Council of the European Union
General Secretariat

Brussels, 07 September 2021

WK 10396/2021 INIT

LIMITE

TELECOM

WORKING PAPER

This is a paper intended for a specific community of recipients. Handling and further distribution are under the sole responsibility of community members.

WORKING DOCUMENT

From:	General Secretariat of the Council
To:	Working Party on Telecommunications and Information Society
Subject:	Artificial Intelligence Act - PowerPoint Presentation : AI Act proposal: Prohibited practices in Art. 5

Delegations will find in annex the PowerPoint Presentation on Artificial Intelligence Act made by the Commission at the Telecommunications and Information Society Working Party on 7 September 2021.

AI Act proposal: Prohibited practices in Art. 5

Kilian Gross
DG CNECT, European Commission

Telecom Council Working Party
September 2021

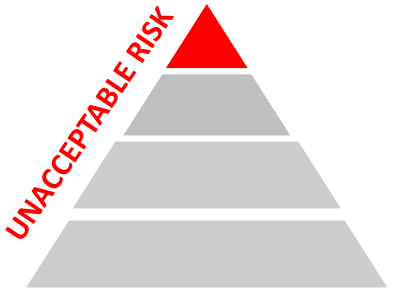
Agenda

This presentation **focuses on AI prohibited practices** in Article 5 of the proposal

1. Rationale
2. Scope of AI Prohibited practices
3. Enforcement
4. a) Harmful subliminal manipulation
5. b) Exploitation of vulnerabilities of specific groups
6. c) 'Social scoring' by public authorities
7. d) Real time RBI for law enforcement purposes

European Commission

Rationale for the AI prohibited practices



- **Top of the risk-pyramid** targeting the AI systems posing unacceptable risks
- Set redlines and clear limits what AI practices we don't want in Europe as **contrary to EU values and fundamental rights**
- Essential for achieving trustworthy AI and preventing misuse of AI for **manipulative, exploitative and social control practices** (recital 15)
- Bring **legal certainty** to providers and users what they should NOT DO with AI
- Follow **well-established regulatory approach** from other pieces of EU legislation (unfair commercial practices)
- **Deliberately narrow** to remain proportionate and not hinder innovation
- **Complementary to other existing EU legislation** (e.g. data protection, consumer protection, non-discrimination)

Scope of the AI prohibited practices

a) Harmful subliminal manipulation

b) Harmful exploitation of vulnerabilities of a specific group of persons

c) 'Social scoring' by public authorities

3 ABSOLUTE PROHIBITED PRACTICES

- '**Placement on the market**', '**putting into service**' and '**use**' of AI systems covered by a), b) and c)
- Applies to both **providers and users**
- Applies to both **public and private** actors (with the exception of social scoring)
- **Protects natural persons** (not only 'consumers' or 'data subjects')
- **No exceptions** from the prohibitions
- **Not affecting research** (before placement on the market/putting into service) if the AI system is not used in human-machine relations that expose people to harm and research is done in line with recognized ethical standards (recital 16)

d) 'Real-time' remote biometric identification for law enforcement purposes in publicly accessible spaces

1 PROHIBITED PRACTICE WITH EXCEPTIONS

- Only '**use**' is prohibited for systems covered by d)
- Applicable only to **'law enforcement authorities'**
- Protects people only in **'physical' publicly accessible spaces**
- **3 exceptions** - MS can decide to allow them or not by national law

Enforcement of the AI prohibited practices

Direct effect of the prohibitions

- The Regulation directly binds all public and private addressees (providers and users)
- The exceptions in art. 5(1)d) do **not** have direct effect and needs to be transposed by national law
- Disciplining **preventive** effect of the high penalties to ensure compliance

May also include injunction to **prevent** imminent placement on the market of an AI application

Public enforcement

- For all practices in art. 5(1) a), b), c) and d) - **ex post** by market surveillance authorities
  - Procedure under Art. 14 and 16 of Regulation (EU) 2019/1020
  - Penalties up to EUR 30 million or 6 % of annual turnover of companies; for public authorities Member States can decide if and to what extent fines can be imposed
- For the exceptions in art. 5(1)d) - **ex ante authorisation** by an independent judicial authority

Private enforcement

- Affected people have no explicit right to complaint and remedies, but complaints can be taken into account by market surveillance authorities (Art. 11(3) of Regulation (EU) 2019/1020)
- The 'direct effect' doctrine would also allow affected people to rely on the prohibitions
- National civil liability rules may also apply

European Commission

Harmful subliminal manipulation - art. 5(1)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'***

Aims to protect:

