Text, logo

Description automatically generatedLogo

Description automatically generated with medium confidence

­­­

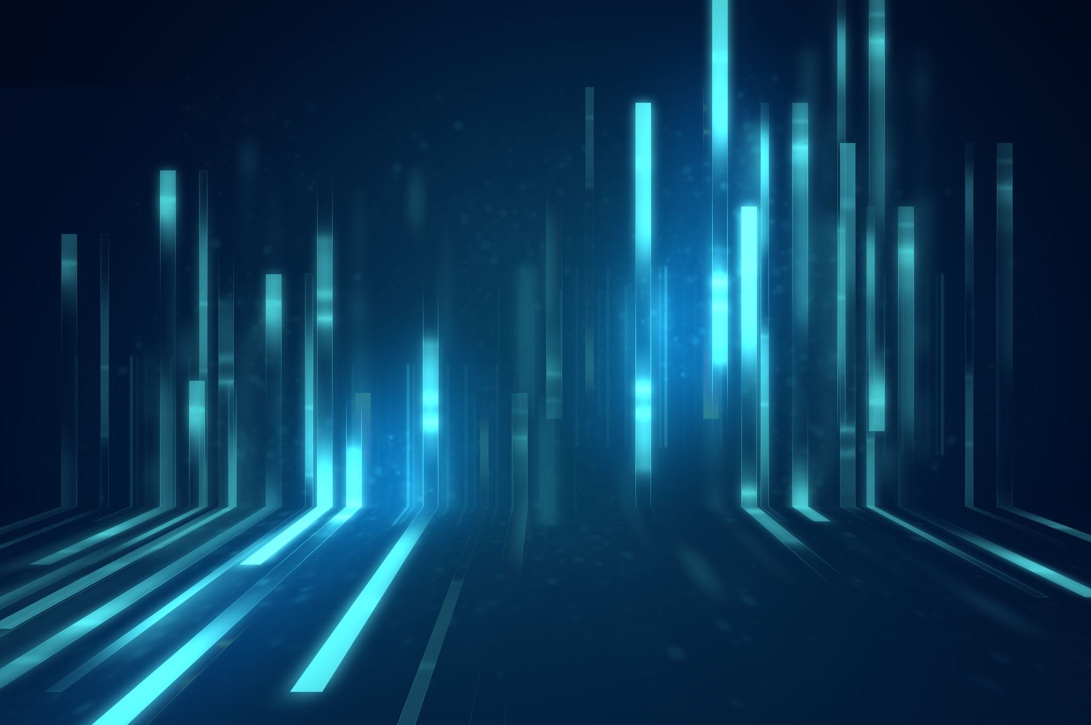
Transparency

Expert

Module 2

**Case Study**

**STAMINA**



# Introduction

For this case study, students will be given a document covering considerations and issues of use of AI in the public sector and an overview of the STAMINA project.

After this theoretical background, there will be a case study exercise.

The hand-out material is complemented by a video interview with a member of TRI’s team from the STAMINA project.

­­­

**Exercise**

This case study will look at the development of a social media listening tools to help you learn to ask: what does transparency mean in practice? What are the implications of uncertainty towards transparency? Which transparency requirements are specific to the public sector and why?

A pandemic planning agency has asked an AI company to develop a basic social media listening tool to help understand public sentiment. They want to know when outbreaks might be happening or when their policy is not being trusted.

The AI developer creates a tool that allows the end-user to provide some search parameters based on which social media posts are gathered from Twitter which then produces some descriptive analytics/summary visualisations on a dashboard. The pandemic planner “End-user would like to use the sentiment analysis insights of the tool to reassure the public in case negative ‘sentiment’ is detected.

Such an activity requires detailed transparency. This includes explainability – the reasons the AI produces specific results or behaves in specific ways. But it also includes the ability to communicate. For example, the end-user needs to be able to communicate to the developer how a parameter they provide relates to the sentiment they are looking for. In addition, the developer needs to be able to communicate the uncertainty in the data, predictions, and results. Such transparency is a starting condition for any assessment of fairness, bias, and representativeness of the AI and decisions made with the outputs. Just as much, it is important for building trust: trust between developer and user, and trust between the public and decision-maker.

In order to refine the tool and identify transparency and explainability requirements, a discussion between developer and stakeholders is needed. One productive place to start is around uncertainty.

Work through the worksheet accompanying this case study on your own. Then revisit your answers -- particularly to the questions in the green boxes -- considering the different “roles” involved in the design and use of the AI outputs.

**Note on the relation between uncertainty and transparency**

Uncertainty is an alternative form of this transparency, e.g., referring to lack of knowledge, of things like margins for error, limits of the data, rate of inaccuracies, or general confidence in results.

Uncertainty is crucial in helping us deal with transparency. It refers to our lack of knowledge about data, process, or outcome, that, if well communicated, can help stakeholders better understand technical outputs, address fairness issues in data, and improve trust in technology.

Mapping uncertainties can help, for example, identify potential disparate benefits or harms by highlighting where gaps in knowledge can lead to missed opportunities to provide aide. Well communicated, uncertainty can also be a way to ensure a human-in-the-loop, helping limit confirmation bias.

Leaving uncertainty unexplained can lead to unintentional bias, unfair outcomes that cannot be justified to the public, or miscalibration between data, models, and goals.

**After completing the worksheet**

Revisit answers through some role playing, and see if you want to add or change your answers:

* Pandemic Planner in multi-cultural city
* Doctor in a wealthy community
* Key worker in an impoverished region
* AI developer new to pandemic decision-making



­­­

**Follow up reflection questions**

Based on the uncertainty and risks you identified, think about the answers to these questions? Are there single answers or could there be multiple? Do you think all stakeholders will agree? If yes, how would you know? If no, what be done about it?

Think about social media analytics:

* What do you think is meant by ‘negative sentiment’?
* What kind of question can you really answer with this tool?
* What are the bias and limitations of such a technique?

How should the uncertainties and assumptions in the model be communicated to:

* model user?
* decision-maker?
* public?

Thinking about the user and their actions using the insights:

* How can a pandemic risk manager use such a tool to not confirm what is already seen/assumed?
* How can such tools provide necessary transparency to understand if what is presented helps support those “most” in need of aid? How do you define “the most”? Would a pandemic planner, a rural community member, or a grocery clerk define it differently?
* How can a pandemic risk manager know their goals are aligned with the tool scope/design? What kind of transparency or explainability indicators can support this awareness?

What else might a user need to have data about to:

* understand what is presented?
* be accountable in their use of what is presented?
* justify their decisions using the data?

Thinking about the designers:

* How can tool designers understand the limitations and scope of the decision-making process and available data?