Trans Atlantic Consumer Dialogue

DOC NO: DIGI 01/23 **DATE ISSUED:** 23 May 2023

TACD POLICY PRIORITIES FOR ARTIFICIAL INTELLIGENCE

Our lives are dominated by products and technologies which are interconnected and increasingly automated and adaptive. The shift towards the use of automated decision-making based on algorithms will change the way in which consumer markets and our societies function.

Artificial intelligence (AI) has the potential to bring many **benefits** to consumers. For example, the connected vacuum cleaner robot or lawn mower liberates consumers from household tasks – time, that consumers can use otherwise. Also, AI applications embedded in goods or simply operating in the background of digital services offer the potential to optimise energy (e.g., smart meter) or easily compare complex offers (price comparison tools).

However, AI also comes with significant **risks and challenges** for consumers. For example, the use of AI can lead to increased risks of algorithmic bias and unfair discrimination among different groups of people on the basis of economic criteria, gender, or a person's health. More broadly, the use of AI can negatively affect consumers' autonomy, freedom of choice, privacy and, ultimately, the ability to hold businesses responsible if something goes wrong:

- For many years, the Dutch government used an automated system that wrongly picked out and accused thousands of parents, mostly from low-income families and ethnic minorities, of fraud and ordered them to repay the child support benefits. As it is now known, the 'Dutch Benefits Scandal' had devastating consequences for many. Because of the unjust claim to repay their benefit (in some cases, the amount went to tens of thousands of euros), people have gone bankrupt or were forced to move house. Several couples divorced.[1]

- Airbnb recently patented an AI system capable of generating social scoring to determine consumers' trustworthiness based on a variety of social media/online data. If operational, this AI is likely to lead to the discrimination of some social groups in the market for holiday homes – and yet, this may be just the tip of the iceberg.[2]

- Spotify has patented technology that will allow it to analyse your voice and suggest songs based on your "emotional state, gender, age, or accent".[3] Emotion recognition technology can be used for other things than adjusting the music you are listening to. For example, Facebook is currently considering how to monetise your emotions (e.g., pupil movements,

[1] Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms, https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/

[2] Booker beware: Airbnb can scan your online life to see if you're a suitable guest, https://www.standard.co.uk/tech/airbnb-software-scan-online-life-suitable-guest-a4325551.html

[3] Dear Spotify: don't manipulate our emotions for profit, https://www.accessnow.org/spotify-tech-emotion-manipulation/

body poses and nose) in the Meta world.[4] Emotion recognition has been shown to be patent "snake oil," as first explained in a book by Lisa Feldman Barrett[5].

Recommendations:
In light of these challenges, TACD supports regulatory efforts throughout Europe and the United States, urging them to include at least the following points:

1. ALL ENTITIES RESPONSIBLE FOR AI SYSTEMS SHOULD COMPLY WITH MANDATORY RULES, INCLUDING TRANSPARENCY, ACCOUNTABILITY, AND FAIRNESS WITH RESPECT TO BOTH ALGORITHMS AND THE DATA FED INTO ALGORITHMIC SYSTEMS.

a) Fair, by avoiding discrimination or causing material distortion of the consumers' economic behaviour, by appreciably impairing their ability to make an informed decision, thereby causing the consumer to take a transactional decision that they would not have taken otherwise.
b) Accountable, by requiring all AI providers to regularly monitor their AI system and assess if it respects the obligations set out in law.
c) Transparent, by requiring AI operators to inform consumers and citizens when they are subject to an AI system and also by requiring the AI to provide an explanation about how it reached its decision affecting a person, e.g., for the way services or products are chosen by a personal assistant, or for access to a service. This explanation should be provided to affected people, regulators, and the public.

2. STRIKE THE RIGHT BALANCE WHEN DEFINING AI.

The definition of AI deserves careful attention. AI should be defined in a way which strikes the right balance between a too-narrow scope that would deprive many consumers of much-needed protection and scope that would virtually encompass any software.

[4] Facebook patents reveal how it intends to cash in on metaverse, https://financialpost.com/fp-finance/facebook-patents-reveal-how-it-intends-to-cash-in-on-metaverse
[5] Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain,* Mariner Books, 2017.

