

Preventing the potential discriminatory effects of the use of artificial intelligence in local services

Policy Brief

October 2020

ePaństwo Foundation

INTERCULTURAL CITIES
CITÉS INTERCULTURELLES

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

The opinions expressed in this work are the responsibility of the author and do not necessarily reflect the official policy of the Council of Europe.

Written by Krzysztof Izdebski

Intercultural Cities Unit,
Council of Europe©

Council of Europe, October 2020

Krzysztof Izdebski. Board Member and Policy Director of ePanstwo Foundation (EPF) and Board Member of Consul Democracy Foundation. He is a lawyer specialized in access to public information and re-use of public sector information. He is the author of publications on freedom of information, technology and public administration including "Transparency and Open Data Principles: Why They Are Important and How They Increase Public Participation and Tackle Corruption" and recently published "alGOVrithms. The State of Play. Report on Algorithms Usage in Govemment-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia, and Slovakia.

This policy brief was produced as a background paper based on the webinar organized by the Intercultural Cities Programme of the Council of Europe and Epaństwo Foundation on 21 September 2020.

This is part 1 of the policy brief prepared for the online course Artificial intelligence and anti-discrimination for local authorities by the Intercultural Cities programme. The full policy brief published online can be found on the Intercultural Cities webpage.

1. Introduction

1.1 Background

Municipalities provide a wide range of public services to their citizens and increasingly this is supported by technologies including Automated Decision Making (ADM) tools and Artificial Intelligence (AI) solutions. The deployment of IT tools in public services has brought new challenges and potential risks of bias, prejudice towards certain categories of citizens, and discrimination. Such risks were, for example, detected in the Dutch SyRI system used by national and local authorities to detect housing or social security fraud, smart water meters in several cities in Europe or AI applications used in staff recruitment.

Some cities - like New York - have already implemented measures to prevent such irregularities, others are only starting to consider what steps should they take. Intercultural cities develop policies and expertise in social inclusion and equality, prevention of discrimination, and raising awareness around important societal challenges. It is useful for decision-makers to also understand the potential biases and risks of AI and learn about ways of mitigating such risks. The experience of advanced cities could help build trustworthy and ethical AI.

The Intercultural Cities Programme held a webinar about the challenges Artificial Intelligence and Algorithmic Decision-making present for local authorities, in particular in relation to (anti-) discrimination, inclusion, and the fight against hate speech. The webinar was prepared and led by Krzysztof Izdebski*, Policy Director of ePaństwo Foundation.

The report reflects the substantial content of the webinar and serves as a short guideline on preventing the potential discriminatory effects of the use of artificial intelligence in local services.

1.2 Glossary

Artificial Intelligence (AI): *Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems.*

- Directive on Automated Decision Making (Canada)

AI is only a type of algorithm which may cause discriminatory risk. As was stated in the Algorithm Charter For Aotearoa New Zealand, *the risks and benefits associated with algorithms are largely* ***unrelated to the types of algorithms*** *being used. Very simple algorithms could result in just as much benefit (or harm) as the most complex algorithms depending on the content, focus and intended recipients of the business processes at hand.*

Therefore, a better term to use is ***Automated Decision [Making] System*** *which according to* the Directive on Automated Decision Making (Canada) *includes any technology that either assists or replaces the judgement of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.*

To put it even simpler, after David Harel and his work Algorithmics - The Spirit of Computing (1987), we can compare an algorithm with a cooking recipe. While ingredients can be compared to input data, and a finished dish is a result, many activities such as selecting appropriate proportions at the right time or applied

methods of thermal processing are just an algorithm. From life experience, one can easily deduce that one mistake at the stage of preparing a dish can lead to failure in its taste and appearance.



According to C. Orwat in (2020) Risks of Discrimination through the Use of Algorithms:

Discrimination *is disadvantageous, unjustified unequal treatment of persons in connection with a protected characteristic.* Such characteristics can include "race" or ethnic origin; ancestry, home country, origin; gender; language; political opinion or viewpoint; religion and belief; disability; trade union affiliation; genetic characteristics or dispositions and health status; biometric characteristics; sex life, sexual identity or orientation.

To differentiate between the "traditional" and AI/ADM discriminatory it is important to take into account the below.

Taste-based discrimination is unequal treatment based on the personal, prejudiced dislikes or preferences of the decision-makers against or for a certain group of people or on dislikes or preferences for certain products.

Statistical discrimination is the unjustified unequal treatment of persons on the basis of surrogate information.

However, it is crucial to understand that, as algorithms are created by humans with all their biases included, the statistical discrimination can originate from the taste-based discrimination. These two phenomena are therefore very rarely independent of each other.