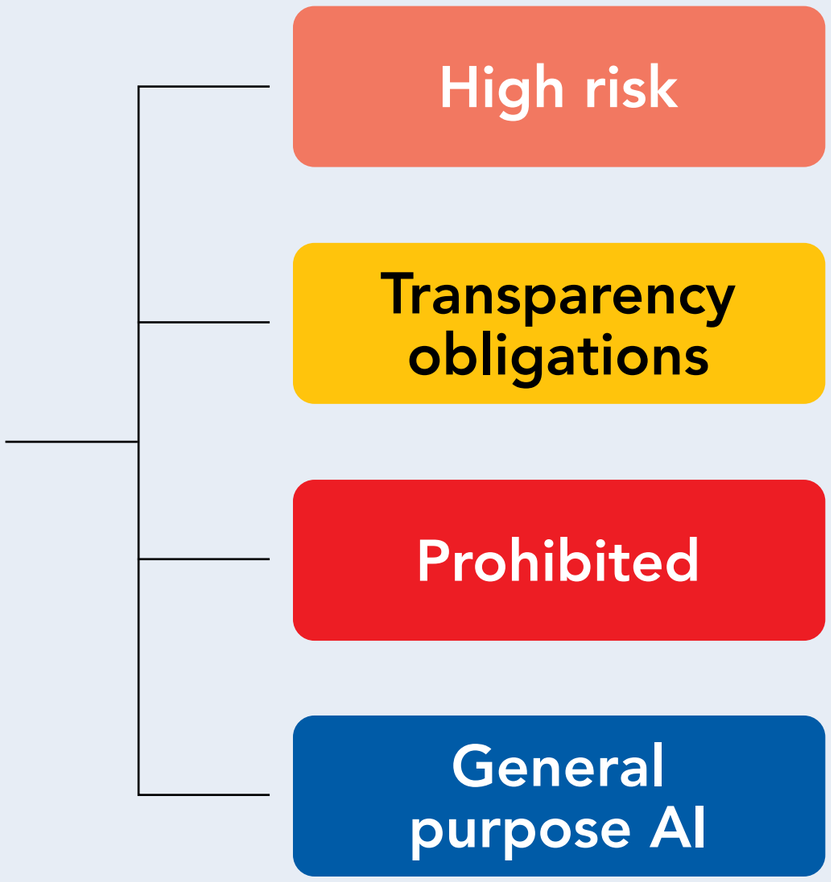
Algorithm Audit

# Risk classification of AI systems under the AI Act

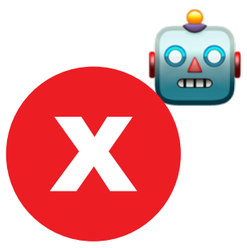



## How the AI Act Implementation Tool is helping you

# Example 1 – Emotion recognition: often incorrectly classified as prohibited AI system

## The front page of a major Dutch newspaper wrongly stated that emotion recognition is prohibited under the AI Act


*"AI in Europe is not allowed to detect emotions. This is strange because I appreciate a chatbot recognizing that I am not feeling well when interacting with a helpdesk"*



A chatbot with emotion recognition is not a prohibited systems under the AI Act …

… but other emotion detection system are prohibited or high-risk AI systems.

# Example 2 – Social scoring: frequently overlooked as prohibited AI system

## Prohibited social scoring is often missed, but also straightforward to avoid

### Guidelines of the European Commission:

> A fraud risk detection systems which "relies on characteristics collected or inferred from social contexts with **no apparent connection or relevance for the assessment of fraud**, such as having a spouse of a certain nationality or ethnic origin, having an internet connection, behaviour on social platforms, or performance at the workplace [...]"

Using data collected in an unrelated context, could mean that inconspicuous forms of profiling become prohibited social scoring. In this respect, compliance with GDPR (data minimalization and purpose limitation) helps with AI Act compliance.

# Dissecting legal definitions of the AI Act is complex

## Correctly classifying AI systems under the AI Act, requires recursively following definitions and translating legal concepts to AI practice

### Art. 5(1)(f) – Prohibited AI Practices

"the placing on the market, the putting into service for this specific purpose, or the use of **AI systems to infer emotions** of a natural person in the areas of **workplace and education** institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons"

Further guidance on what counts as 'workplace' and 'education' is given in the EU Commission's Guidelines on prohibited artificial intelligence practices

Emotion recognition systems are defined in Article 3(39). This definition in turn depends on the definition of biometric data as defined in Article 3(34). Further guidance is given in the EC Guidelines.

