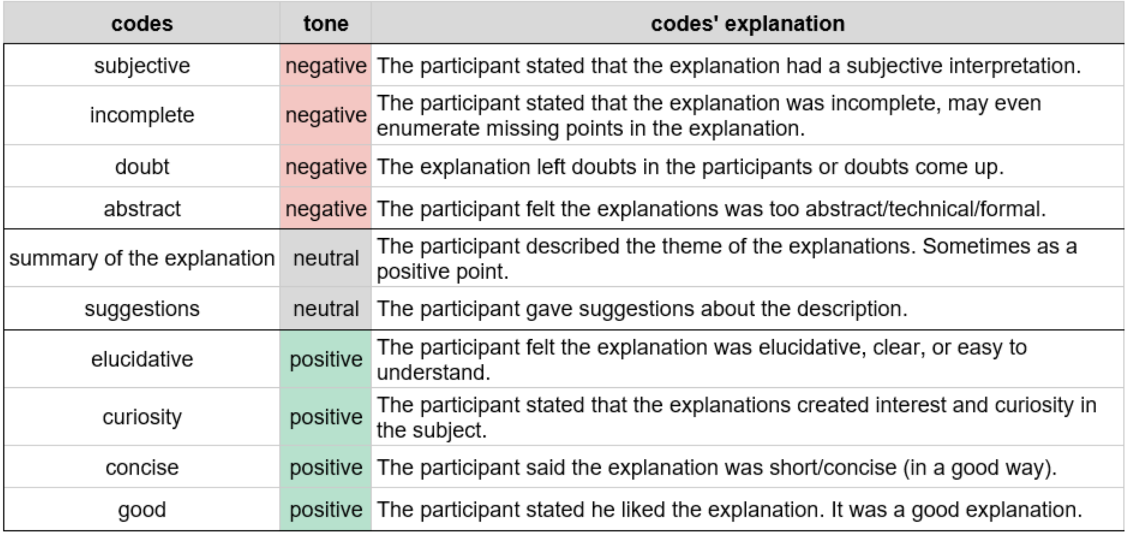
As we can see, we have 16 participants out of the 69 who said they were familiar with the formal verification concept. We still considered these participants in the analysis of the questionnaires, checking for each question to see if there were any cases where the number of correct answers was related to familiarity with the concept. We could not conclude that there was a direct relationship between them.

### 3.3.2 Questionnaires' answers

Regarding the number of responses, we obtained 69 completed answers in total. In the questionnaire with description A, we obtained 17 complete answers. The questionnaire with description B had 16 responses, and the one with description C had 15. Finally, the questionnaire with description D had 21 complete responses. The description corresponding to the letter (A, B, C, D) can be found in Figure 3.2.

We collected and analyzed information for the four questionnaires individually since they had different descriptions. For each questionnaire, we evaluated the answers to the first question (participants' subjective knowledge) and the true or false questions (participants' objective knowledge). We also studied the participants' feedback on the open-ended questions.

Regarding open-ended questions, we have developed a codebook, as in the paper [8]. In this process, two coders were involved. After collecting all answers to the open questions, the two coders analyzed all the data individually to find codes in the responses. After the first analysis, the coders met online and discussed the codes found to reach a consensus. Both coders agreed that responses that did not answer the question or were irrelevant to the study, as they were of low quality, were excluded from the analysis. We present the partial codebook in Figure 3.9. The complete codebook can be found in Appendix C.



| codes | tone | codes' explanation |
|---|---|---|
| subjective | negative | The participant stated that the explanation had a subjective interpretation. |
| incomplete | negative | The participant stated that the explanation was incomplete, may even enumerate missing points in the explanation. |
| doubt | negative | The explanation left doubts in the participants or doubts come up. |
| abstract | negative | The participant felt the explanations was too abstract/technical/formal. |
| summary of the explanation | neutral | The participant described the theme of the explanations. Sometimes as a positive point. |
| suggestions | neutral | The participant gave suggestions about the description. |
| elucidative | positive | The participant felt the explanation was elucidative, clear, or easy to understand. |
| curiosity | positive | The participant stated that the explanations created interest and curiosity in the subject. |
| concise | positive | The participant said the explanation was short/concise (in a good way). |
| good | positive | The participant stated he liked the explanation. It was a good explanation. |

Figure 3.9: Participants responses' partial codebook

The two coders coded all participants' responses. In order to assess the reliability of the coding, we calculated Cohen's Kappa value.

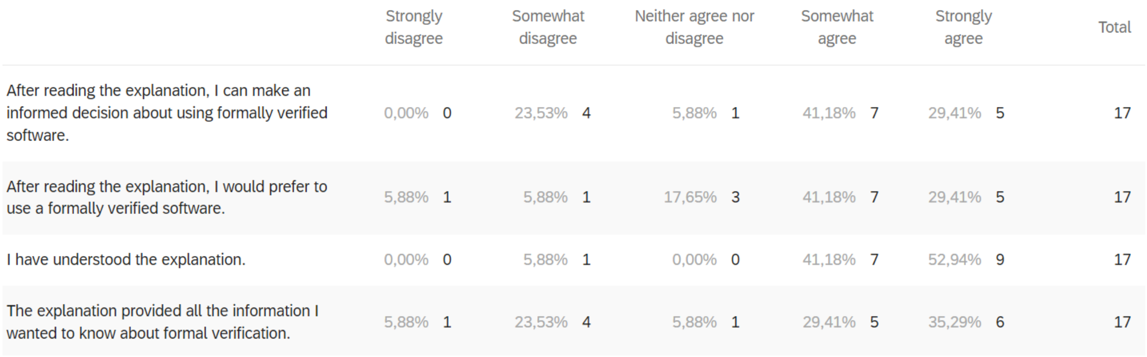
To obtain Cohen's kappa, we calculated the observed agreement, Pr($\alpha$), the percentage of times the ratings the two raters have agreed [28]. We calculated this by dividing the number of agreements (64) by the total ratings (75). We obtained the value 0,8533333333.

We created a table to obtain expected agreement by calculating the probability that the coders would randomly agree. For each row, we calculated the number of responses of that category given by the first coder divided by the total responses multiplied by the number of responses given by the second coder divided by the total responses. We obtained the value 0,164.

The value of Cohen's Kappa was 0,8246173469 (near-to-perfect agreement), which checks reliability since it is above 0.60 [20].