GENERATIVE AI

Generative AI is a type of AI system that generates text[6], images[7], audio or video. Since the launch of ChatGPT in November 2022, these systems gained worldwide attention. For example, chatbots similar to ChatGPT are widely used by consumers[8] and they are increasingly incorporated into products and services by businesses.[9]

Such technologies can bring benefits to our economy and society but also come with big challenges for consumers that become more evident each day. In the EU, recently, the Italian Data Protection Authority temporarily banned ChatGPT in Italy.[10] Other data protection authorities have decided to investigate it further[11] and the European Data Protection Board set up a task force to coordinate between authorities on possible enforcement actions conducted by data protection authorities.[12] Evidence shows that this technology can pose serious risks to safety[13] and consumer rights.[14]

In the US, this topic is also in the spotlight. President Bident recently underlined the importance of AI protecting people's rights and safety.[15] In a recent op-ed, Federal Trade Commission Chair Lina Khan called for regulation on AI.[16] Finally, civil society organisation CAIDP (Center for AI and digital Policy) has filed a complaint to the FTC, in which ample evidence is provided not only of the various consumer risks, but also of the public safety and health risk and concerns for consumers.

We urge regulators from both sides of the Atlantic to take action against companies deploying generative AI applications that do not respect current regulations protecting consumer rights, safety, privacy and personal data protection. We urge policy-makers to fill in the regulatory gaps in governing generative AI, including with the adoption of new regulation and the update of existing legislation.

---

[6] ChatGPT, Bard,
[7] Midjourney, DALL-E, Stable Diffusion,
[8] Reuters, ChatGPT sets record for fastest-growing user base - analyst note (2 February 2023)
[9] The Verge, Hands-on with the new Bing: Microsoft's step beyond ChatGPT (8 February 2023)
[10] Garante, Artificial intelligence: stop to ChatGPT by the Italian SA (31 March 2023) || Garante, ChatGPT: OpenAI reinstates service in Italy with enhanced transparency and rights for european users and non-users (28 April 2023)
[11] Reuters, Italy's ChatGPT ban attracts EU privacy regulators (3 April 2023)
[12] European Data Protection Board, EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT (13 April 2023)
[13] BEUC, Urgent all for action regarding generative AI systems and concerns related to their safety (12 April 2023)
[14] BEUC, Call for action to open an inquiry on generative AI systems to address risks and harms for consumers (21 April 2023)
[15] https://twitter.com/POTUS/status/1643359035398184960
[16] The New York Times, Lina Khan: We Must Regulate A.I. Here's How (3 May 2023)

3. CERTAIN APPLICATIONS OF AI MUST BE BANNED OR PLACED ON AN INDEFINITE MORATORIUM.

For certain applications of AI, the risks that they convey outweigh their potential benefits. It is particularly the case when such technologies directly put fundamental rights at risk, including consumers' physical or emotional integrity.

The ban should cover the following AI applications:

a) **Emotion recognition** – Emotion recognition technology is a growing industry that uses AI to detect emotions from various facial expressions and cues**.** Its use in various settings, including security, education, and employment contexts, causes harm that results in substantial injury. A 2019 meta-analysis of the relevant scientific literature revealed that there is no reliable evidence that an individual's emotional state can be inferred from their facial movements.[17] Emotion recognition technology is unable to "confidently infer happiness from a smile, anger from a scowl, or sadness from a frown" because it glosses over cultural and social contexts.[18]

b) **Facial analysis in publicly accessible spaces** because it is invasive, ineffective, and biased. Harms result from the photos initially fed into the database, or how the algorithm informing the analysis and recognition phases fail to accurately recognize certain faces. A study[19] from the National Institute of Standards and Technology ("NIST") analysed the facial recognition algorithms of a "majority of the industry" and found the software up to 100 times more likely to return a false positive for a non-white individual than for a white individual.[20] Additionally, individuals rarely have the opportunity to provide explicit and informed consent on whether they want to be identified, especially in a public crowd.

c) AI that has the objective or effect of materially **distorting a person's behaviour beyond a person's consciousness** with the effect of causing or is likely to cause that person or another person harm.

d) AI used for the **exploitation of a person's vulnerabilities** with the objective or effect to materially distort that person's behaviour in a manner that causes or is likely to cause that person harm. For instance, in the past, social media have suggested content related to self-harm to teenagers suffering from depression. In the UK, these recommendations contributed to the death of a 14-year-old teenager.[21] From a consumer perspective, such algorithms can become dramatically harmful if they repetitively present to people suffering from addiction content related to this very same addiction (e.g., advertising alcohol to alcoholics etc.).

