**REVIEW PROTOCOL**

**1. SMS’S OBJECTIVE**

The objective of this systematic mapping study is to summarize the state-of-the-art on the verification of intelligent systems based on formal methods.

**2. RESEARCH QUESTIONS**

This study aims to answer the following research questions (RQ):

* What formal approaches are used (or most common) in the verification of intelligent systems? (RQ1)
* What are the benefits of using formal methods based techniques in the verification of intelligent systems? (RQ2)
* What are the challenges of verification? (RQ3)
* How is 'responsible AI' defined across articles, and what are the key similarities and differences in these definitions? (RQ4)
* What are the trends for the future? (RQ5)

**3. SEARCH STRATEGY**

The search strategy consists in determining what databases, keywords and inclusion/exclusion criteria will be used to find and filter studies. The developed strategy is later used during the selection procedure.

**3.1. DATA SOURCES**

* IEEE Xplore;
* ACM Digital Library;
* SpringerLink;
* Science Direct;

**3.2. SEARCH TERMS**

The keywords chosen for this study are Formal Methods, Responsible, and AI. We have opted to use the term “Formal Methods” exclusively, as we aim to focus on formal methods as a whole, rather than considering alternative terms. The term “Responsible” is deliberately broad, as it encompasses the overarching concept of explainable artificial intelligence (XAI). The goal is to enable humans to understand how AI decisions and predictions are made, which is crucial in fields such as healthcare, finance, and law, where AI decisions can have significant consequences and, therefore, must be justifiable. The term “AI” was chosen for its broad scope, as it is central to the field of artificial intelligence and its applications.

**3.3 INCLUSION AND EXCLUSION CRITERIA**

This study will consider the following criteria for selecting the studies.

* **Inclusion criteria**
  + Studies published from 2019 to 2024;
  + Studies written in English;
  + Studies that we are able to get access to the full text;
  + Studies regarding verification of intelligent systems based on formal methods.
* **Exclusion criteria**
  + Short papers are not considered;
  + Studies published before 2019;
  + Studies written in other languages than English;
  + Studies not aligned to this study goal.

**4. SELECTION PROCEDURE**

The objective of this procedure is to identify and include studies that are specifically relevant to the research questions. This process helps ensure that the review focuses on studies that contribute meaningfully to the analysis of the topic, in an organized way. The selection procedure is based on the following steps:

1. The search string will be applied to the databases;
2. Results will be combined to remove duplicates;
3. Based, first on the Title, the inclusion/exclusion criteria will be applied.
4. The abstract will be analyzed to resolve doubts.
5. The introduction will be analyzed to filter the final results.

**5. DATA EXTRACTION**

**Table 1 - Data extraction table**

| **ID** | **DATA INFORMATION** |
| --- | --- |
| **D01** | DOI |
| **D02** | Title |
| **D03** | Publication year |
| **D04** | Author(s) |
| **D05** | Type: Journal/Conference |
| **D06** | What languages and techniques were used? (RQ1) |
| **D07** | What are the motivations to use formal methods? (RQ2) |
| **D08** | What are the identified challenges? (RQ3) |
| **D09** | What is the definition of “responsible AI”? (RQ4) |
| **D10** | What are the trends of validation with formal methods? (RQ5) |

**6. THREATS TO VALIDITY**

**Study Selection Validity**:

The validity of the study selection process is crucial to ensure that all relevant primary studies are identified and included. Potential threats to this validity include the adequacy of the search strategy, the comprehensiveness of the selected sources, and the systematic application of inclusion and exclusion criteria. To address these concerns, this study utilized well-established and widely recognized digital libraries, ensuring access to a broad range of publications. Additionally, the review was compared to other secondary studies to evaluate the consistency and coverage of the selected studies. All studies filtered during the final stages of the selection process were accessible, mitigating the risk of excluding relevant work due to unavailability of full texts. Furthermore, the exclusion of gray literature was a deliberate decision to maintain the focus on peer-reviewed, high-quality sources. The criteria for inclusion and exclusion were applied systematically to minimize potential biases and inconsistencies.

**Data Validity**:

Ensuring the validity of data is critical for the reliability of the study’s results. One important consideration is whether the sample size is sufficient for the findings to be considered robust and representative. Another is the selection of appropriate variables for data extraction, as incorrect choices could undermine the study’s conclusions. Additionally, data extraction bias presents a potential risk, as it may lead to the omission of relevant information or the inclusion of irrelevant data. To address these challenges, the study employed a broad search process across general-purpose search engines and digital libraries, ensuring comprehensive identification of relevant publication venues. Artificial intelligence tools were utilized to mitigate the risk of bias further and assist in the systematic extraction of relevant data. This technology facilitated consistency and minimized errors in the data collection process. Moreover, two additional researchers were available to review the process and provide feedback whenever uncertainties or doubts arose. This collaborative approach reinforced the rigor and reliability of the study, enhancing the credibility of its findings.

**Research Validity**:

The validity of the research process is essential to ensure that the study is reliable, repeatable, and aligned with its objectives. Key considerations include whether the chosen research method is appropriate for addressing the research questions, whether the research questions effectively capture the study’s goals, and whether researchers were sufficiently familiar with the field before conducting the review.To enhance the reliability and repeatability of the process, the study documented the review details in a structured protocol. The authors critically discussed the suitability of the selected research method—systematic mapping study (SMS)—in addressing the objectives of the study and research questions. These discussions ensured that the chosen approach was well-matched to the research goals. Additionally, the research questions were carefully analyzed to confirm that they holistically covered the intended objectives of the study. To further mitigate potential validity threats, artificial intelligence tools supported the research process, ensuring systematic execution and reducing the likelihood of human error. Moreover, before conducting this systematic mapping study, the lead researcher extensively engaged with foundational literature on secondary studies in software engineering and searched related work to familiarize with the field.

**7. SYNTHESIS OF EXTRACTED DATA**

.

In this step the data obtained will be analysed, aiming at identifying similarities, differences, and recurring patterns across the studies. The goal of this process is to address each research question outlined in the study, providing clear and evidence-based answers.