# The AI Act Implementation Tool supports quick, structured risk classification

## Uniform approach counters paralysis and legal uncertainty

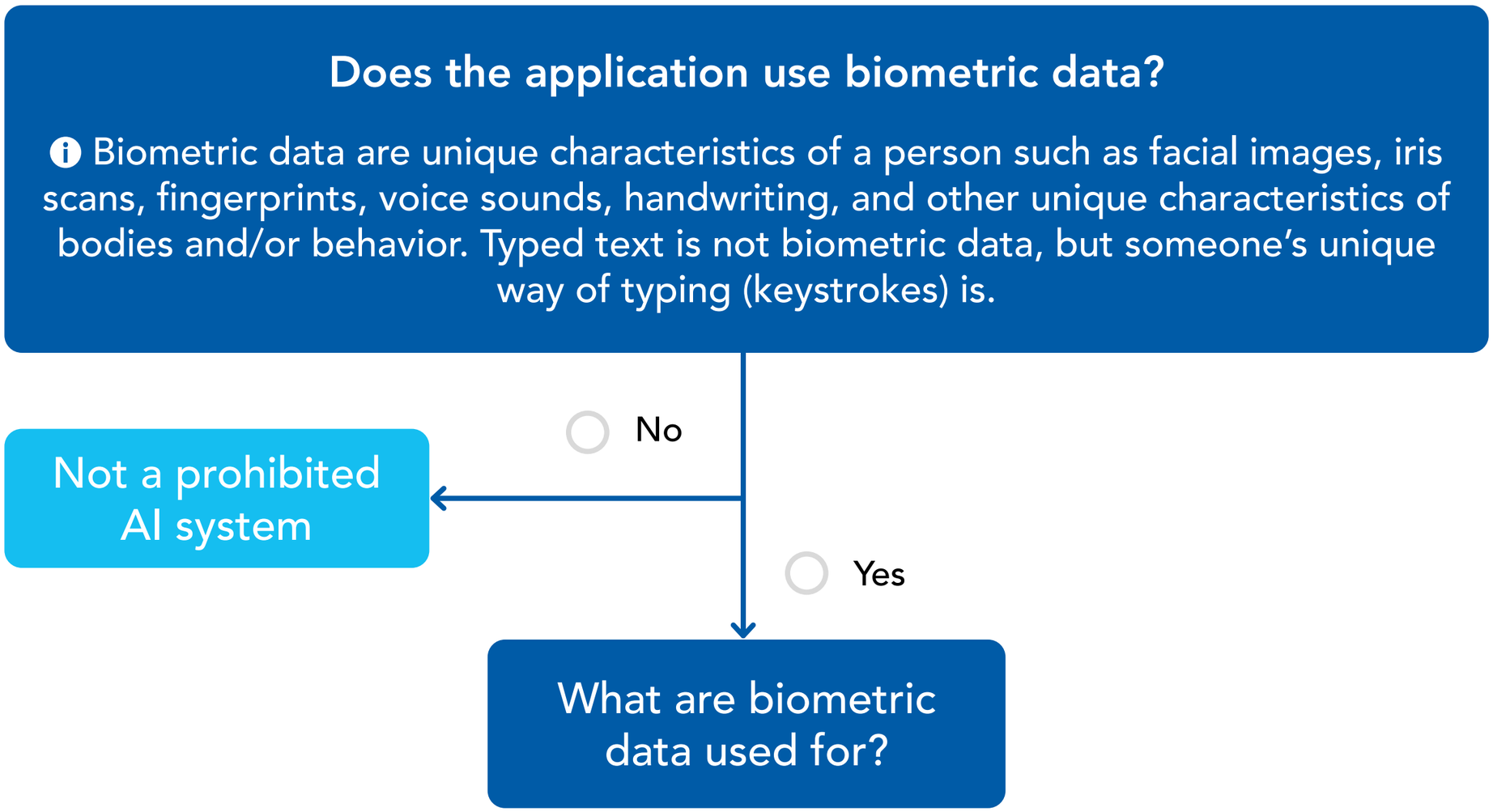
In the AI Act, there are 8 different prohibited AI practices and 25 different high-risk use-cases in 8 categories in Annex III only. This gives use well over a **100 dependencies, terms and definitions** to dissect…

… operational managers or tech leads should **be empowered to quickly and reliably determine the AI Act risk classification** of their systems by the themselves

Open-source **AI Act Implementation Tool** assists with risk classification through multiple-choice questions presented in clear, accessible language

# Example 1 – Emotion recognition: Dynamic questionnaires and multiple-choice minimize efforts to understand the AI Act

## The definitions are translated to yes/no questions, ordered to make sure only a minimal number of questions needs to be answered




🤔 If no biometric data is used, it is not prohibited emotion recognition. Therefore, first answering this questions ensures that a correct assessment is made as quickly as possible.

# Example 2 – Social scoring: Dynamic questionnaires and multiple-choice minimize efforts to understand the AI Act

## The definitions are translated understandable questions, ensuring legal expertise is only required after a first assessment is made

**Could the system lead to unfavourable treatment in a different context than the behavior or personal characteristic on which the score is based?**

ⓘ Unfavourable treatment in an unrelated context is, for example, a fraud risk score that leads to an investigation, where the fraud risk score is based on characteristics without clear relevance to the assessment of fraud, such as having a partner with a certain nationality, having an internet connection, behavior on social platforms, or performance at work.