#### 3.3.2.1 Questionnaire with Description A

Regarding the first question, where participants had to indicate the level of agreement with the sentences, most answers were *"Somewhat agree"* or *"Strongly agree"*. For the first sentence, *"After reading the explanation, I can make an informed decision about using formally verified software."*, we obtained 23.53% of the 17 participants who answered that they disagreed with the sentence, and 70.59% agreed with it. For the expression, *"After reading the explanation, I would prefer to use a formally verified software."*, 11.76% of participants disagreed, and 70.59% agreed that they would prefer to use formally verified software after reading the description. Regarding understanding the description, only one participant demonstrated that he did not understand, while 94.12% agreed that they had understood the explanation. For the last sentence, *"The explanation provided all the information I wanted to know about formal verification."*, 29.41% of the answers revealed that the participants disagreed with the sentence, and more than half, 64.7%, agreed that they got the information they wanted with the explanation. We had, in total, 5 "Neither agree nor disagree" answers to the four statements (Figure 3.10).



| | Strongly disagree | | Somewhat disagree | | Neither agree nor disagree | | Somewhat agree | | Strongly agree | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| After reading the explanation, I can make an informed decision about using formally verified software. | 0,00% | 0 | 23,53% | 4 | 5,88% | 1 | 41,18% | 7 | 29,41% | 5 | 17 |
| After reading the explanation, I would prefer to use a formally verified software. | 5,88% | 1 | 5,88% | 1 | 17,65% | 3 | 41,18% | 7 | 29,41% | 5 | 17 |
| I have understood the explanation. | 0,00% | 0 | 5,88% | 1 | 0,00% | 0 | 41,18% | 7 | 52,94% | 9 | 17 |
| The explanation provided all the information I wanted to know about formal verification. | 5,88% | 1 | 23,53% | 4 | 5,88% | 1 | 29,41% | 5 | 35,29% | 6 | 17 |

Figure 3.10: Questionnaire 1- Participants' subjective understanding

We obtained the results below regarding the true or false questions that the participants had to answer. For each question, we analyzed the true and false versions of the sentences.

1. **Question 1**

   - *Formal verification relies on mathematical models and logical analysis.*

     A total of 7 participants responded to this question, of which 4 responded as a true statement. The remaining 3 participants selected the option *"Don't know"*. Since this statement is effectively true, we could conclude that 57.14% of the 7 participants answered correctly.

   - *Formal verification relies on random testing techniques.*

     Regarding this statement, we had 10 responses, of which 4 were marked as *"False"*, 1 as *"True"* and 5 as *"Don't know"*. The correct answer to this statement is that it is false, so we could see that 40% of the participants answered correctly, and 10% got it wrong. We had 50% of participants answering *"Don't know"*, which may indicate that the explanation did not provide enough information to evaluate this statement. In fact, the description that the participants read focuses on the code properties, not mentioning which verification methods are used by formal verification.

2. **Question 2**

   - *Formal verification can be used to verify the compliance of a system with a given set of requirements or specifications.*

     A total of 12 participants answered this question. Of this total, 10 chose the true option, and the remaining 2 said they did not know the answer. This statement is correct, so 83.33% answered correctly, and no participant made the wrong choice.

   - *Formal verification can not be used to verify the compliance of a system with a given set of requirements or specifications.*

For this sentence, there were 5 responses. Of these participants, 4 responded that the statement was false, and only 1 answered that he did not know. The statement is false, so 80% got it correct, and there were no wrong answers.

The description addresses this topic, so having 14 participants out of 17, correctly answer the question demonstrates that the description communicates the concept of formal verification concerning the theme **code properties** well.

3. **Question 3**

   - *Formal verification requires expert knowledge of formal methods to apply formal methods effectively.*

     There were a total of 8 responses to this statement. In this case, 2 participants answered that the statement was true, 4 chose false, and 2 did not know the answer. The statement is true, so only 25% responded correctly, and 50% of the participants got it wrong. It may show that the description needs more information to answer this question. Nothing is said in the description about the knowledge necessary for applying formal verification methods.

   - *Formal verification does not require expert knowledge of formal methods to apply formal methods effectively.*

     We had 9 responses to this question, where 1 response was true. There were 5 false answers, and 3 participants replied: *"Don't know"*. The sentence is false, so 55.56% answered correctly and 11.11% incorrectly.

   We can verify that the true and false versions of the sentence got similar results. The description needs to provide more information to evaluate these statements accurately.

4. **Question 4**

   - *Formal verification can be used to guarantee the absence of errors in specific contexts.*

     This statement regarding formal verification is true. We got 7 answers, with 100% of the participants getting the option right. The description clearly says: *"Formal verification helps prevent errors[...]"*, which could justify all correct answers.

   - *Formal verification can not be used to guarantee the absence of errors in specific contexts.*

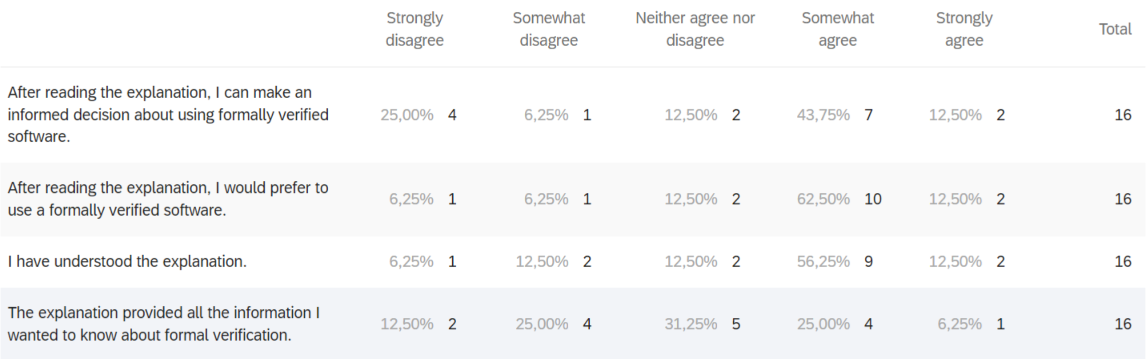
     We had 10 answers, of which 5 were false, 3 were true, and 2 were *"Don't know"*. The statement is false, showing that 50% of the participants selected the correct answer and 30% failed. These results showed that possibly the fact that the description says that *"Formal verification helps **prevent** errors[...]"*, but not that it **guarantees** the absence of errors, may raise doubts regarding the terms *prevent* and *guarantee* since preventing does not mean that there can be no errors.

This description fits the theme of **code properties**. It explicitly states that *"the system will always meet its requirements"*. The second question made it possible to assess whether the participants understood this point accurately, so we could conclude that there were no wrong answers. This description effectively communicated the formal verification concept concerning the associated theme. Furthermore, the description stated that it prevented errors from occurring. Although most participants answered correctly, the term used may have led some participants to question the statements in the third question.

Regarding the feedback from the participants collected in the open responses of this questionnaire and the codebook presented in Figure 3.9, they generally considered the description concise and enlightening. However, they also thought it was incomplete, and it raised doubts. Some participants suggested adding more detail to the explanation or use cases. One participant said, *"It does not mention if formal verification considerably increases development efforts by developers"*, (P16), which raised the question: Does formal verification increase development efforts? The answer to this question should be clarified in a formal verification description.