- Right to human dignity – Art. 1 of the Charter
- Right to physical and mental integrity – Art. 6 of the Charter
- Freedom of thought, conscience and religion – Art. 10 of the Charter

Harmful subliminal manipulation - art.5(1)a)

3 CUMULATIVE ELEMENTS

- The AI system must deploy **subliminal techniques**
  - Affecting people 'beyond their consciousness', in ways that cannot be perceived (recital 16)
  - Concept of 'subliminal' already known and prohibited in art. 9(1)b) AVMSD
- **Intention to materially distort one's behavior**
  - Intention should be to impair one's personal autonomy and ability to act without undue influence, thereby causing him or her to behave in a way he or she would not have behaved otherwise (see also art. 2(e) UCPD)
  - Intention not necessarily to inflict harm, other (commercial) purposes also possible
  - The intention may not be presumed if the distortion results from factors external to the AI system which are outside of the control of the provider or the user (recital 16) – to be interpreted as events that could not be 'reasonably foreseen' by the provider/user, or even if foreseeable, the latter can do nothing to prevent them
- The intended distorted behaviour should cause or be likely to cause **physical or psychological harm**
  - Not necessary for the harm to have occurred, it may be just 'likely' for the operator;
  - It does not need to be only one-off event, harms may also be cumulative and reinforce over time
  - May be physical or psychological to protect the fundamental right to physical and mental integrity
  - May be affecting that or another person, incl. collective harms affecting many people

Harmful subliminal manipulation - art.5(1)a)

FOR EXAMPLE

- Even though the source of the experience is virtual, Extended Reality (XR) applications provide real personal experience that can be highly intensified and persuasive. Misuse of such applications by people who control the sensory experience to incite someone to do something harmful they would not naturally do should be prohibited.

FOR EXAMPLE

- AI-enabled ‘smart’ personal assistant that is optimized to increase economic benefits for certain companies that gives advice for unhealthy diet and/or unhealthy daily regime to consumers.

FOR EXAMPLE

- AI-enabled personalised ‘dark pattern’ embedded in the design interface of a video game reacting in real-time to users’ behaviour that keeps consumers effectively playing excessive time, thus causing them sleep deprivation and anxiety.

Harmful exploitation of vulnerabilities of a specific group of persons- art.5(1)b)

'the placing on 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****'.*

Aims to protect:

- Right to human dignity – Art. 1 of the Charter
- Right to physical and mental integrity – Art. 6 of the Charter
- Freedom of thought, conscience and religion – Art. 10 of the Charter
- Specific rights of vulnerable groups (children, disabled, elderly) – Artt. 24, 25 and 26 of the Charter

Harmful exploitation of vulnerabilities of a specific group of persons - art.5(1)b)

3 CUMULATIVE ELEMENTS

- The AI system must **exploit any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability**
  - Only specific groups are protected – defined by age (children or elderly) and disability (physical or mental)
  - Exploitation of *any* vulnerabilities of those groups is covered - e.g. credulity, inexperience, immaturity, dependency, lack of attention or reflection, lack of self-control, risk-taking behavior, crave for attention, physical incapacity, fragility of the psyche etc.
  - The vulnerabilities can be assessed from the perspective of the average member of that group, but they can be also specific to the concrete AI system and specific context of use in an individual case
- **Intention to materially distort the behavior** of a member of that group
  - Identical as for prohibited practice art.5(1) a)
- The intended distorted behaviour should cause or be likely to cause **physical or psychological harm**
  - Identical as for prohibited art.5(1) a)

Harmful exploitation of vulnerabilities of a specific group of persons - art.5(1)b)

FOR EXAMPLE

- Addictive and compulsive design of AI-enabled applications intended for children i.e., gambling-like random rewards or sending systematic push-notifications when 'off', thus making children progressively dependent and threatening their well-being
  (vulnerability: lack of self-control, dependency, fragility of the psyche, immaturity)

FOR EXAMPLE

- A care robot optimized to make old persons follow their daily routine irrespective of their will, thus causing them psychological harms when applying coercion to this end
  (vulnerability: dependency, fragility of the psyche, reduced physical capacity)

FOR EXAMPLE

- AI system embedded in assistive sensor and visual analysis technologies for disabled persons that misguides the disabled person, thus putting his or her health and life at risk
  (vulnerability: physical incapacity, dependency)

Distinction between practices a) and b)

a) Subliminal manipulation

Any natural person is protected

Techniques are **'subliminal'** affecting people's consciousness in covert/hidden ways

