universität
wien

The Draft Artificial Intelligence Act

A preliminary analysis against the background of Part 3 (Regulating private use of AI) of the ELI's 2020 response to the public consultation

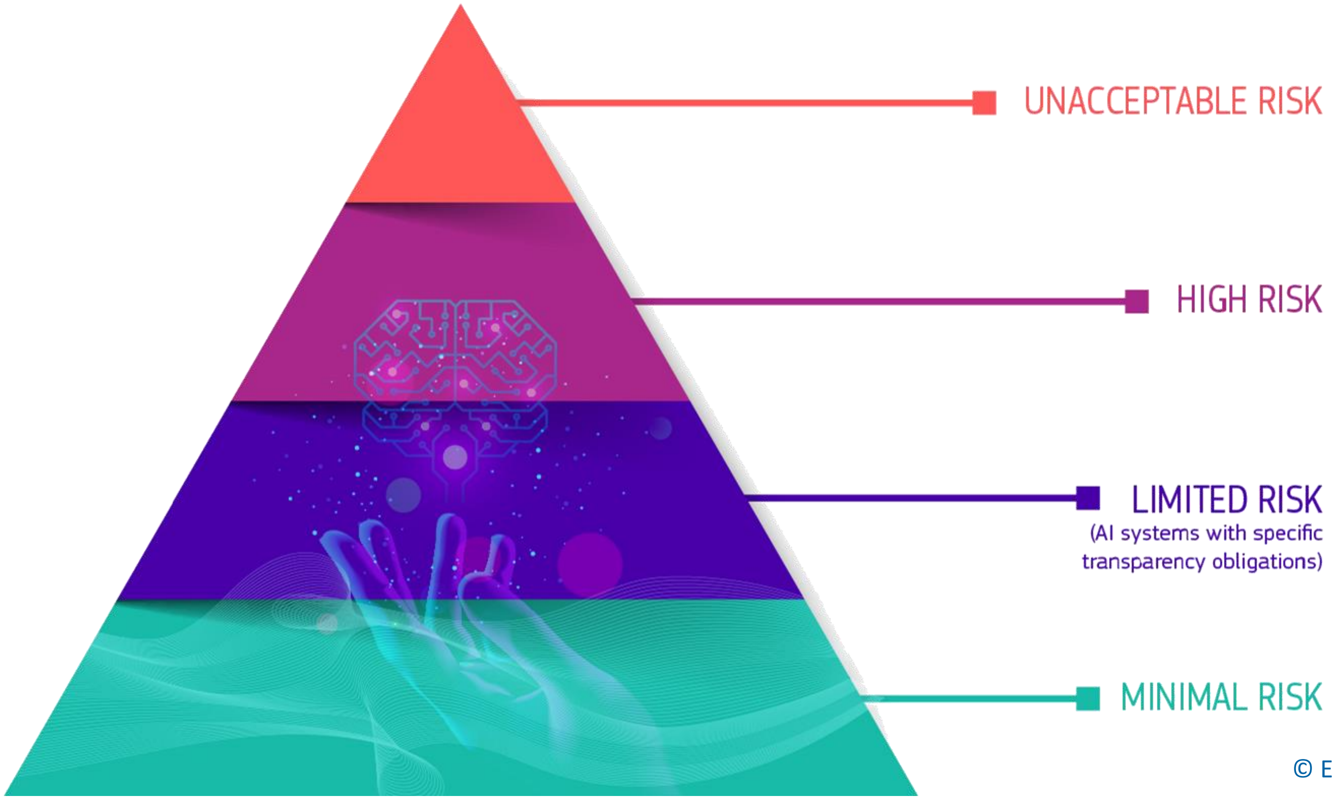
Christiane C. Wendehorst

European Law Institute, 29 April 2021

1

ELI‘s 2020 response had focussed on “squaring the circle of a high level of protection that avoids too much red tape”

The risk-based approach



UNACCEPTABLE RISK

HIGH RISK

LIMITED RISK
(AI systems with specific transparency obligations)

