How to detect and manage AI errors and risks; human baseline and engagement

Problematic AI: Finding the Best Way Forward

February 10, 2023

ACT AUTONOMY THROUGH CYBERJUSTICE TECHNOLOGIES

CENTRE DE RECHERCHE EN DROIT PUBLIC
Université de Montréal

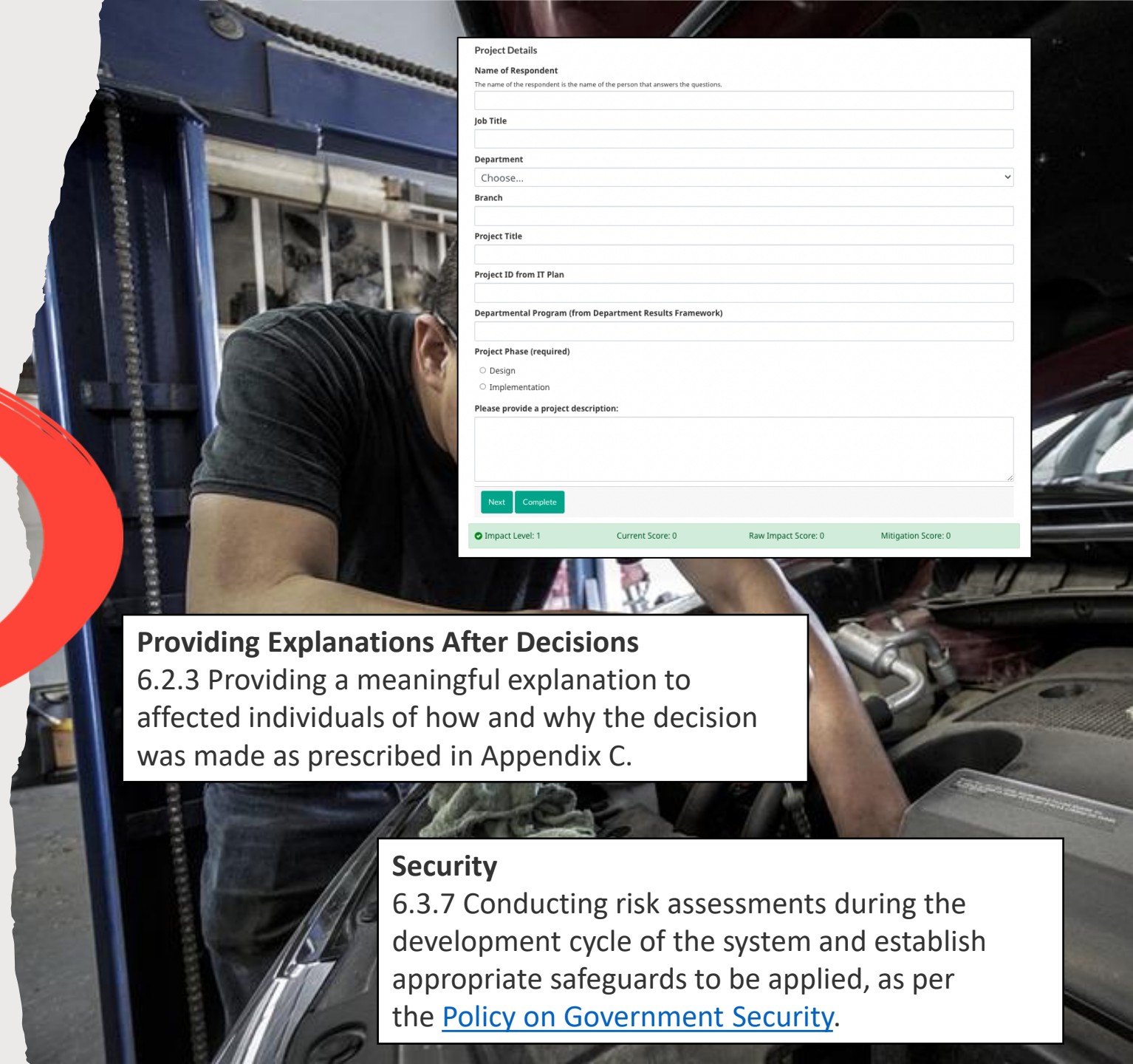
Laboratoire de CYBERJUSTICE Laboratory

Faculté de droit
Université de Montréal

Directive on Automated Decision-Making

Requirements

- Algorithmic Impact Assessment
- Transparency
- Quality Assurance
- Recourse
- Reporting



