Non-discriminatory artificial intelligence – mission impossible?

Presented by: Aurelija Šerniūtė
Commentators: Prof. Urs Gasser, Dr Yaniv Benhamou

Geneva Digital Law Research Colloquium
June 19, 2020
Aurelija Šerniūtė

Research field – the interaction between emerging technologies & human rights

✓ PhD Student at Vilnius University, Faculty of Law (supervised by Prof. Tomas Davulis)

✓ Junior Assistant at Vilnius University, Faculty of Law, Department of Private Law

✓ Project Coordinator at Vilnius University Tech Hub
Discriminatory

Problematic features

Negative practical examples

Ready to fix discrimination

Promises in various spheres

Positive practical examples

Artificial intelligence
What is our goal behind the idea of artificial intelligence?
What is our goal behind the idea of artificial intelligence?

**ARTIFICIAL HUMAN**

AI = HUMAN INTELLIGENCE

*Idea:* resembles human intellectual abilities

*Talking point:* strong AI

**ARTIFICIAL SUPERINTELLIGENCE**

AI > HUMAN INTELLIGENCE

*Idea:* focuses on the optimal performance of the tasks

*Talking point:* singularity

**SUPER-INTELLIGENT HUMAN**

AI as a way to HUMAN INTELLIGENCE+

*Idea:* expands human capacities

*Talking point:* transhumanism
1. **Ethical principles, guidelines**
   “Diversity, non-discrimination, fairness”

2. **A new set of rules for the AI**
   Due to exceptionally high risk of discrimination, the use of artificial intelligence is prohibited in situations: [x], [y], [z].

3. **Technology-specific legal provisions**
   “The producer should be strictly liable for defects in emerging digital technologies even if said defects appear after the product was put into circulation, as long as the producer was still in control of updates to, or upgrades on, the technology. A development risk defense should not apply.”

4. **Self-regulation**
   \[
   r(x) \geq \frac{b_{00} + c_{10}}{b_{00} + b_{11} + c_{01} + c_{10}} = c_{det}/b_{crime}.
   \]
<table>
<thead>
<tr>
<th>Regulatory options</th>
<th>Ethical principles, guidelines</th>
<th>A new set of rules for the AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+ balanced interests of multiple stakeholders</td>
<td>+ space for creativity, clear protective limits</td>
</tr>
<tr>
<td></td>
<td>- lack of detail</td>
<td>- potential unawareness of positive future examples</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><strong>Technology-specific legal provisions</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ additional certainty to the current regulatory landscape</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- disregarded technological neutrality</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><strong>Self-regulation</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ highly adaptable standards created by industry participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- values decided by a relatively small group of people</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

I am looking forward to your feedback!

Please do not hesitate to contact: aurelija.serniute@tf.vu.lt or