MINIMAL RISK

© European Commission

The risk-based approach

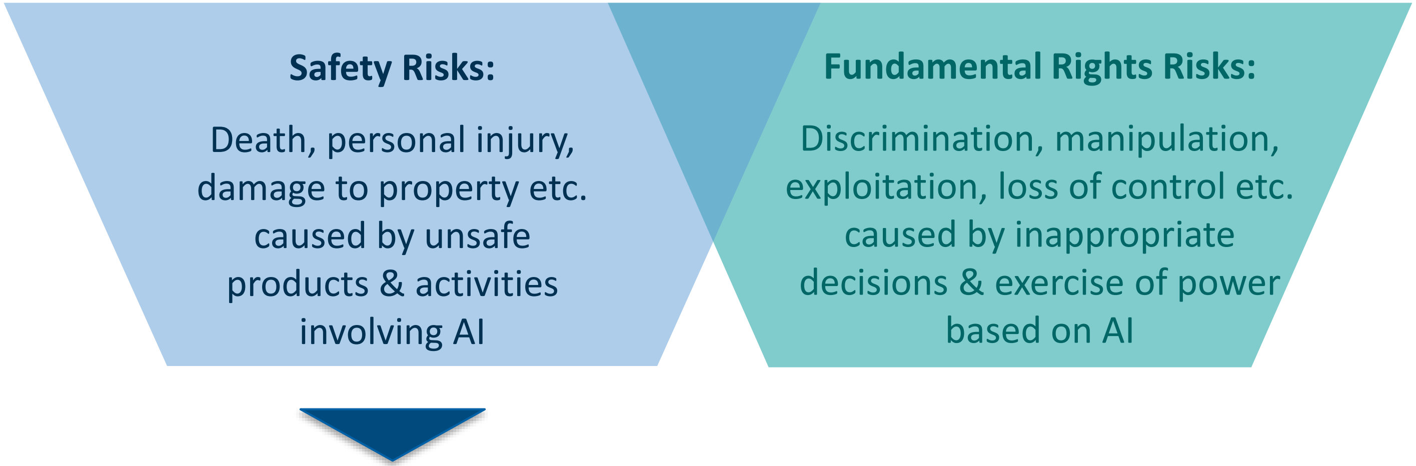
Risk-based approach of the draft Regulation is very positive

However, the details still require discussion, e.g. ‘emotion recognition systems’ as such are only subject to a transparency obligation under Article 52

2

ELI‘s 2020 response stressed the need to link the ‘safety risk’ dimension of AI to existing product safety legislation

AI and safety legislation



Safety Risks:

Death, personal injury, damage to property etc. caused by unsafe products & activities involving AI

Fundamental Rights Risks:

Discrimination, manipulation, exploitation, loss of control etc. caused by inappropriate decisions & exercise of power based on AI

Needs alignment with 'digital fitness check' of existing safety legislation

AI and safety legislation

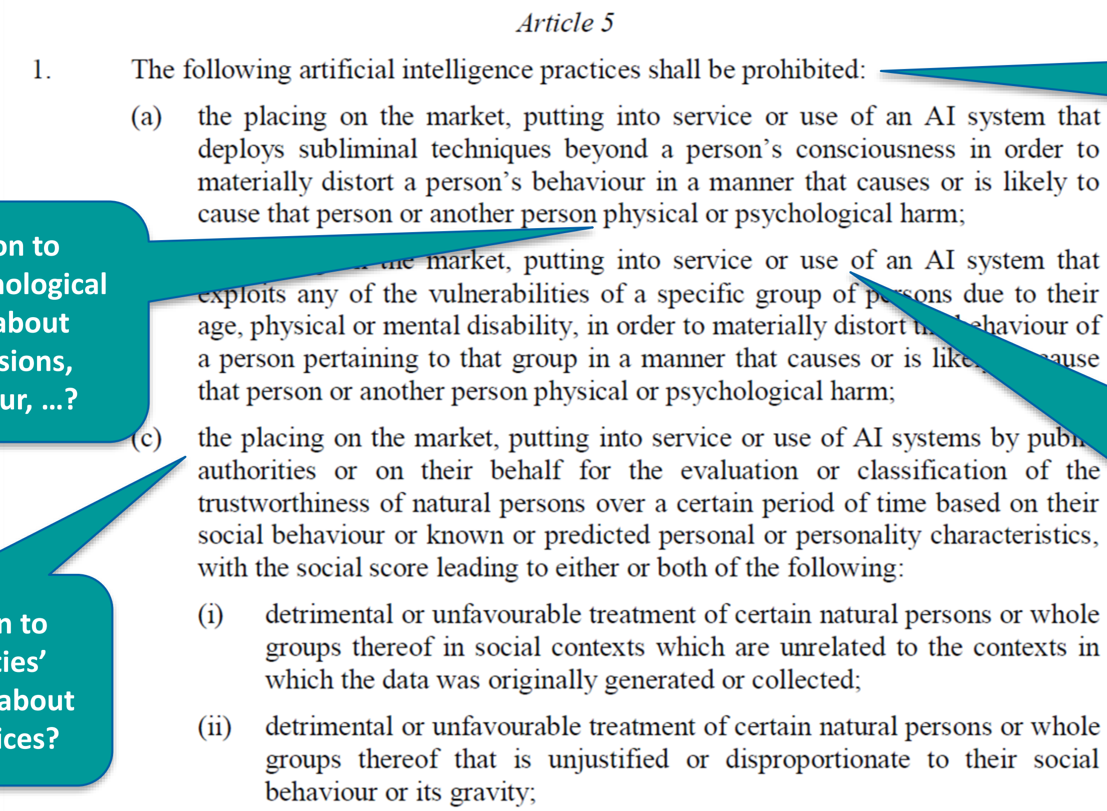
Draft Regulation rightly takes a ‘product safety’ approach for AI, irrespective of whether the AI is embedded or non-embedded software, and also for software-as-a-service