#### 3.3.2.2 Questionnaire with Description B

Most responses to the first question, which asked participants to rate how much they agreed with the statements after reading the formal verification description, were *"Somewhat agree"* and *"Strongly agree"*, except for the sentence *"The explanation provided all the information I wanted to know about formal verification."*, where the majority responded *"Somewhat disagree"* and *"Strongly disagree"*. We got 31.25% of the 16 participants to disagree with the first phrase, *"After reading the explanation, I can make an informed decision about using formally verified software."*, and 56.25% agreed. After reading the explanation, 75% of participants agreed, and 12,50% of participants disagreed with the statement, *"After reading the explanation, I would prefer to use a formally verified software."*. Three participants explicitly stated that they did not grasp the description, whereas 68.75% said they understood the explanation. In response to the final statement, where participants had to evaluate if the description gave all the needed information, 37.50% disagreed. In comparison, 31.25% said the explanation gave them the needed knowledge. Here, we had 31.25% answer *"Neither agree nor disagree"*, which is a considerable percentage only to one option. We had, in total, 11 "Neither agree nor disagree" answers to the four statements (Figure 3.11).



| | Strongly disagree | | Somewhat disagree | | Neither agree nor disagree | | Somewhat agree | | Strongly agree | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| After reading the explanation, I can make an informed decision about using formally verified software. | 25,00% | 4 | 6,25% | 1 | 12,50% | 2 | 43,75% | 7 | 12,50% | 2 | 16 |
| After reading the explanation, I would prefer to use a formally verified software. | 6,25% | 1 | 6,25% | 1 | 12,50% | 2 | 62,50% | 10 | 12,50% | 2 | 16 |
| I have understood the explanation. | 6,25% | 1 | 12,50% | 2 | 12,50% | 2 | 56,25% | 9 | 12,50% | 2 | 16 |
| The explanation provided all the information I wanted to know about formal verification. | 12,50% | 2 | 25,00% | 4 | 31,25% | 5 | 25,00% | 4 | 6,25% | 1 | 16 |

Figure 3.11: Questionnaire 2- Participants' subjective understanding

We had the results below concerning the true or false questions.

1. **Question 1**

   - *Formal verification relies on mathematical models and logical analysis.*

     This question received 8 responses. The statement is true, so 50% of the participants answered correctly, and 50% did not know what to answer. It could mean insufficient information in the explanation.

   - *Formal verification relies on random testing techniques.*

     The statement is false. We received 8 replies to this question, where no one gave the wrong answer. 62.50% of the participants responded correctly, and 37.50% answered that they *"Don't know"*.

     Although this description does not mention the formal verification methods, there were no wrong answers. In total, 9 participants got the answer right. Of these 9, only 3 said they were familiar with the formal verification concept before the study.

2. **Question 2**

   - *Formal verification can be used to verify the compliance of a system with a given set of requirements or specifications.*

     We had 4 participants responding to this question. All the participants said the affirmation was true, so 100% responded correctly.

   - *Formal verification can not be used to verify the compliance of a system with a given set of requirements or specifications.*

     There were 12 answers to this statement. Of these participants, 50% responded correctly, saying the statement was false, and just 1 answered wrong, saying it was true. 41.67%, corresponding to 5 responses, were the *"Don't know"* option.

Of the 16 participants, 10 answered this question correctly, and only 1 gave the wrong answer, so we could conclude that the description was enough to answer the question. Of the 10 participants who got the answer right, 3 said that the concept of formal verification was familiar to them before the study, and 1 responded that they were unsure. Although in the description it is not directly written that *"Formal verification can be used to verify the compliance of a system with a certain set of requirements or specifications."*, this understanding can be inferred from this part of the description *"It is more thorough than other methods, meticulously analyzing all possible system states guaranteeing its behavior."*.

3. **Question 3**

    - *Formal verification requires expert knowledge of formal methods to apply formal methods effectively.*

      This statement received a total of 7 replies. In this instance, 4 respondents said the statement was true, 1 said it was false, and 2 did not know the response. More than half answer correctly by saying that the affirmation is true. It was 57,14% of the participants.

    - *Formal verification does not require expert knowledge of formal methods to apply formal methods effectively.*

      This question received 9 answers. The more significant percentage went to the *"Don't know"* answer, with 44.44% corresponding to 4 participants. 3 (33.33%) participants responded false, hitting the question, and 2 gave the wrong answer (*"true"*).

    The description says nothing about the level of knowledge that is required to apply formal methods. Hence, most participants did not select the correct option. 6 of the 16 participants said they did not know the answer, and 3 were wrong. It corresponds to 56.25% of the participants.

4. **Question 4**

    - *Formal verification can be used to guarantee the absence of errors in specific contexts.*

      We got 9 responses in total. Of the 9, 4 are correct with the *"True"* option selected, and 3 are incorrect. Also, 2 participants answered that they did not know how to answer the question.

    - *Formal verification can not be used to guarantee the absence of errors in specific contexts.*

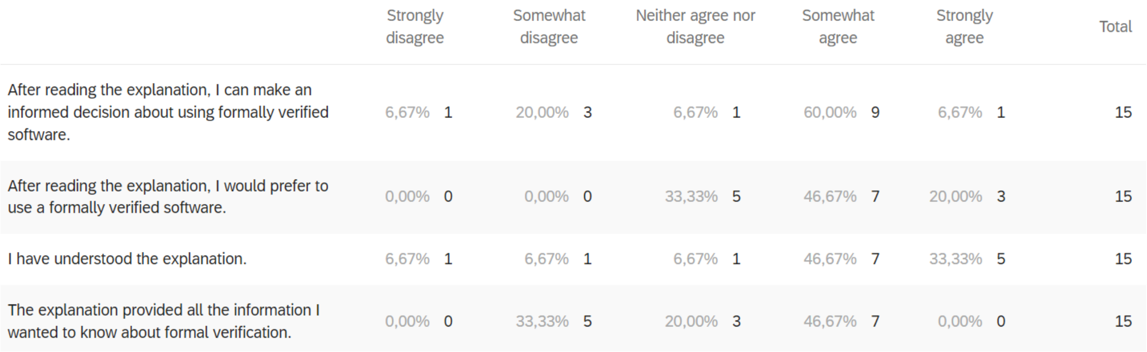
      We had 7 participants respond to this question. 2 got it right, saying the statement was false, and 2 got it wrong when selecting the *"True"* option. 42.86% (3 participants) selected the *"Don't know"* option.

The results for this question were balanced between the three answer options, demonstrating that the description does not give enough information to know whether the formal verification guarantees the absence of errors.

Concerning the feedback from the participants gathered in the open questions, the positive aspect they highlighted was that the description was concise. However, most of the participants considered that the description had missing points. A participant suggested adding a scenario to the description. He said, *"Add final recommendation about when/which scenarios is worth to use these methods (i.e. benefits >costs)"*, (P21), which should be considered in the future formal verification descriptions. Another participant suggested an improvement in the last sentence of the description. He stated, *"I don't understand the last sentence. However, it (formal verification) can be challenging, (...) increasing the difficulty of applying it (formal verification). If I understood correctly, I would write "However, it can be challenging, time-consuming, and expensive, which makes it hard to apply effectively.""*, (P31).

