

Social Trust in AI Mapping

A simple set of criteria and steps can be used to self-assess social trust in AI systems which can be adapted to the specific area of geospatial mapping.

Background

The research team at The James Hutton Institute (JHI) has an interest in social aspects of AI responsibility and trustworthiness and has researched historical and current attempts at assessing trust in AI systems. This research provides an opportunity to test and improve on previously available guidance and adapt it to the context of geospatial mapping which is the specific focus of Trustable AI-generated Mapping project (TAiM) as shown in Figure 1 below.

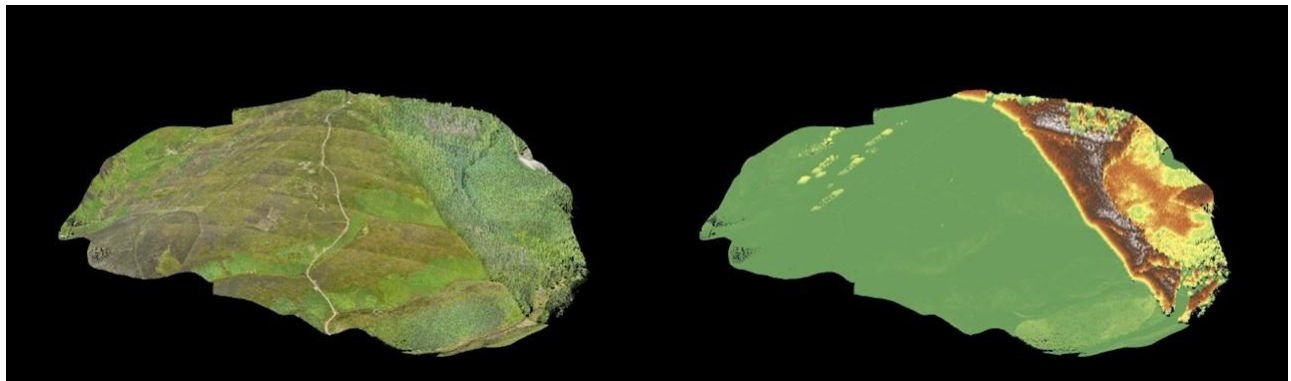


Figure 1: A rendering of the 3-D model generated from LiDAR point cloud and RGB orthomosaic data (left) captured from the survey of the Loch Hill area at Glensaugh, and a derived Height Above Ground model (right) illustrating tree height. Acknowledgement: James Hutton Institute ICMS team.

While there is a notable earlier history of ethical advances around AI developments, the concept of AI trustworthiness is relatively recent with its importance growing in the late 2010s, particularly following the 2019 Assessment List for Trustworthy Artificial Intelligence (ALTAI) developed by the EU High-Level Expert Group on AI (see Figure 2 below for further developments and framing with respect to geospatial mapping developments).

