# Instruction for using the COMPASS Framework

Navigating sociotechnical challenges of AI(-based) systems through evaluation.

## Introduction

The COMPASS framework enables AI development teams to critically evaluate both technical innovation potentials and societal impacts of AI systems to ensure a responsible and trustworthy AI system. COMPASS provides a roadmap for AI developers and auditors to navigate the complex landscape of AI system and its impact on society. This framework offers flexibility for AI practitioners to tailor an evaluation process to the industries' specific needs and priorities through a self-assessment process. It helps identify where to leverage necessary skills by making use of the SPATIAL design principles.

The COMPASS framework is intended for use by teams as a self-evaluation, but COMPASS may also be used by internal evaluators (e.g., managers, AI champions), or by external evaluators (e.g., auditors) to gain quick insight into a team's process and focus areas. It is designed as a framework for ongoing, iterative use.

## COMPASS stands for seven key components:

- **CONTEXT:** Defining and determining the context of AI system, including who is involved in its lifecycle, and who is affected by the system.
- **OPENNESS:** Ensuring AI systems and algorithms are transparent and understandable for all stakeholders, considering access and documentation.
- **MEASURES:** Iteratively developing mechanisms for evaluating AI systems so that they operate fairly and reliably.
- **PRIVACY POTENTIALS:** Safeguarding user privacy and data protection throughout the AI life cycle.
- **ACCOUNTABILITY:** Highlighting trustworthiness and holding AI systems and developers accountable for their actions.
- **SECURITY and SAFETY:** Implementing security measures and safety precautions throughout the AI life cycle to minimize the potential attacks.
- **SUSTAINABILITY:** Integrating mechanisms to maintain the reliability and performance of AI systems while providing sustainable and environmentally friendly AI solutions.

## Instruction for use

Using the COMPASS framework is straightforward. It consists of a set of questions that evaluators can go through in relation to their project developments. The projects can be defined as narrow or as broadly as is desired.

The value of COMPASS is related to the effort invested in responding to and reflecting on the questions. By following the steps, evaluators can effectively leverage the COMPASS framework to navigate the sociotechnical challenges of AI systems and promote responsible and trustworthy AI development and deployment.

***Step 1: Enter relevant information on the landing page***

Begin by filling out the required information on the landing page. These details will help when looking back at the evaluation at a later date.

***Step 2: Respond to Closed-ended Questions***

The closed-ended questions are designed to help you assess and understand various aspects of your AI system such as ethical implications and societal impacts. For each of the seven components, you can rate close-ended questions on a scale from 0 to 5, where 0 means not implemented at all, but relevant, and 5 means fully implemented. You can also grade with 'NA' which stands for 'Not applicable' and then this score will not be included in the calculation (see Step 3). It may be beneficial to add comments in the appropriate column to provide context on why a particular score was provided. For example, why does a particular question not apply to this situation?

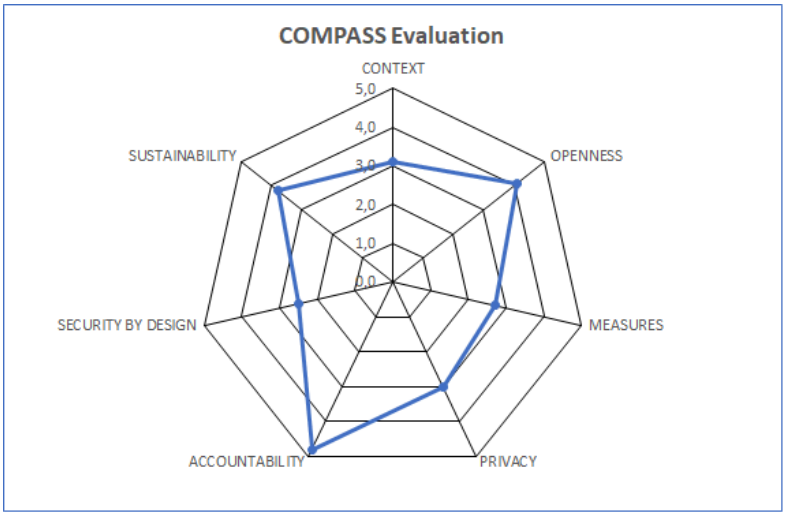
This set of questions is not prescriptive of where the AI system should be, as this may vary based on the goal of the AI system, the goal of the team, and the stage of the development process. **However,** the further the AI system is in its development, the higher the expectation is regarding higher scores.

Once you've completed the closed-ended questions, you can move on to interpreting the results, or you may first continue with Step 4.

***Step 3: Results – Radar Chart***

The closed-ended questions will result in a radar chart. This provides direct visual insight into the areas where the AI system scored well and where the scores were not as high. So, strong and weak points are determined by considering the results obtained because of the evaluation process. It may depend on the goal of the AI system, the goal of the team, and the stage of the development process, how high each component scores.

See the figure below for an example of a radar chart.




***Step 4: Open-ended questions***

Next, you can continue with the open-ended questions, which are provided in the same Excel file. This set of questions is not graded but should help the evaluator(s) understand where more extensive work is necessary. This part gives an opportunity to delve deeper into specific challenges or considerations related to AI systems.

It may be clear directly what directions a team could take to improve for each component. However, when this is less clear, it may be helpful to revisit the complete set of questions for the relevant component and see where more detail is needed and helpful.

This step ends with a set of resources to support teams in taking the next steps:

- SPATIAL guidelines and best practices for trustworthy AI. This guideline will provide the actions required to effectively address the challenges.
- SPATIAL design principles and patterns.
- Open-source resources.

***Step 5: Identify lessons learned and action points***

After completing the evaluation, it is helpful to identify where the AI system could improve and what topics the development team should focus on in next stages or future projects. On the landing page in the Excel file, there is space to fill these out.

First, there is a space to list the lessons learned and reflect on how the scores relate to the expectations. For example, if a component scores low, what is potentially the reason for this? And how does the score reflect expectations? Similarly, if a component scores

relatively high or higher than expected, you can critically reflect with the team why this is the case.

Second, the evaluator(s) and/or the development team can make a list of both short-term and long-term action points. These can be very specific, based on the questions of the COMPASS tool, or they can be extrapolated from the lessons learned.

## Final recommendations

1. Ensure that the Context component is well considered and has sufficient detail.
2. Be critical but constructive, try to consider points beyond the obvious. The more effort you invest in responding to the questions, the more relevant the evaluation outcomes will be.
3. Conduct the evaluation with a team, either at the same time or independently, and discuss the results together. An interdisciplinary team is most suitable for this.
4. COMPASS can be used in various stages of the process. The framework can be helpful if used iteratively, so at multiple moments in the development process.