#### 3.3.2.3 Questionnaire with Description C

Most participants chose "somewhat agree" and "strongly agree" to describe their level of agreement with the statements presented after reading the formal verification description. Regarding making an informed choice about using formally verified software, most participants (66.67%) felt that the description was sufficient to allow an informed choice, while 26.67% felt it was not. Regarding making an informed choice about using formally verified software, most participants (66.67%) felt that the description was sufficient to allow an informed choice, while 26.67% felt it was not. After reading the description, most respondents prefer formally verified software (66.67%). The remaining 33.33% neither agree nor disagree. 80% of the participants understood the explanation, and only 13.34 did not. Finally, 46.67% of the participants considered that the explanation provided all the necessary information. 33.33% thought not. We had, in total, 10 "Neither agree nor disagree" answers to the four statements (Figure 3.12).



| | Strongly disagree | | Somewhat disagree | | Neither agree nor disagree | | Somewhat agree | | Strongly agree | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| After reading the explanation, I can make an informed decision about using formally verified software. | 6,67% | 1 | 20,00% | 3 | 6,67% | 1 | 60,00% | 9 | 6,67% | 1 | 15 |
| After reading the explanation, I would prefer to use a formally verified software. | 0,00% | 0 | 0,00% | 0 | 33,33% | 5 | 46,67% | 7 | 20,00% | 3 | 15 |
| I have understood the explanation. | 6,67% | 1 | 6,67% | 1 | 6,67% | 1 | 46,67% | 7 | 33,33% | 5 | 15 |
| The explanation provided all the information I wanted to know about formal verification. | 0,00% | 0 | 33,33% | 5 | 20,00% | 3 | 46,67% | 7 | 0,00% | 0 | 15 |

Figure 3.12: Questionnaire 3- Participants' subjective understanding

The results below concern the true or false questions of the questionnaire.

1. **Question 1**

- *Formal verification relies on mathematical models and logical analysis.*

  There were a total of 7 answers, all correct, with the "true" option selected. Of the 7 participants who responded, 2 stated that the formal verification concept was familiar to them before the study, and another 2 said they were unsure. It may justify the totality of correct answers since nothing in the description says that formal verification is based on mathematical methods.

- *Formal verification relies on random testing techniques.*

  The statement is false. We received 8 replies to this question, where 2 participant gave the wrong answer. 62.50% of the participants (5 participants) responded correctly, and 1 said, *"Don't know"*.

2. **Question 2**

   - *Formal verification can be used to verify the compliance of a system with a given set of requirements or specifications.*

     The sentence is true. We got 11 answers to this question, of which 81.82% are correct. There were no wrong answers. 2 of the participants did not know the answer.

   - *Formal verification can not be used to verify the compliance of a system with a given set of requirements or specifications.*

     This sentence is false. There were a total of 4 participants answering this question. All participants correctly answered this question.

   The results for this question showed that this description, centered on **security**, communicates well that a system's adherence to a given set of criteria or specifications can be confirmed through formal verification.

3. **Question 3**

   - *Formal verification requires expert knowledge of formal methods to apply formal methods effectively.*

     There were 8 responses. Of these 8, 6 were correct (*"True"*), and 2 *"Don't know"*.

   - *Formal verification does not require expert knowledge of formal methods to apply formal methods effectively.*

     7 participants responded in total. Since the sentence was false, 3 answered correctly, 2 chose wrong, and 2 did not know the answer.

4. **Question 4**

   - *Formal verification can be used to guarantee the absence of errors in specific contexts.*

     We had 4 responses, of which 50% were *"True"* and 50% *"Don't know"*. Nobody chose the wrong option.

- *Formal verification can not be used to guarantee the absence of errors in specific contexts.*

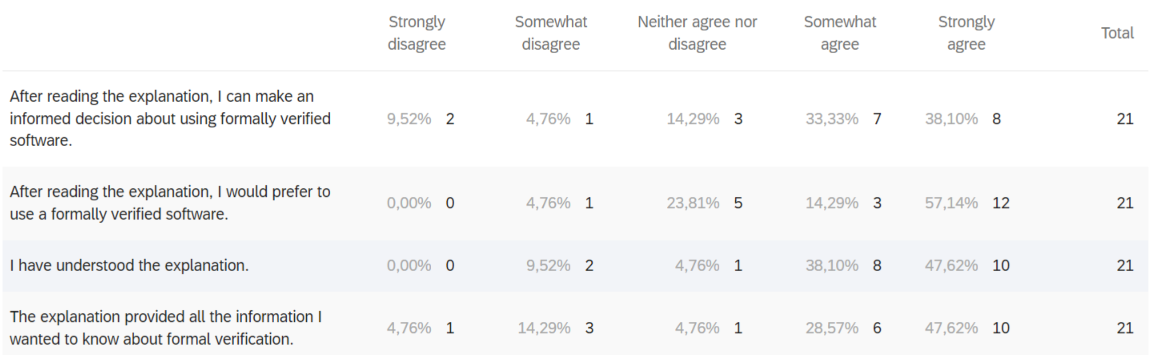
  There were a total of 11 responses. The majority (63.64%) did not know how to answer the question. Only 1 participant answered correctly, choosing the *"False"* option, and 3 chose the wrong alternative.

  The description needs more information to evaluate these statements since most of the percentage was to the answer "Don't know".

Considerable participants indicated that the description used in this questionnaire was elucidative and concise in the open-ended questions. However, several participants' suggestions indicated issues that could be added and clarified in the description. One participant said, *"The explanation could benefit from including concrete examples to illustrate the concepts of formal verification. Examples could help clarify how mathematical modeling, theorem proving, and model checking are applied in practice, making the explanation more relatable and tangible for readers."*, (P41). Another expressed, *"Maybe slightly explain the constraints or its disadvantages."*, (P44). A participant also asked, *"How is this applied to real-world systems? How big is the impact of the chosen verification methods?"*, (P47).

#### 3.3.2.4 Questionnaire with Description D

The options *"Somewhat agree"* and *"Strongly agree"* were the most chosen. Concerning making an informed choice about using formally verified software, 71.73% of participants agreed that the description was sufficient to allow an informed choice, while 14.28% did not agree. Regarding the statement, *"After reading the explanation, I would prefer to use a formally verified software."*, 71.43% prefer formally verified software, and 4.76% responded: *"Somewhat disagree"*. 85.72% of the participants understood the explanation, and 9.52% did not. Ultimately, 76.19% of the participants agreed that the explanation provided all the information they wanted, and 19.05% thought not. We had, in total, 10 *"Neither agree nor disagree"* answers to the four statements (Figure 3.13).



| | Strongly disagree | | Somewhat disagree | | Neither agree nor disagree | | Somewhat agree | | Strongly agree | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| After reading the explanation, I can make an informed decision about using formally verified software. | 9,52% | 2 | 4,76% | 1 | 14,29% | 3 | 33,33% | 7 | 38,10% | 8 | 21 |
| After reading the explanation, I would prefer to use a formally verified software. | 0,00% | 0 | 4,76% | 1 | 23,81% | 5 | 14,29% | 3 | 57,14% | 12 | 21 |
| I have understood the explanation. | 0,00% | 0 | 9,52% | 2 | 4,76% | 1 | 38,10% | 8 | 47,62% | 10 | 21 |
| The explanation provided all the information I wanted to know about formal verification. | 4,76% | 1 | 14,29% | 3 | 4,76% | 1 | 28,57% | 6 | 47,62% | 10 | 21 |

Figure 3.13: Questionnaire 4- Participants' subjective understanding

The true or false questions results are presented below.

1. **Question 1**

    - *Formal verification relies on mathematical models and logical analysis.*

      For this sentence, we had 13 participants respond. 5 got the answer right, saying the statement was true, and 1 got it wrong. Most participants did not know how to respond to the question (53.85%).