---

[17] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movement", Psychological Science in the Public Interest 2019 20:1, 1-68.

[18] Idem.

[19] NIST, "NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software", 19 December 2019

[21] Morgan Meaker, "How A British Teen's Death Changed Social Media", The Wired, 5 October 2022

e) **Social scoring**: AI used by both public and private bodies to evaluate the trustworthiness of an individual based on their social behaviour or other personal attributes, such as someone's preferences, emotions, health or intelligence.

f) **Biometric categorization** systems because they are unscientific, artificially assign a person to a category (essentialization) and are often used for surveillance or targeted marketing purposes.[22]

g) **Predictive policing** because it is ineffective, increases the likelihood of police interactions, and provides justification for increased police presence in historically over-policed communities.

4. BEFORE BEING PLACED ON THE MARKET OR USED CONTINUALLY, AI SYSTEMS THAT CONVEY *HIGH-RISK* MUST UNDERGO A THIRD-PARTY AUDIT THAT MEET RIGOR AND INDEPENDENCE STANDARDS, INCLUDING AN AUDIT OF THEIR IMPACT ON FUNDAMENTAL RIGHTS.

The independent third-party evaluators should perform an impact assessment, including a fundamental right assessment as part of the audit. These impact assessments should be made public and focus on the impact on vulnerable persons, such as children and marginalized communities. The third-party audit should also evaluate the safety and security risks posed by the AI system.

The impact assessment should include at least the elements mentioned in Annex I.

5. INVOLVE CIVIL SOCIETY STAKEHOLDERS IN THE DESIGN OF STANDARDS FOR AI SYSTEMS.

The requirements applicable to AI systems, as well as the evaluation methods, will largely rely on standards. The standardization work has already started for years, notably at the international level in ISO/IEC JTC 1/SC 42.

It is worth noting that civil society and consumers' interests are not sufficiently represented in standardisation bodies. These bodies are essentially driven by industry and there is an urgent need to facilitate the participation of civil society actors in standardization through simplified procedures and increased funding. AI is a technology with such life-changing potential that all interested parties should have a say in the way technology is concretely implemented.

This is particularly true when it comes to international standardization bodies. Overall, it is important to be careful when it comes to the use of international standards for public policy objectives. In international standardization, there is significant participation from countries which do not share the democratic values and principles of the EU and the US.

Furthermore, standards should not substitute the work of democratically elected legislators. The definition of key legal concepts, such as the definition of what constitutes unfair bias, requires

[22] Access Now, "Prohibit remote biometric categorisation in publicly accessible spaces, and any discriminatory biometric categorization", January 2023.

democratic scrutiny. Standardization bodies are not subject to such scrutiny and hence are not well-placed to take important decisions relating to fundamental rights and consumer protection.

EU-US Roadmap on AI: In the context of the Trade and Technology Council (TTC), the EU and US published a roadmap on trustworthy AI. One of the objectives of this roadmap is to establish a working group of 'AI terminology and taxonomy', where basic concepts such as what should be considered unfair biased would be discussed. In line with what was mentioned, the definition of such concepts must be carried out domestically, by legislators, with sufficient democratic scrutiny.

For more information concerning the EU-US roadmap on AI, TACD provided detailed comments to the published draft[23].

6. RIGHT TO OBJECT AND REQUEST HUMAN REVIEW OF A DECISION WHICH CAN HAVE A SIGNIFICANT IMPACT.

For systems that can have a significant impact on people (e.g., a system granting social benefits or a credit scoring system), citizens and consumers should have the right to object and the right to request a review of the AI-made decision by a human operator.

7. HUMAN OVERSIGHT REQUIREMENT

All AI systems that convey high risks should comply with a human oversight requirement. Human oversight means that the AI system is built in such a way that it can be overseen by natural persons. For example, this means that a human operator should be able to interpret the output of the AI system and understand on which grounds the decision was taken. This should also mean that the natural person should be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system. Entities should log or otherwise document outcomes of human oversight, in order to assess the usefulness of the system, as a best practice.