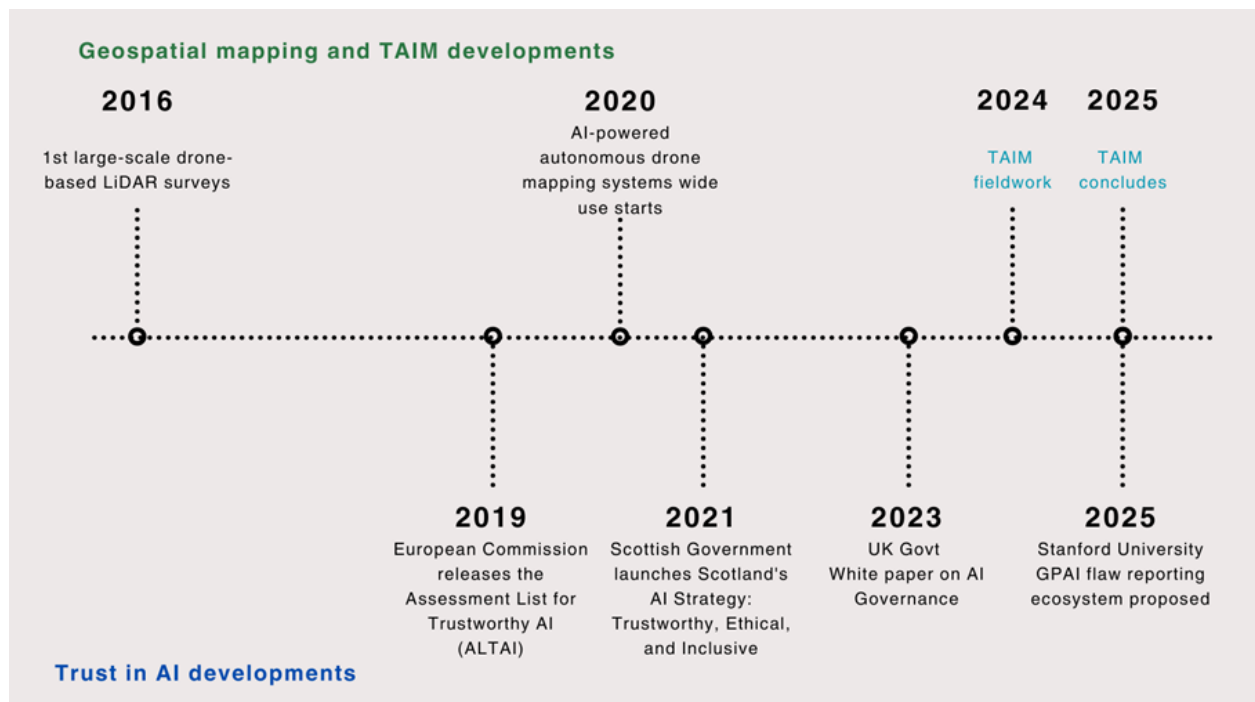


Figure 2: Timeline of AI trust and geospatial mapping developments.

Survey on trustworthiness of AI mapping

An internal questionnaire was conducted to gather expertise and views on aspects of trustworthiness and the stages at which the different ALTAI criteria are relevant – shown in Figure 3 below.

Questionnaire responses show that while the original criteria proposed is clearly relevant, some of the elements are not a priority in the context of geospatial mapping. Impact on work and skills, and Environmental wellbeing are the two items with the fewest 'Very important' responses, while General safety and Resilience to attack and security are the items considered of greatest importance.

AI Technical Robustness was considered important by most of the respondents, but they were mostly cautious about Accuracy. One respondent stated: "It should be accepted that AI systems will not be perfect, just as human decision making is not. What is more important to me is that we have clear guidelines on what to do if moderate or serious AI errors occur." It was also suggested that "being able to communicate levels of confidence is more important than levels of accuracy."