    - *Formal verification relies on random testing techniques.*

      We had 8 responses. 50% were correctly the *"False"* option. 2 answers were wrong, and 2 were *"Don't know"*.

    Regarding this first question, with the two statements, it was expected that nearly all participants would be able to answer correctly since the description begins with *"Formal verification relies on mathematical and logical techniques to analyze and verify a system."*. However, of the 21 responses to these two sentences, 9 were "Don't know", 9 were correct, and 3 were wrong. The percentage of participants not knowing how to respond was significantly high. It may show that the participants did not fully comprehend the formal verification description.

2. **Question 2**

    - *Formal verification can be used to verify the compliance of a system with a given set of requirements or specifications.*

      We had 11 answers, of which 8 were the *"True"* option, which is correct. 2 participants answered wrongly, and 1 did not know what to answer.

    - *Formal verification can not be used to verify the compliance of a system with a given set of requirements or specifications.*

      10 participants assessed this statement. 5 evaluated it as being false, having answered correctly. 2 gave the wrong answer, and 3 selected the *"Don't know"* option.

    Regarding these two sentences, most participants selected the correct option. Although this description addresses the topic of verification methods more directly, it mentions this at the end: *"to ensure system reliability and adherence to specifications"*, which may explain most of the correct answers.

3. **Question 3**

    - *Formal verification requires expert knowledge of formal methods to apply formal methods effectively.*

      We got 10 responses. 40% of the participants got the option right (*"True"*), and the remaining 60% did not know what to answer. There were no wrong answers.

- *Formal verification does not require expert knowledge of formal methods to apply formal methods effectively.*

  This sentence had 11 responses. 6 were correctly the *"False"* option, 2 were wrong, and 3 were the *"Don't know"* option.

  The answers to these two sentences reinforced that the explanation does not provide enough information to respond since it says nothing about the knowledge required to apply formal verification methods.

4. **Question 4**

   - *Formal verification can be used to guarantee the absence of errors in specific contexts.*

     We had 14 responses, of which 64.29% (9 answers) were *"True"* and 21.43% (3 answers) were *"False"*. The correct option is *"True"*. 2 participants did not know what to answer.

   - *Formal verification can not be used to guarantee the absence of errors in specific contexts.*

     It was a total of 7 responses to this sentence. Most participants selected the *"Don't know"* option. There were 4. The rest mistakenly selected the option *"True"*. None got the answer right.

   In this case, the participants' responses also demonstrated that the description does not have enough information to give an informed response to these two sentences.

Regarding the open-ended questions for this questionnaire, the participants considered the description concise. In addition, one participant mentioned that *"For some people (probably persons related to IT it can act as a trigger to know more about "Formal Verification""*, (P59), which was a relevant issue since it suggests that this explanation encourages the search for more information on the subject. However, we had several participants mention that it was an incomplete description. One of the participants raised some questions that he would like to see clarified in the description. He said that *"Yes. It lacks an explanation of what is formal verification. It seems that it can be great but without knowing what is actually being verified it is not possible to know. Is there any kind of system to determine what should be verified based on the type of system we want to verify? Are we just listing some verifications of the top of our heads and then running them? What is the method?"*, (P49). The same participant also suggested that *"Instead of such a big list of things that formal verification can prevent maybe explain how it is conducted. Explaining how the thought process works will let the reader get to what the benefits are. So I would give a wider view on the benefits and be more specific on the method."*, (P49).

### 3.3.3 Results overview

For description A, considering all the answers to the true or false questions, we had a total of 68. Of these 68, 41 were the correct option, 9 the wrong one, and 18 the "Don't know" answer (Figure 3.14).

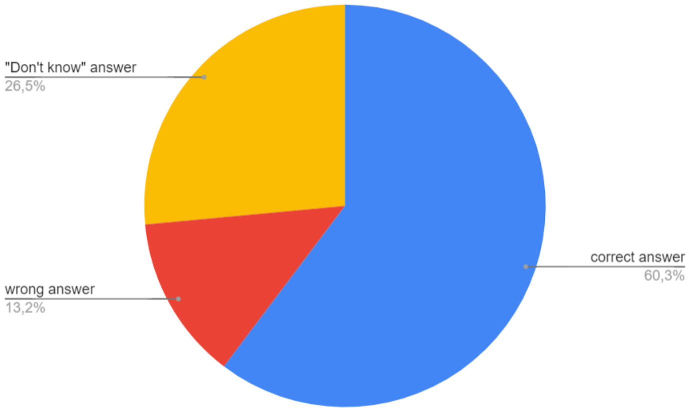


Figure 3.14: Questionnaire 1- Answers overview

For description B, considering all the answers to the true or false questions, we had a total of 64, where 32 were the correct option, 9 were the wrong one, and 23 were the "Don't know" answer (Figure 3.15).

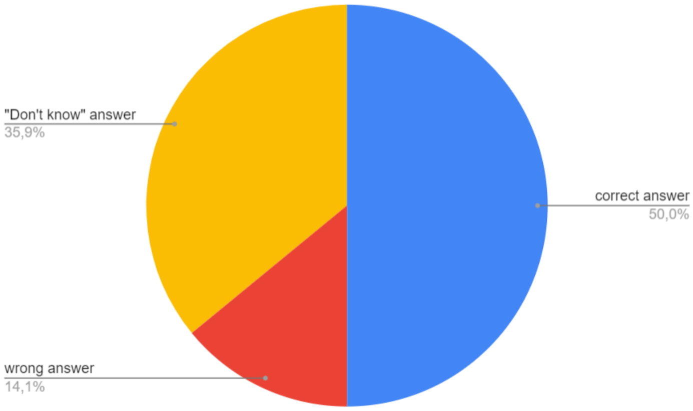


Figure 3.15: Questionnaire 2- Answers overview

For description C, we had a total of 60, where 37 were the correct option, 7 were the wrong one, and 16 were the "Don't know" answer (Figure 3.16).

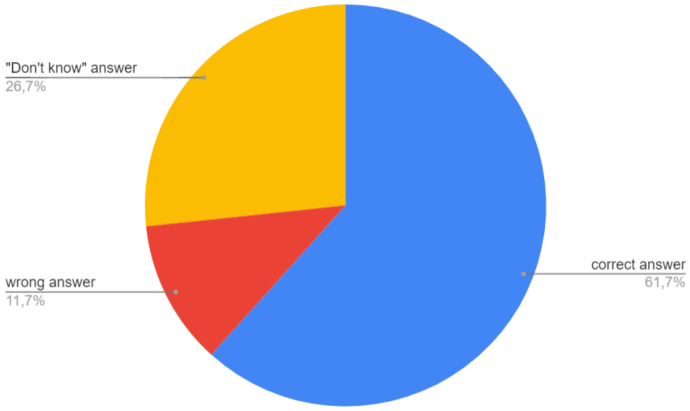


Figure 3.16: Questionnaire 3- Answers overview

For description D, we had a total of 84. Of these 84, 41 were the correct option, 15 the wrong one, and 28 the "Don't know" answer (Figure 3.17).

