



Technical Report, WP5

---

# An abbreviated assessment list to support the Responsible Development and Use of AI

---

Authors: V. Dignum, J.C. Nieves, A. Theodorou, A. Aler Tubella

Responsible Artificial Intelligence Group,  
Department of Computing Sciences, Umeå University

Design and typesetting by T. Sarayeva



UMEÅ UNIVERSITY

To ensure that requirements for Trustworthy AI are fully integrated in the development and use of AI systems, means to support the assessment of tools, development processes, applications and systems are needed. As part of our contribution to the AI4EU, we have developed an abbreviated assessment framework, based on the [Assessment List for Trustworthy AI \(ALTAI\)](#), developed by the [European High-Level Expert Group on AI](#).

This abbreviated assessment list is mainly meant to assess the AI applications shared through the AI4EU catalogue, but it also can support organisations perform a 'quick scan' of the AI-application they want to develop, procure, deploy, or use. This quick-scan list is a self-assessment tool to quickly identify the relevant elements of responsible AI and the level of adherence to these elements. It will help determine the level of impact of the AI applications and provide options to balance different tensions and interests. The assessment is based on a set of 15 questions grouped into 3 topics:

- Impact on Citizens and Society
- Technology
- Governance

All questions are answered with a simple 3-point scale, indicating to what extent the criteria is met. The assessment levels are (from high to low):

- (2)** the criteria are considered to be sufficiently and appropriately addressed, and evidence can be provided if necessary
- (1)** the criteria have been considered but not fully satisfied
- (0)** the criteria have not been addressed, or are considered not relevant to the application

Obviously, this method does not fully measure the 'quality' of the AI-based system but merely reflects the understanding and concern regarding its trustworthiness by the organisation sharing it in the platform. This quick-scan list should also not be taken as a complete in-depth assessment or by any means as evidence of legal compliance.

This abbreviated assessment can provide developers, auditors, and users a quick overview of the potential impact of an AI system but is not meant to replace the full assessment. We have used our experience with the evaluation of how [AI contributes to the Sustainable Development Goals \(SDGs\)](#) and [the assessment of covid19 tracing apps](#).

The abbreviated assessment list below is aligned with the 7 requirements for trustworthy AI as defined by the [European High-Level Expert Group on AI](#), as follows:

- 1 Human Agency and Oversight (questions 1, 5, 8, 9)
- 2 Technical Robustness and Safety (questions 7, 8, 10)
- 3 Privacy and Data Governance (questions 2, 6, 11)
- 4 Transparency (questions 3, 12, 13, 14)
- 5 Diversity, Non-discrimination, and Fairness (questions 1, 4)
- 6 Societal and Environmental Well-being (questions 5, 9, 14)
- 7 Accountability (questions 13, 15)

## Evaluation criteria

### Impact on users

#### **1. Inform on how it is respecting fundamental rights of individuals:**

How are you dealing with the effect of the application on the rights to safety, health, non-discrimination, and freedom of association?

- a. We've performed a clear analysis in response to these principles and can provide details. (2)
- b. We have partially/informally considered these principles but no specific details can be provided. (1)
- c. We have not considered these issues yet. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

**2. Privacy and data protection:** Is data collection compliant with the General Data Protection Regulation (GDPR) and does it respect the privacy of the individual? Note that A Data Protection Impact Assessment (DPIA) must be carried out before the deployment of any system.

- a. The purpose of the AI system and the mechanisms to assess its usage are clearly defined and compliant with GDPR, a DPIA has been performed and privacy of individuals is guaranteed. We can provide further information. (2)
- b. We have only done a partial/informal analysis and/or not all aspects of data and privacy protection are clear. (1)
- c. We cannot guarantee privacy and data protection. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

## Evaluation criteria

### Impact on users (continuation)

#### 3. **Transparency rights:** Do you include the right of users to:

- be notified that their data is being processed/collected,
  - access information on which personal data are collected,
  - control their own data,
  - access explanations of results produced by the system,
  - be informed of who, when and how the system can be audited.
- a. All of the above are fulfilled. (2)
  - b. Only some of the above are fulfilled or partially addressed. (1)
  - c. We cannot guarantee any transparency aspects. (0)
  - d. We consider that these issues are not applicable to our case. (N/A)

#### 4. **Accessibility:** Can your app/system/resource be used by all regardless of demographics, language, disability, digital literacy, and financial accessibility?