○ No

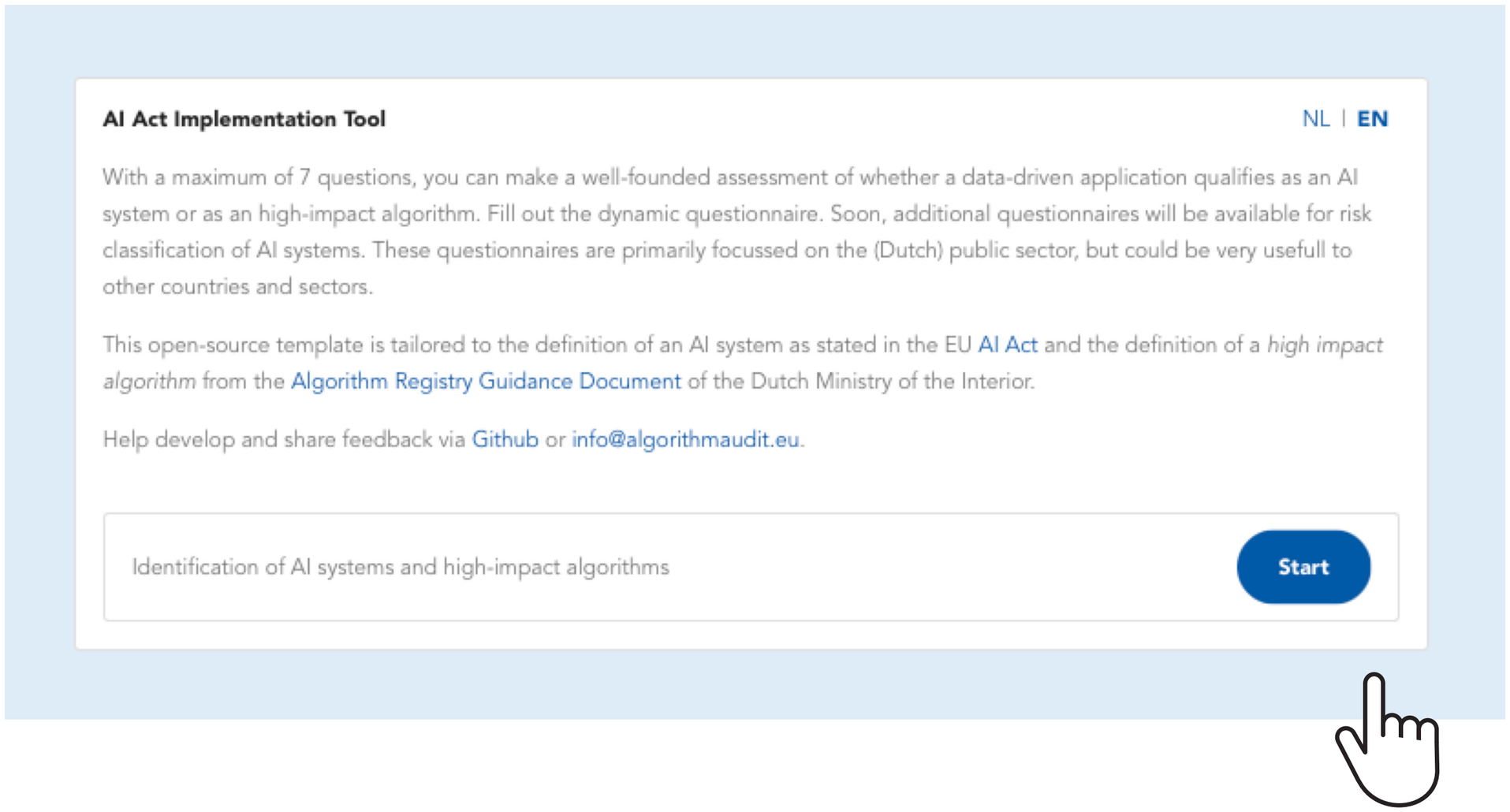
Not a prohibited AI system

○ Yes

Potentially a prohibited AI system – consult a legal expert

🤔 Prohibited social scoring is defined (among other things) by using personal data from different contexts than what is predicted by the AI system.

# Use our open-source AI Act Implementation Tool to assign the right risk classification for your AI systems




https://algorithmaudit.eu/technical-tools/implementation-tool/#tool

- Open-source developed under EUPL-1.2 license
- Developed and tested in collaboration with the Municipality of Amsterdam

Gemeente Amsterdam

Building *public knowledge*

for *ethical algorithms*

## Join the discussion!

Algorithm Audit

www.algorithmaudit.eu

info@algorithmaudit.eu

www.linkedin.com/company/algorithm-audit

www.github.com/NGO-Algorithm-Audit