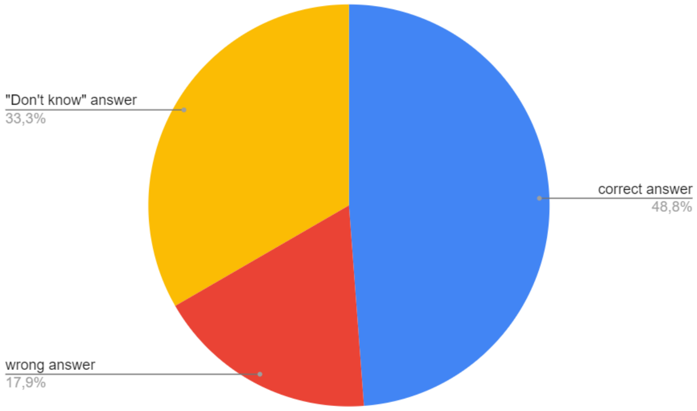


Figure 3.17: Questionnaire 4- Answers overview

In all the True/False questions of the four questionnaires, we verified a pattern in which the highest percentage corresponds to correct answers, followed by a percentage for the answers with the *"Don't know"* option, and finally, the lowest percentage goes to the wrong answers. In addition, the questionnaires with descriptions B and D, whose topics were **advantages/disadvantages** and **verification methods**, respectively, obtained a percentage of 50% or less correct answers and a more considerable percentage of "Don't know" answers. In the questionnaires with descriptions A and C, whose topics were **code properties** and **security**, respectively, more than 50% of the answers were correct. We could conclude that addressing these two themes is more effective in communicating the concept of formal verification.

## 3.4 Limitations

Our study had limitations, such as concerning questionnaires, which were consistently associated with biases because of the questions' design or order. The words used in the questions could introduce biases. However, we did our best to reduce the bias as much as possible. We randomized the order of the statements for the question used to evaluate the subjective comprehension of the participants. All sentences had two versions for the true/false questions, one true and one false. We also randomized the exhibition of these versions. We performed two cognitive interviews to detect possible errors and biases in the questionnaire.

Another point was the size of the study, as it was carried out only in the personal network, which is a relatively small sample. Nevertheless, it was possible to draw relevant conclusions too.

## 3.5 Threats to validity

Only one person did the description selection process, which could be biased.

In the thematic analysis, only two coders coded the descriptions and the participants' responses, which may not be ideal. We would have more reliability using multiple coders.

Although we did not conclude a direct relationship between participants familiar with the concept and correct answers, the fact that we had participants familiar with the concept of formal verification may introduce bias to the results.

## 3.6 Conclusions

This study gave us relevant insights to answer the research questions.

Regarding how formal verification is currently described (RQ1), when we code the collected descriptions, we verified that they are currently based a lot on code properties. This theme appears in more than half of the sample we analyzed. The descriptions also cover the mathematical methods used by formal verification and mention the guarantees of the correctness of the system. We found that many descriptions may be too complex for non-experts users. Some current descriptions require background knowledge or understanding of other technical concepts used in the formal verification descriptions.

The performed user study allowed us to answer the third research question, *"How can we better communicate the concept of formal verification?"* (RQ2). With the user study, we concluded that the most compelling descriptions in communicating the formal verification concept addressed the theme of code properties and security. Thus, both of these themes should be included in the descriptions of this concept. In addition, many participants mentioned that the descriptions lacked practical examples, so adding use cases to the explanations can be an asset for a better understanding of the concept.

The descriptions that better communicate the concept of formal verification to people with no background in this domain (RQ3) were the description of the code properties theme and the security theme since, by the performed user study, we saw that they were the most well-comprehended descriptions by the participants. These descriptions had more correct answers in the questionnaires.

# Chapter 4

# Final considerations

We saw that in the existing literature, we could find several different descriptions of the formal verification concept. However, there were no human-centered studies on better communicating the concept of formal verification and the formal guarantees it offers to users without knowledge in that area.

Therefore, we carried out a user study about formal verification. We consider performing this study vital to ensure that people fully understand the concept of formal verification and its guarantees to avoid these users exposing themselves to unnecessary risks when using unknown software.

Regarding this investigation, we better understood how to communicate the concept of formal verification and its guarantees to non-expert users to improve their understanding. We have seen that non-experts best understand formal verification descriptions associated with security and code properties. As many participants suggested, adding practical examples could be an asset to the descriptions.

## 4.1 Future work

In our study, we created an explanation of the main themes used in formal verification descriptions. In the literature, an explanation of the formal verification concept can address more than one theme. We propose as future work the realization of studies to evaluate the effectiveness of communicating the concept of formal verification with different combinations of themes in the explanations.

With the completion of these studies, we can increasingly understand how to communicate the formal verification' concept to users so that they understand it well and know its guarantees. Thus, users will be able to make more informed decisions when choosing between software that uses formal verification and software that is not formally verified.

# References

[1] Custom Search JSON API: Introduction | Programmable Search Engine | Google Developers.

[2] Internet Archive: Digital Library of Free & Borrowable Books, Movies, Music & Wayback Machine.

[3] Local SEO Tools, Local Citation Finder and Local Rank Checker & Tracker by GeoRanker.

[4] SerpApi: Google Search API.

[5] Web Search API.

[6] Christiane Atzmüller and Peter M. Steiner. Experimental Vignette Studies in Survey Research. *Methodology*, 6(3):128–138, January 2010.

[7] Jeremy Avigad and John Harrison. Formally verified mathematics. *Commun. ACM*, 57(4):66–75, apr 2014.

[8] Wei Bai, Michael Pearson, Patrick Gage Kelley, and Michelle L. Mazurek. Improving non-experts' understanding of end-to-end encryption: An exploratory study. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*, pages 210–219, 2020.

[9] Frank Berntsen. Formal verification 27.03.

[10] Per Bjesse. What is formal verification? *SIGDA Newsl.*, 35(24):1–es, dec 2005.

[11] Matthew L. Bolton, Ellen J. Bass, and Radu I. Siminiceanu. Using formal verification to evaluate human-automation interaction: A review. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(3):488–503, 2013.

[12] Virginia Braun and Victoria Clarke. Thematic analysis. In Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher, editors, *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, pages 57–71. American Psychological Association, Washington, 2012.

[13] Carolina Carreira, João F. Ferreira, Alexandra Mendes, and Nicolas Christin. Exploring usable security to improve the impact of formal verification: A research agenda. *Electronic Proceedings in Theoretical Computer Science*, 349:77–84, nov 2021.

[14] Marco Chiappetta. MIT Researchers Create Crash-Tolerant File System Guaranteed Not To Lose Data. *Forbes*, 2015.

[15] Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. "i need a better description": An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, page 3037–3052, New York, NY, USA, 2021. Association for Computing Machinery.

[16] Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. Am i private and if so, how many? In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. ACM, nov 2022.

[17] Maximilian Hailer. New Approaches and Visualization for Verification Coverage. 2022.

[18] Aditya Harbola, Deepti Negi, and Deepak Harbola. Infinite automata and formal verification. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2012.

[19] Kevin Hartnett. Hacker-proof code confirmed. *Quanta Magazine*, 2016.

[20] Jonathan Lazar. *Research methods in human computer interaction*. Elsevier, Cambridge, MA, 2nd edition edition, 2017.

