This document describes the per-question codes used in the ***"These results must be false": A usability evaluation of constant-time analysis tools*** paper. The format is the following:

```
{question_code}: {question_text}
    • {code_name}: {code_description}
    • {code_name}: {code_description}
    ...
```

Some answers were coded into different sub-categories. In these cases, we start by defining the category code, and give the codes associated with the elements of this category as follows:

```
{question_code}: {question_text}
    - [Category] {category_description}
        • {code_name}: {code_description}
        ...
```

Given the similarity of answers to some questions, multiple questions for the survey were coded under the same question code. In this case, they are listed together in the following codebook.

Codebook - Repair tasks

AbilityToSolve Were you able to solve the subtask? *{Here, we assess the participant's ability to find potential issues.}*

- **Yes** *The participant managed to conclude on the CTness of the code.*
- **No** *The participant did not manage to conclude on the CTness of the code.*

CommentOnSolving Were you able to solve the subtask? [Comment] *{This is an open question, where participants were invited to share any thoughts on how they solved the task.}*

- **ToolAnalysis** The participant used the tool. This is assumed to be the case as it was their initial instructions.
- **ManualAnalysis** The participant states that they resorted to manual analysis to conclude.
- **MixedAnalysis** The participant states that they relied partly on manual analysis to conclude.

IsTaskCT Was the example piece of code in the subtask constant-time with respect to the secret?

- **Yes** *The participant concludes that the code is already ct.*
- **No** *The participant concludes that the code is not ct.*
- **NotSure** *The participant manifests explicit confusion or doubts about their conclusion.*

AbilityToRepair Were you able to fix the timing attack vulnerability in the code example?

- **Yes** *The participant provided a fixed version of the code.*
- **No** *The participant did not reach a solution they are satisfied with.*
- **NA** *The question is irrelevant if the participant concluded the task was already CT.*

MajorIssue Please describe any issues that you may have encountered while using the tool for solving the subtask.
What about this task was hard to solve?
Which parts of solving this subtask were the biggest time drains?

— [**Familiarization**] *The participant discusses issues they had during their first contact with the tool.*

- **LackDocumentation** *Documentation does not allow for an end-to-end usage of the tool*
- **LackExamples** *Small practical examples would have helped but were not available.*

— [**Preprocessing**] *The participant discusses issues they had while preparing the code for analysis (e.g., specific build or secret designation).*

- **Annotation** *Annotations are difficult to get right or take too much time.*
- **Setup** *Using the tool requires additional setup, besides the initial installation provided by the authors.*
- **Wrapper** *Using the tool requires implementing a wrapper, which was a problem for the participant.*
- **AdditionalKnowledge** *Using the tool requires additional knowledge unrelated to CT programming skills (e.g. how to develop in another language).*

- **Runtime** *The participant expressed difficulties in running the tools the tool efficiently (e.g, struggle to run the tool without error).*

— [**ProblemResolution**] *The participant discusses issues they had when trying to identify the source of the leakage and fix it.*

- **FixingIssue** *Besides identifying the issue, fixing it while preserving the functional aspect of the code was difficult.*
- **OutputInterpretation** *The participant expressed difficulty or incapacity to interpret the output of the tool.*
- **OutputUnclear** *The output is inconsistent, or doesn't state clearly state whether the code is constant-time or not.*

— [**Reliability**] *The participant discusses trust issues and limitations they encountered during their analysis.*

- **Limitation** *The participant noticed clear limitations in how the tool works (e.g., soundness, completeness).*
- **TrustInTool** *The participant expressed doubt about the tool's results and preferred to trust their instincts/manual analysis.*
- **FalseNegative** *The participant thinks the tool gives false negative.*
- **FalsePositive** *The participant thinks the tool gives false positive.*

- **NoIssue** *No issue reported. This is the only answer which is mutually exclusive with all the others.*

AdditionalResources Please name (or even paste links to!) all resources that you used to solve this subtask.

- **OfficialDocumentation** *The participant used some documentation directly related to the tool (e.g., the tools' github page, or academic paper)*
- **ProvidedTutorial** *The participant used or tutorial or crafted examples.*

— [**ThirdPartyDocumlentation**]

- **CTProgramming** *The participant used documentation related to CT programming practices.*
- **CTVerification** *The participant used documentation related to CT verification tools, unrelated to the tool they were using.*
- **CDocumentation** *The participant used documentation related to C programming.*

AdditionalComment Please feel free to provide any further comments and experiences for using the tool to solve this subtask. You don't need to repeat comments you've already submitted. There will be a place to provide general feedback on the tool later on, please collect it elsewhere for the time being.

- *The answer to this open question was coded using all the code above, as participants mostly shared their various struggles and complaints on the tasks.*