8. PUT PRIVACY-BY-DESIGN AT THE HEART OF ALL AI SYSTEMS.

All AI systems should comply with privacy principles in their design and use. This means, for instance, that AI systems should not be trained on data that resulted from indiscriminate scraping. A prime example of such indiscriminate scraping is the Clearview AI application that constituted a database of portraits, collecting images from social media and denying any form of consent to the persons whose face is now present in the database. As a general rule, the collection of data should always follow the best practices in terms of data protection, needless

[23] The Consumer Perspective on the joint EU-U.S. Roadmap on Artificial Intelligence, February 2023 https://tacd.org/wp-content/uploads/2023/02/20230213-The-consumer-perspective-on-the-joint-EU-US-roadmap-on-AI.pdf

to say, it must also comply with existing legislation, such as the General Data Protection Regulation (GDPR) for EU citizens.

Biometric data are particularly sensitive because they relate to what a person is and remain relatively stable over time. If these biometric data end up in the wrong hands, little can be done by the victim to minimize the damage (you can change address, even name, it is a lot more complex to change face or fingerprints). For this reason, as a principle, the collection and processing of biometric data should be purpose-limited and collected only through informed consent. Exceptions must be few, duly justified and regulated by legislation.

In addition, the AI system should collect as little data as needed during its use, following the principle of data minimization. In cases where data is provided to an entity (e.g., health insurance company), that data should only be used for that purpose.

9. IMPLEMENT SECURITY BY DESIGN PRINCIPLES.

Security by design is a prerequisite to privacy by design. If the AI system is vulnerable to cyber-attacks, there is a risk of data leak amounting to a privacy breach. An AI system - especially those based on machine learning - can be vulnerable to data poisoning, meaning that the data that it uses for continuous learning is intentionally corrupted by a malicious actor. Attackers typically use the data poisoning attack to deceive the system, for instance, to confuse a spam filter and render it useless.

It is currently difficult to make a machine unlearn. Machine unlearning is a field of exploration. Therefore, it is even more important that providers of AI systems implement adequate preventive measures to detect and block attack attempts before the next training cycle takes place. Providers of AI should also implement strong access control measures for the model and the training data.

10. INDIVIDUALS AND CONSUMERS SHOULD BE GRANTED A RIGHT TO COMPLAIN AND A RIGHT TO COLLECTIVE REDRESS WHEN DEALING WITH AI SYSTEMS.

It is first and foremost the responsibility of public authorities to put in place effective enforcement mechanisms. AI deployers should be held liable for their violations and be responsible for facilitating more effective enforcement through transparency. For enforcement to be effective, public authorities should:

- Have a dedicated body with the required expertise.
- Allocate sufficient funding to enforcement activities.
- Implement discouraging measures, including fines and penalties.
- Cooperate with other countries for cross-border cases, but enforcement should first and foremost be the responsibility of the country where the breach takes place, not where the headquarter of the AI provider is located. The EU-US should cooperate on enforcement, including in the context of the EU-US Trade and Technology Council (TTC).
- Implement the possibility for consumers to lodge a complaint with a public authority if they believe that an AI system breaches the law.

Additionally, consumers should also have the possibility to undertake collective redress/class action, especially if they have suffered immaterial or material harm. Overall, consumers should be entitled to receive compensation for harm suffered by an AI system. Regulators, like the FTC in the *Kurbo* case, should also explore equitable remedies like model deletion and disgorgement.

ANNEX I – IMPACT ASSESSMENT

i. What decision(s) the system will make or support;
ii. Whether the system makes final decision(s) or supports decision(s);
iii. The system's intended benefits and research that demonstrates such benefits;
iv. A detailed description of the system's capabilities, including capabilities outside of the scope of its intended use and when the system should not be used;
v. An assessment of the system's impact on fundamental rights, not only for users but for the general public at large (including bystanders)
vi. An assessment of the relative benefits and costs to the consumer given the automated decision system's purposes, capabilities, and enforcement
vii. Inputs and logic of the system
viii. Data use and generation information, including:
    1. How the data is populated, collected, and processed as inputs;
    2. The type(s) and amount of data the system is programmed to generate[24];
    3. If the data serving as inputs for the system or outputs generated by the system are used downstream for any purpose not articulated in Sec. 1;
ix. Reference to the yearly validation studies and audits of accuracy, bias, and disparate impact;
x. A detailed use and data management policy.

[24] This should be done in descriptive terms (e.g., a number on a scale of 1–100 or a rating of low, medium, or high).