[21] István Majzik. Formal verification.

[22] Microsoft. Web Search API | Microsoft Bing.

[23] Yannick Moy, Emmanuel Ledinot, Hervé Delseny, Virginie Wiels, and Benjamin Monate. Testing or formal verification: Do-178c alternatives and industrial experience. *IEEE Software*, 30(3):50–57, 2013.

[24] Budi Rahardjo. Formal methods in hardware (circuit) design. 2004.

[25] Elissa M. Redmiles, Yasemin Gülsüm Acar, Sascha Fahl, and Michelle L. Mazurek. A summary of survey methodology best practices for security and privacy researchers. 2017.

[26] Erik Seligman, Tom Schubert, and M V Achutha Kiran Kumar. Chapter 1 - formal verification: From dreams to reality. In Erik Seligman, Tom Schubert, and M V Achutha Kiran Kumar, editors, *Formal Verification*, pages 1–22. Morgan Kaufmann, Boston, 2015.

[27] B. H. Thacker, S. W. Doebling, F. M. Hemez, M. C. Anderson, J. E. Pepin, and E. A. Rodriguez. Concepts of model verification and validation (la–14167). Technical report, United States, October 2004.

[28] Susana M. Vieira, Uzay Kaymak, and João M. C. Sousa. Cohen's kappa coefficient as a performance measure for feature selection. In *International Conference on Fuzzy Systems*, pages 1–8, 2010.

[29] Natalie Wolchover. In computers we trust? *Quanta Magazine*, 2013.

[30] Natalie Wolchover. Introduction to formal verification. *EEWeb*, 2019.

## Appendix A

# User study - Questionnaire

# User study about Formal Verification

## Privacy Policy

Q1 Thank you for your interest in participating in this survey. Below you find information about this research project, conditions for participation, and handling of the collected data. Please read everything carefully. If you agree and wish to participate in this study, please confirm your consent below.  **General information about the research project:**  In this study, you will be asked to complete a questionnaire about formal verification and how to explain it to a wider audience. We will ask questions related with how to explain formal verification. We will also ask you to provide a few demographic data and background information. The survey will take approximately 10 minutes to complete. Please note that you must be at least 18 years old to participate in this study. No particular burdens or damages from participating in this research project are to be expected.  **Voluntariness:**  Your participation in this research project is voluntary. You can revoke your consent to participate at any time, without providing reasons, and without any disadvantages.  **Privacy and anonymity:**  No personal identifiable data will be collected. The data collected in the context of this research project are exclusively assessed to investigate the statistical effects of demographic data on our research question. Analyses will be based on group statistics. Individual measurements are not relevant to our research questions. In the open-ended answers, please refrain from providing any personally identifiable information or sensitive data about yourself or others. This survey is designed to maintain your anonymity, and any responses containing such information will be removed or anonymized.  **Use of data:** The results of this study may be published for teaching or research purposes (e.g., theses, scientific publications, or conference proceedings). This will be done in an anonymous form, i.e. without the data being attributable to a specific person.  **Responsible management of the research project:**  If you have any questions regarding the research project or if you wish to make use of your right of revocation, please contact:
Carolina Carreira  *carolina.carreira@tecnico.ulisboa.pt*
Mariana Soares  up201605775@g.uporto.pt

○ I agree with the terms above and I am willing to voluntarily participate.

## Formal verification description

Please, read carefully the formal verification description below.

**Formal verification focuses on verifying the correctness of a system, this is, the system will always meet its requirements. Formal verification helps prevent errors and promotes robust software development, thus ensuring the system is predictable.**

Q1
If needed read the description again.

Pay attention **we will ask questions about this description in the next sections.**

Have you read the description carefully?

○ Yes

Understanding

Q2 How much do you agree with the following statements?

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The explanation provided all the information I wanted to know about formal verification. | ○ | ○ | ○ | ○ | ○ |
| After reading the explanation, I can make an informed decision about using formally verified software. | ○ | ○ | ○ | ○ | ○ |
| I have understood the explanation. | ○ | ○ | ○ | ○ | ○ |
| After reading the explanation, I would prefer to use a formally verified software. | ○ | ○ | ○ | ○ | ○ |

Objective Understanding #A

AT Formal verification relies on mathematical models and logical analysis.

○ True

○ False

○ Don't know

AF Formal verification relies on random testing techniques.

○ True

○ False

○ Don't know

Objective Understanding #B

BT Formal verification **can** be used to verify the compliance of a system with a given set of requirements or specifications.

○ True

○ False

○ Don't know

BF Formal verification can **not** be used to verify the compliance of a system with a given set of requirements or specifications.

○ True

○ False

○ Don't know

Objective Understanding #C

CT Formal verification requires expert knowledge of formal methods to apply formal methods effectively.

○ True

○ False

○ Don't know

CF Formal verification does **not** require expert knowledge of formal methods to apply formal methods effectively.

○ True

○ False

○ Don't know

Objective Understanding #D

D1 Formal verification can be used to guarantee the absence of errors in **specific contexts**.

○ True

○ False

○ Don't know

D2 Formal verification can **not** be used to guarantee the absence of errors in specific contexts.

○ True

○ False

○ Don't know

Strengths and Weaknesses

The description you saw was:

**Formal verification focuses on verifying the correctness of a system, this is, the system will always meet its requirements. Formal verification helps prevent errors and promotes robust software development, thus ensuring the system is predictable.**

The next questions are about this specific explanation and not about the concept of "Formal Verification" in general.
*Pode responder em português se preferir.*

Q3 Are there any strengths or positive aspects of this explanation? If so, please elaborate.

________________________________________________________________

________________________________________________________________

________________________________________________________________

________________________________________________________________

________________________________________________________________

Q4 Are there any weaknesses or negative aspects of this explanation? If so, please elaborate on how to improve it.

________________________________________________________________

________________________________________________________________

Q5 If you have further feedback about this explanation, please share them below.

Demographic Information

Q6 Have you worked professionally in a field related to computers?

○ Yes, please specify what you did professionally

○ No

Q7 Before this study were you familiar with the concept of formal verification?

○ Yes

○ No

○ I'm not sure

Q8 How old are you?

○ Under 18

○ 18-24 years old

○ 25-34 years old

○ 35-44 years old

○ 45-54 years old

○ 55-64 years old

○ 65+ years old

Q9 How do you describe yourself?

○ Male

○ Female

○ Non-binary / third gender

○ Prefer to self-describe ________________________________________________

○ Prefer not to say

Q10 What is the highest level of education you have completed?

○ Some high school or less

○ High school diploma or GED

○ Some college, but no degree

○ Associates or technical degree

○ Bachelor’s degree

○ Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)

○ Prefer not to say

Q11 In which country do you currently reside?

▼ Afghanistan ... Zimbabwe

Q12 In which sector do you primarily work? Select all that apply:

□ Education

□ Industry

□ Academia

□ Government

□ Non-profit organization

□ Prefer not to answer

□ Other: ________________________________________________

Q13 Which job title best represents your responsibilities right now?

□ Programmer / Software Developer / Software Engineer

□ System Administrator / Network Engineer

□ Project Manager

□ Technical Lead / Team Leader

□ Researcher

□ Prefer not to answer

□ Other: ________________________________________________

Q14 If you have any additional feedback or suggestions, please feel free to share them in the box below.