Draft Regulation very elegantly links the new AI-specific requirements with existing safety regulation, while covering also fundamental rights risks (although the relationship with liability remains unclear)

3

ELI‘s 2020 response suggested the prohibition of a set of blacklisted ‘unfair algorithmic practices’, mentioning discrimination, exploitation of vulnerabilities, total surveillance, manipulation …

Prohibited AI Practices



Article 5

1. The following artificial intelligence practices shall be prohibited:

(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;

the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;

(c) the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;

Should maybe 'discrimination' also have been mentioned?

Why restriction to 'physical or psychological harm'? What about economic decisions, voting behaviour, …?

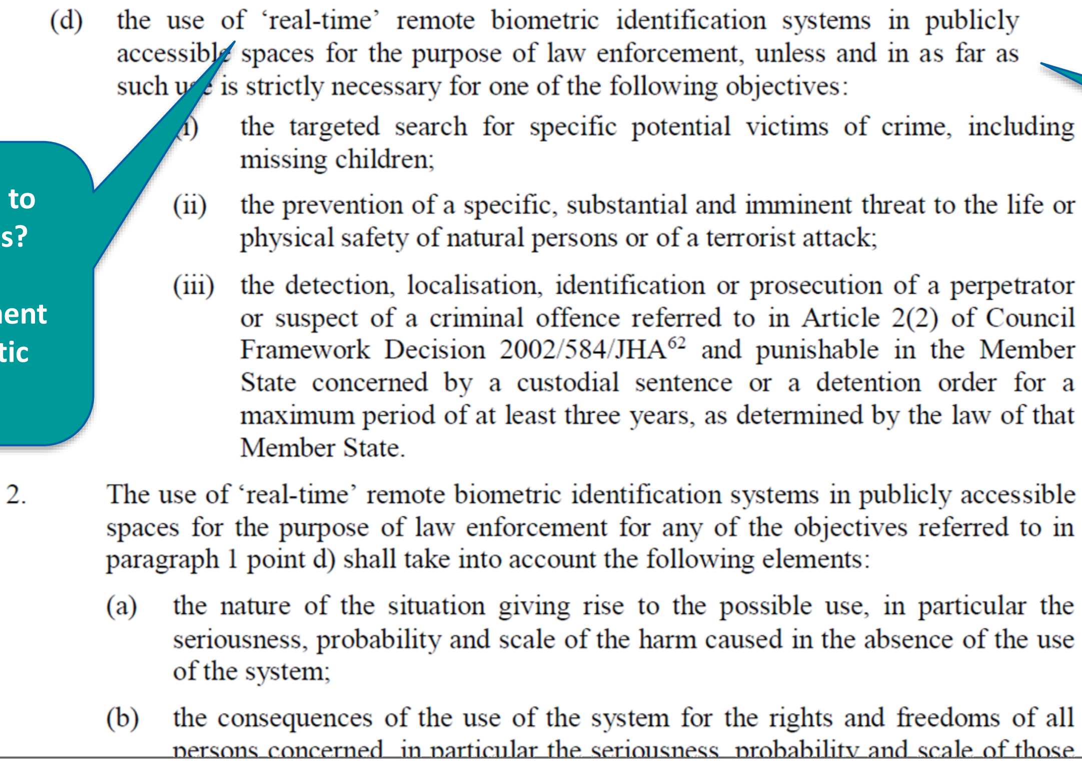
Why only some group-specific vulnerabilities?
Is not exploitation of very individual vulnerabilities at least as dangerous?
And why the restriction to physical or psychological harm?

Is the restriction to 'public authorities' adequate? What about gatekeeper services?

Prohibited AI Practices

Why the restriction to ‘real time’ practices?

And is law enforcement the only problematic purpose?



(d) the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:

(i) the targeted search for specific potential victims of crime, including missing children;

(ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack;

(iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA[62] and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.

2. The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:

(a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system;

(b) the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those

Use of real time remote biometric identification is not really ‘prohibited’ but rather heavily regulated and seems somewhat an alien element in Article 5