

## **RELX response to EU High Level Expert Working Group – Ethics Guidelines – Trustworthy AI Assessment List Pilot**

### **Introduction**

RELX welcomes the opportunity to provide feedback on the EU's High-Level Expert Working Group Trustworthy AI Assessment List. RELX is a European company and a global provider of information and analytics to a range of sectors, including financial services, science, technology, medical, healthcare and energy. We employ 30,000 people worldwide and support customers in 180 countries. We utilise technology and data to help our customers improve their decision making across the sectors we serve. For example we help:

- scientists make new discoveries
- doctors and nurses improve the lives of patients
- lawyers win cases
- prevent online fraud and money laundering and
- insurance companies evaluate and predict risk.

RELX is therefore very involved with Artificial Intelligence, and is delighted that one of its employees, Elisabeth Ling, is a member of the High-Level Expert Group. The feedback provided here should be considered separate to the input Elisabeth has made to the HLEG thus far.

RELX is a diverse organisation, as outlined above. The feedback below has been collected from senior individuals from across the organisation, representing different business areas and product lines.

In order to provide structured feedback on the Assessment List, we have broken down our response into four primary areas, which have been inspired by the HLEG's Ethics Guidelines for Trustworthy AI document:

1. Interpretability
2. Implementability
3. Conflicts with existing regulation
4. Completeness

### **1. Interpretability**

When considering the Assessment List, the questions posed are generally understandable, however there is also a feeling that most of the questions use generic language which could be interpreted in different ways which has the potential to confuse teams. It was less clear what 'trustworthy' meant in the context of the document. There was a strong feeling that ethical and trustworthy AI are two separate things. Indeed, it is quite possible for trustworthy AI to be unethical and for ethical AI to be untrustworthy. Whilst the "ethics guidelines for trustworthy AI" articulate the components for trustworthy AI, we think that as a standalone Assessment List to share with technologists (who are not likely to read the longer document), a definition of trustworthy AI is needed, up front.

The EU has well established definitions and expectations for fundamental rights, however the individuals most likely to interact with this Assessment list may not be so familiar with what they are. In future drafts of the list it would be helpful to include a brief definition of fundamental rights.

There was also confusion as to the meaning of question 11.2 (Did you provide the necessary information in case of a risk for human physical integrity?) and we would suggest making the language clearer.

Some further specific feedback on questions within the list included the following:

Question 3.2, which considers whether users should be informed that an algorithm was involved in making a decision, is particularly complicated in the context of an *Artificial Intelligence* assessment list as algorithms predate AI by some decades. This would mean alerting users to most products already in use today, for example auto-pilots in aircraft. Businesses already face obligations under the GDPR Article 22 where data subjects have a right not to be subjected to automated decisions where there is a legal or similarly significant effect. Given the widespread use of algorithms in both modern and more traditional systems, the current requirement under GDPR achieves a suitable balance providing redress to users but not overwhelming businesses.

Question 7 might be better worded as *“Did you perform a formal threat modelling exercise against the AI system and identify points of risk to the overall system (e.g. cyber-attacks, physical infrastructure, data leakage, data pollution, etc.)?”*. Additionally, Question 7 and 7.1 are very similar and with this amendment to question 7, question 7.1 becomes less necessary.

It would be helpful to include an example with question 8, such as *“For example, how does the system behave with extraneous input? With boundary conditions? And invalid or impossible input?”*

In relation to question 10, sufficient should be replaced with ***fail-safe***.

Question 45 uses the term ‘preferences’ which is vague and unhelpful without further definition, particular in relation to the interplay between what is considered a preference compared to a need. We would suggest rephrasing the question as follows - *“Did you ensure that the AI system accommodates a wide range of individual **needs** and abilities”*.