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

## Appendix B

# Thematic Analysis - Descriptions' codebook

| Codes | Theme | Description | *In Vivo* |
|---|---|---|---|
| confidence/trust | advantages | A description suggests we can have a reliable system by applying formal verification. | "Once formal verification is done on the model, one can assert with the confidence afforded by formal proofs that the system as implemented and modeled preserves privacy in addition to knowing that the algorithms implemented by the system preserve privacy." [E22] |
| more thorough than other methods | advantages | A description that describes formal verification as being more thorough than other methods | "Formal verification — the process of using mathematical methods to “inspect” a program or smart contract across any number of inputs — is generally seen as the more concise, more comprehensive alternative to traditional testing for writing higher quality, more secure code. But in reality, formal verification is an open-ended and interactive process. Much like unit testing, developers must dynamically define and layer on formal specifications, iterating on their approach as their code and analyses evolve. Further, formal verification is only as effective as its specifications, which can be time consuming to write (and often come with a steep learning curve)." [E39] |
| correctness | code properties | A description that makes clear that the formal verification guarantees the correctness of the system. If a program or algorithm generates suitable results for the supplied inputs, it is said to be correct. | "formal verification algorithms for machine learning aim to formally prove or disprove desired properties of machine learning models, including safety, fault tolerance, fairness, robustness, and correctness." [E42] |
| bugs | code properties | A description that says that formal verification is used to detect bugs in a system. | "Formal verification (FV) ensures that mission-essential software is free from disruptive errors and security vulnerabilities, but requires human experts that can be quickly overwhelmed by the increasing number, size, and complexity of software systems." [E41] |
| predictability | code properties | A formal verification description suggests that we can predict the behavior of a system. | "Furthermore, although formal verication can assist in ensuring behavioral predictability, it is known to be time-consuming." [E24] |
| properties/specification | code properties | A description defines that the formal verification guarantees a system satisfies the desired properties/rules or fulfills determined specifications. This code only applies to descriptions that talks about generic properties (e.g. "safety properties" would be the code "safety" and not this one) | "In other words, verification consists in verifying the satisfaction of a set of properties" [E17] |
| safety | code properties | A formal verification description says the system is safe when formally verified. In computer science, "safety" typically refers to a system's or program's capacity to function without unintended or harmful consequences. | "Formal verification guarantees that a model is safe w.r.t. a safety property. The remaining task is to validate whether those models are adequate, so that the verification results transfer to the system implementation." [E3] |

| | | | |
|---|---|---|---|
| liveness | code properties | A formal verification description says formal verification can ensure liveness | "In formal verification, the requirements are formulated as a logical formula. A theorem prover then creates a mathematical proof showing that all possible executions—usually infinitely many—of the model are correct (safety proof), or showing that the model has a way to achieve a goal (liveness proof). The mathematical proof is the correctness certificate." [E7] |
| complexity | disadvantages | A description that states that formal verification has the disadvantage of being very complex, particularly in sophisticated and large systems | "Ideally, formal verification enables fully automatic proofs of system properties under any input conditions. Practically, its application so far has been quite limited due to the complexity of the associated computation. Computations in formal verification are often linear (or low polynomial) in the number of states of the system, but that number is often too large." [E45] |
| expensive | disadvantages | A description that states that formal verification has the disadvantage of being very expensive | "There are three main components involving in the formal verification of a hardware or software system: • A formal model of the system. Models used in formal verification vary in the level of abstraction, from an automaton describing status changes of the system, to source code or machine code of the system. • A formal specification, often described in a formal languages. These formal languages also have different power of expressiveness. • A formal method, implemented in a fully or partially automated tool, to prove or disprove the conformance of the formal model to the formal specification." [E28] |
| vulnerabilities | security | A description of formal verification says that it turns the system less vulnerable. | Formal verification (FV) ensures that mission-essential software is free from disruptive errors and security vulnerabilities, but requires human experts that can be quickly overwhelmed by the increasing number, size, and complexity of software systems. [E41] |
| privacy | security | A description explains that formal verification preserves the privacy of a system. | "[...]formal verification methods to ensure that personal information is not leaked due to mistakes or carelessness. The ability to verify subtle algorithms should be coupled with the ability to infer most of the proofs of correctness to reduce the programmer burden during the development and subsequent maintenance of a privacy-preserving code base." [E32] |
| mathematical methods | verification methods | A description clarifies that formal verification uses mathematical methods to check the system's behavior. | "A systematic process that uses mathematical reasoning and mathematical proofs (i.e., formal methods in mathematics) to verify that the system satisfies its desired properties, behavior, or specification (i.e., the system implementation is a faithful representation of the design)." [E34] |

|  |  |  |  |
|---|---|---|---|
| verification methods | verification methods | A description clarifies that formal verification uses algorithms to check if the system has the desired behavior. | "Model checking is a powerful method widely explored in formal verification. Given a model of a system, e.g., a Kripke structure, and a formula specifying its expected behaviour, one can verify whether the system meets the behaviour by checking the formula against the model." [E76] |
| abstraction | verification methods | A description clarifies that formal verification uses abstractions to check if the system has the desired behavior. | "There are two approaches to formal verification of separation kernels at the implementation level:: theorem proving the implementation model by abstraction from source/binary code, and software model checking" [E9] |

## Appendix C

# Thematic Analysis - Participants answers' codebook

| codes | tone | codes' explanation | *in vivo* |
|---|---|---|---|
| subjective | negative | The participant stated that the explanation had a subjective interpretation. | "a bit abstract and with a subjective interpretation" [P15] |
| incomplete | negative | The participant stated that the explanation was incomplete, may even enumerate missing points in the explanation. | "Feels abstract and incomplete just highlights the high level aspects and not the deep reasons" [P12] |
| doubt | negative | The explanation left doubts in the participants or doubts come up. | "If someone is not well versed in Formal Verification may be left with some questions." [P69] |
| abstract | negative | The participant felt the explanations was too abstract/technical/formal. | "Feels abstract and incomplete just highlights the high level aspects and not the deep reasons" [P24] |
| summary of the explanation | neutral | The participant described the theme of the explanations. Sometimes as a positive point. | "The explanation accurately describes formal verification as a process that relies on mathematical and logical techniques. It emphasizes the goal of ensuring system reliability and adherence to specifications, which are key objectives of formal verification." [P41] |
| suggestions | neutral | The participant gave suggestions about the description. | "Perhaps adding a simple example would complement it quite well without adding too much information." [P48] |
| elucidative | positive | The participant felt the explanation was elucidative, clear, or easy to understand. | "It's clear and easy to understand" [P63] |
| curiosity | positive | The participant stated that the explanations created interest and curiosity in the subject. | "For some people (probably persons related to IT it can act as a trigger to know more about "Formal Verification"" [P59] |
| concise | positive | The participant said the explanation was short/concise (in a good way). | "It's a brief and to the point explanation on formal verification." [P69] |
| good | positive | The participant stated he liked the explanation. It was a good explanation. | "Numa primeira impressão, penso ser suficiente." [P43] |