- a. This resource is fully accessible, and we can provide information on accessibility accommodations. (2)
- b. This resource partially complies with accessibility requirements. (1)
- c. This resource is not accessible to all. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

#### 5. **Education and tutorials:** Do you ensure that users are informed and capable of using the system correctly?

- a. We provide complete in-system help (2), or
- b. We provide support through external materials, e.g. website. (1)
- c. We do not provide user support. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

## Evaluation criteria

### Technology

**6. Data management:** Do you comply with the data-minimization principle, i.e. usage of local and temporary storage and encryption, based on principles of data protection by design? Do you ensure that only strictly necessary data are captured and processed?

- a. We use local and temporary storage and data encryption methods. We only collect and process strictly necessary data. We can provide further information. (2)
- b. We partially comply with the above, and some documentation can be provided. (1)
- c. We do not comply with these data management aspects. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

**7. Security:** Do you have user authentication in place to prevent risks such as access, modification, or disclosure of the data? Do you use unique and pseudo-random identifiers, renewed regularly and cryptographically strong?

- a. Strong security elements are in place, e.g. user authentication, unique identifiers regularly renewed. We can provide further information. (2)
- b. Some security features are in place. (1)
- c. No security features are in place. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

## Evaluation criteria

### Technology (continuation)

**8. Ease to deactivate/remove:** How easy is it to deactivate or remove the system and data once users are no longer interested or need the system?

- a. Very easy, either through clear instructions or automatically by the sunset clause. (2)
- b. Instructions on how to deactivate or remove the system and data are unclear. (1)
- c. There are no instructions or automated procedures to remove the system and the data. (0)
- d. We consider that these issues are not applicable to our case (N/A)

**9. Ease to access services without using the AI system:** In the case of AI systems aimed to replace or complement public services, are there full non-system alternatives?

- a. Yes, there is an easily accessible full non-system alternative. (2)
- b. There is a partial alternative or access to the full alternative is unclear. (1)
- c. There is no alternative to the AI system for this service. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

**10. Open-source code:** Is the development participatory and multidisciplinary?, What kind of access to the code and development is there?

- a. The code and development are open-source. (2)
- b. The code is open-source code without the possibility of contributing. (1)
- c. Non-open-source code. (0)
- d. We consider that these issues are not applicable to our case.



## Evaluation criteria

### Governance

11. **Ownership:** Is the ownership of the resource clear?
  - a. Ownership of the resource (including code, data, use) is clear and explicit. (2)
  - b. Some ownership aspects are made clear. (1)
  - c. Ownership information for the resource and related code or data is unavailable. (0)
  - d. We consider that these issues are not applicable to our case. (N/A)
  
12. **Openness over Data governance:** How open is Data governance?
  - a. Open data governance. (2)
  - b. Intermediate openness of data governance. (1)
  - c. Private/opaque settings. (0)
  - d. We consider that these issues are not applicable to our case. (N/A)
  
13. **Legislation and Policy:** Are there explicit legislation and/or other policies relevant to your system/resource?
  - a. The system is covered by an explicit clear, legal framework or sectorial formal policies, and we address these explicitly. (2)
  - b. We are aware of policy partially relevant to our system and address these sufficiently. (1)
  - c. We are not aware of any relevant legislation or policy and do not address these. (0)
  - d. We consider that these issues are not applicable to our case. (N/A)

## Evaluation criteria

### Governance (continuation)

#### 14. **Design Impact Assessment and Open Development Process:**

How explicit is the design process leading to this resource?

- a. Explicit information on the design process is available, including a clear description of aims and motivation, stakeholders, public consultation process and impact assessment. (2)
- b. Some information on the design process, aims and motivation, and impact assessment is available. (1)
- c. There is no information on the design and impact of the resource. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

#### 15. **Right to contest/liability:** Are users able to contest decisions/actions or demand human intervention?

- a. Processes for contesting and/or demanding human intervention are set up and clearly available. (2)
- b. Some contestability or intervention processes are available. (1)
- c. It is not possible to contest the system's output nor to demand human intervention. (0)
- d. We consider that these issues are not applicable to our case. (N/A)

## Examples (fictive)