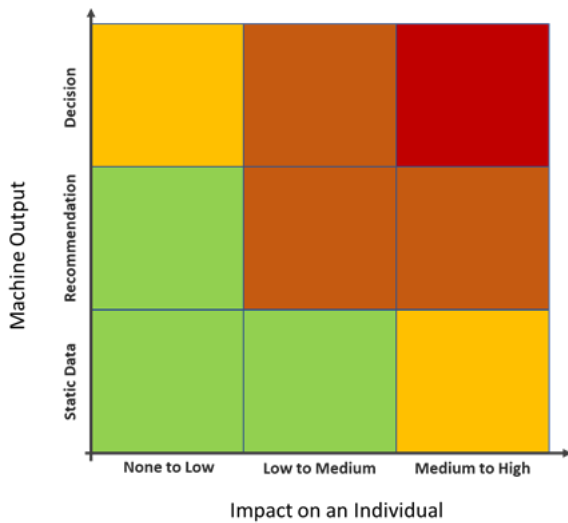
It was felt that question 46 was of limited value because user analysis is the key driver of product development and there would be no product without taking the user into account.

## **2. Implementability**

Across RELX we had different views as to whether the AI Assessment list was actionable, notwithstanding the specific questions outlined above where there was confusion.

It is worth noting that one of the business areas within RELX applied the Assessment List to a live project and found that a considerable number of the questions simply did not apply to the project. However, the same team recognised that if the scope of the list is targeted at all forms of AI projects the questions would likely be relevant in some cases and not in others.

Whilst not all questions will be relevant to all AI projects, as a whole RELX takes the view that a single set of questions is probably the most efficient way of implementing such a procedure. Attempting to differentiate for different projects would likely lead to additional compliance and administrative confusion and distract from the ultimate objective of achieving trustworthy AI. That said, an important contextual element is the extent to which an AI-based decision impacts one or more individuals in a significant way. When this is the case assessment should be more stringent.



Even though RELX believes that the Assessment list is unlikely to apply in its entirety to every project, the list's strength is that it has a broad applicability in terms of technology. This allows teams to draw on questions where helpful/appropriate from the Assessment list.

### 3. Conflict with existing regulations

RELX notes that there could be overlap with a number of existing regulations which would cause confusion when developing products. In particular there are overlaps with GDPR, security regulations and the FCRA in the US.

It would be helpful for the Commission to produce guidance on how this AI Assessment List, when finalised, should interact with the legal obligations' organisations are already under, and how any conflicts should be resolved.

Additionally, specifically relating to accessibility concerns, the questions should refer to existing accessibility guidelines as a reference to ensure accessibility requirements are properly implemented.

### 4. Completeness

RELX believes that the questionnaire is already lengthy and there are potentially concerns that implementing the entire list against all systems which might be considered AI could be too onerous and therefore limit adoption. We would recommend cutting down the number of questions and eliminating those that overlap.

However, while bearing that consideration in mind, there are a number of very specific additions RELX believes could add to the Assessment List's completeness.

Within the resilience section we would suggest adding a question that addresses penetration testing against AI systems, such as *'Have you performed a penetration test against the AI system to ensure that it has no cyber vulnerabilities? Have you had an independent assessment performed against the AI system to ensure that it has no cyber vulnerabilities?'*

With regard to accessibility, additional questions could include probing on whether an AI system provides any notable benefit to people with disabilities and whether the AI system meets the Web

Content Accessibility Guidelines (WCAG 2.0) International Accessibility Standard or other relevant accessibility standards.

## **Conclusion**

RELX hopes this feedback on the AI Assessment List is helpful. The EU HLEG has produced the first comprehensive list of questions developers of AI should be considering, for which it should be congratulated. RELX understands that the HLEG does not expect the version which is currently being piloted to remain static, particularly given the systems being discussed are evolving at pace and stands ready to provide further input to the work of the HLEG and the Commission as appropriate.

Submitted by:

Elizabeth Crossick

Head of Government Affairs, EU

RELX

EU Transparency Register: [338398611148-62](#)

On 29 November 2019