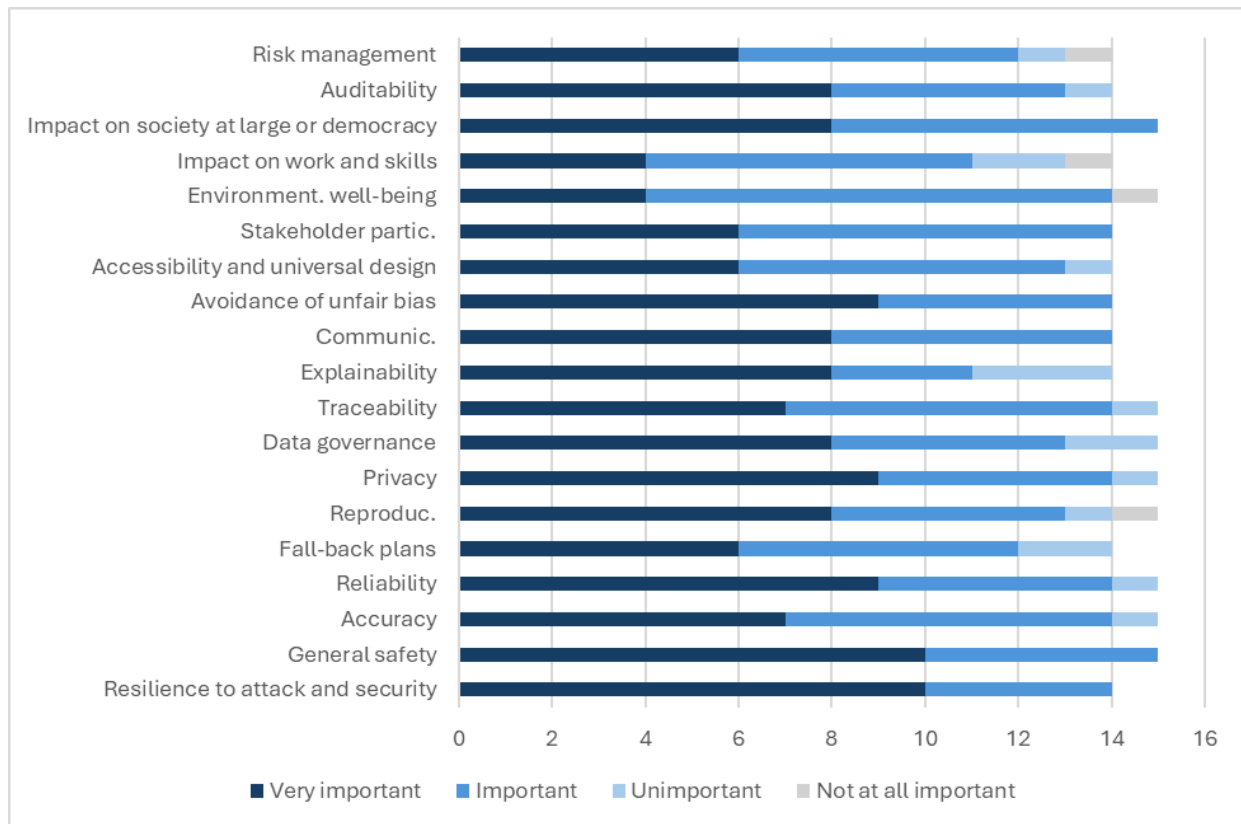


Figure 3: AI trustworthiness criteria according to importance rating by questionnaire respondents.

On the wider topic of Human agency and autonomy, respondents argued for a “human-in-the-loop” approach where a balance needs to be struck between long-term goals and targets set by humans and the specific requirements of the domain for AI to make decisions and take actions within clearly defined constraints but without human intervention.

While Privacy and Data Governance was considered important on the whole and relevant to building trust, it was highlighted that “most important is the protection of information that could be used to disadvantage vulnerable individuals or groups.” It was also pointed out that “much of the requirements are dealt with not in the AI itself but in the conditioning rules for use and in the curation and cleansing of input data sets (which may itself be an AI-supported process).”

Regarding Transparency, it was recognised that “understanding the process the AI uses to arrive at an output is key to trusting said output.” However, depending on the level of AI-expertise, responses varied: “I understand AI system usage and the process and so I’m coming at it from that perspective. I think as long as accuracy is high the other elements don’t matter so much. Communication is very important as once you understand AI and its limitations you feel less concerned about the other factors.”

Similarly, another respondent emphasised: “A critical need is for the generation of confidence measures in both the inputs to the system and the conclusions and outputs. That goes hand-in-hand with generating the understanding in users of how to engage with AI-supported systems.” It was also noted that “this is very important at the start, but likely to become less important once any AI system becomes more integrated into daily use and becomes a familiar component of the user's environment.”

On Diversity, Non-discrimination and Fairness, respondents recognised that the solutions have to be for the users and individuals must trust them, and they further argued that while “some geospatial applications are largely invariant to demographic bias, but many - particularly those using or inferring social/spatial data and outcomes, require rigorous pre-validation, in both input/training data and in any baseline assumptions made in the design of the system.”

The team concluded that the TAIM self-assessment web guidance tool would benefit from adapted criteria to the geospatial mapping, dynamic assessment at different stages, and associated recommendations for trustworthiness assessment via stakeholder engagement.

Stakeholders' views on Benevolence and Integrity of AI mapping

Specific AI-generated mapping solutions will only be adopted if they satisfy users' needs and reflecting these will help building trust. In this spirit, further discussion with relevant stakeholders during the project's final workshop weighed up the specific value of Benevolence and Integrity versus Ability (as defined in Trust in AI: progress, challenges, and future directions), and widely considered the importance of algorithmic biases and ethical use and assessment, including independent audits, while acknowledging that integrity is inherently easier to audit than benevolence. It was agreed that these audits would allow for trustability to be interrogated, and models validated. There was also consensus around the need for ethical frameworks to exist at the international level.

The following themes were raised for future focus:

- The importance of uncertainty visualisation on trust in AI, in particular around spatial uncertainty representation.
- Environmental impacts of AI mapping, i.e. environmental leakage (where are the data centres, what damage it is creating and how it is counteracted by the positive impact of the mapping).
- Impact of developments on countries that are under mapped, and other country-specific constraints (e.g., countries where government doesn't trust the AI solutions and those where the government data is not of high enough quality for the AI to be trusted).



Figure 4: TAIM final workshop encouraged lively discussions on trust in AI mapping solutions.

Throughout, the project team considered co-benefits of AI mapping, from climate change mitigation to other social impacts, and aimed to reflect these considerations in our tools. While the field evolves rapidly, we will continue to engage in the conversation and build upon TAIM's responsible approach to AI development.