Providing Explanations After Decisions

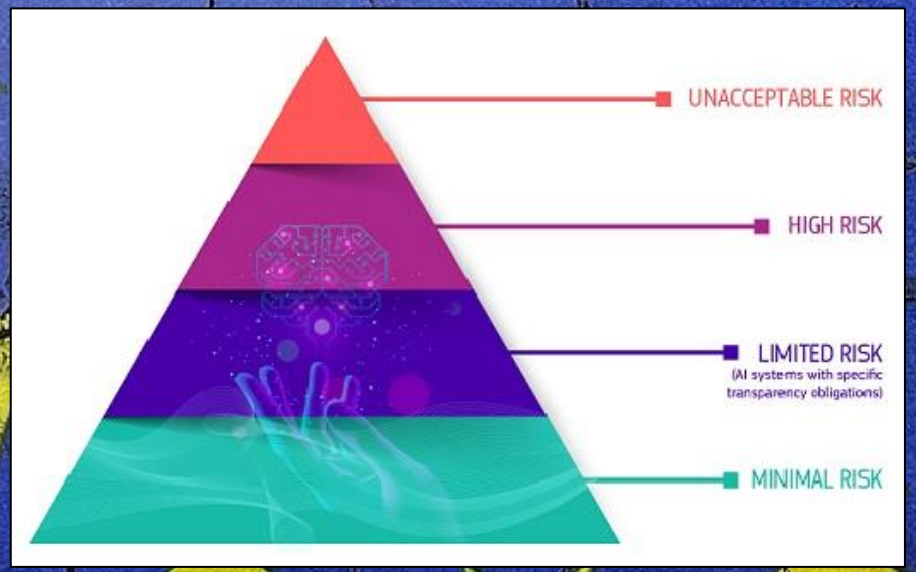
6.2.3 Providing a meaningful explanation to affected individuals of how and why the decision was made as prescribed in Appendix C.

Security

6.3.7 Conducting risk assessments during the development cycle of the system and establish appropriate safeguards to be applied, as per the Policy on Government Security.

Treasury Board of Canada Secretariat

The technical robustness is a key requirement for high-risk AI systems. They should be resilient against risks connected to the limitations of the system (e.g. errors, faults, inconsistencies, unexpected situations) as well as against malicious actions that may compromise the security of the AI system and result in harmful or otherwise undesirable behaviour.



Source: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts

Article 1

Subject matter

This Regulation lays down:

(a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union;

(b) prohibitions of certain artificial intelligence practices;

(c) specific requirements for **high-risk AI systems** and obligations for operators of such systems;

(d) harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content;

(e) rules on market monitoring and surveillance.

Artificial Intelligence and Data Act

8 A person who is responsible for a **high-impact system** must, in accordance with the regulations, establish measures to identify, assess and mitigate the risks of harm or biased output that could result from the use of the system.

American Data Privacy and Protection Act

SEC. 207. CIVIL RIGHTS AND ALGORITHMS.

(A) IMPACT ASSESSMENT.—Notwithstanding any other provision of law, not later than 2 years after the date of enactment of this Act, and annually thereafter, a large data holder that uses a covered algorithm in a manner that **poses a consequential risk of harm** to an individual or group of individuals, and uses such covered algorithm solely or in part, to collect, process, or transfer covered data shall conduct an impact assessment of such algorithm in accordance with subparagraph (B).

Contact information

Cyberjustice Laboratory

Pavillon Jean-Brillant (B-2215)
Université de Montréal
3200 rue Jean-Brillant
Montreal (Quebec)
H3T 1N8

Nicolas.vermeys@umontreal.ca

www.cyberjustice.ca

ACT AUTONOMY THROUGH CYBERJUSTICE TECHNOLOGIES

Faculté de droit Université de Montréal

CENTRE DE RECHERCHE EN DROIT PUBLIC Université de Montréal

Laboratoire de CYBERJUSTICE Laboratory