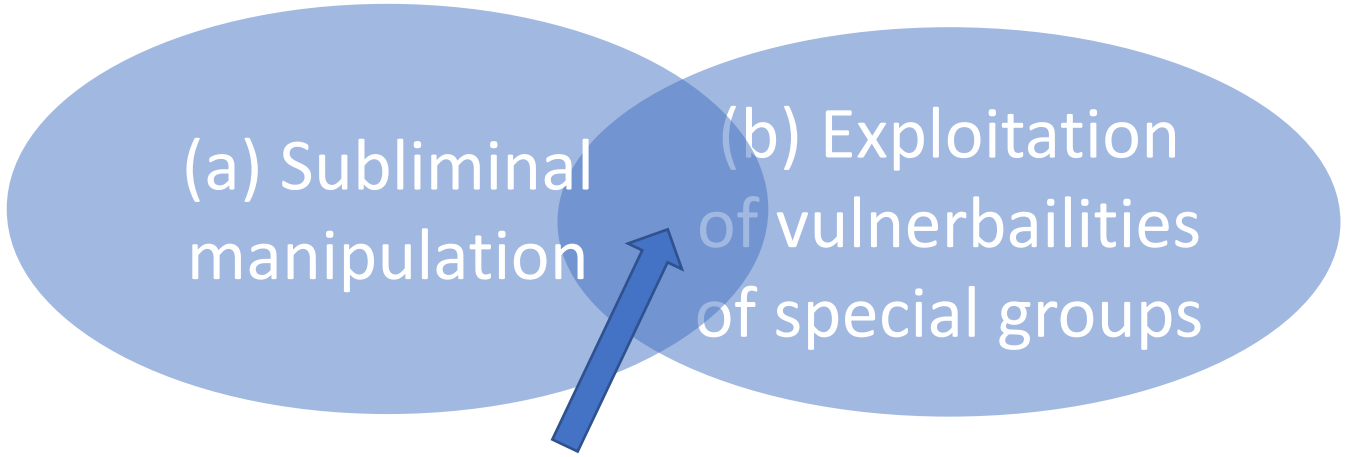
b) Exploitation of vulnerabilities of specific groups

Only **members of specific vulnerable groups** are protected (children, elderly, disabled)

Techniques do not have to be hidden/covert, their **exploitative character** makes them unacceptable even if the person is aware about their use

Relationship between practices a) and b)

- In principle, practices a) and b) have different personal scope and deploy different techniques ('subliminal' or 'exploitative'), but some practices can fulfil simultaneously both conditions



FOR EXAMPLE

- An AI system deployed in a music platform covertly recognises emotions of teenagers and micro-targets those in 'low mood' with recommendations for depressive songs to keep them longer in the platform, thus exacerbating their psychological distress and depression

Relation between the prohibitions in art.5(1)a) and b) and other legislation

COMPLEMENTARITY WITH THE UNFAIR COMMERCIAL PRACTICE DIRECTIVE 2005/29 (UCPD)

- UCPD prohibitions can also apply to AI, but the scope of the directive is limited to business-to-consumer practices and aims to protect primarily the economic interests of consumers
- Art. 3(3) UCPD explicitly states it is without prejudice to EU or national rules relating to the health and safety aspects of products → AIA and UCPD are complementary and apply simultaneously. In case of conflict AIA as lex specialis will prevail.

COMPLEMENTARITY WITH THE DIGITAL SERVICES ACT (DSA)

- The AIA applies also to online platforms when acting as providers/users of AI systems falling within the prohibitions a) and b)
- The DSA includes due diligence and risk management obligations which can help platforms comply with existing prohibitions in EU safety and consumer protection law; transparency obligations for targeted advertising and recommender systems will avoid the risk of subliminal/covert manipulation

COMPLEMENTARITY WITH NATIONAL CIVIL AND CRIMINAL LAWS

- AI prohibited practices can involve conduct already covered by other national legislation
- The AIA adds an uniform layer of protection which can be activated by public enforcement means without necessarily having to start criminal or civil law proceedings.

‘Social scoring’ by public authorities - art.5(1)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’.

Aims to protect:

- Right to human dignity – Art. 1 of the Charter
- Rights to privacy and data protections – Art. 7 and Art. 8 of the Charter
- Rights to equality and non-discrimination – Art. 20 and Art. 21 of the Charter
- Solidarity rights (to social assistance, healthcare etc.) – Art. 34, 35, 36 of the Charter
- Right to good administration – Art. 41 of the Charter

‘Social scoring’ by public authorities- art.5(1)c)

4 CUMULATIVE ELEMENTS

- The AI system must be used for the ‘**evaluation or classification of the trustworthiness** of natural persons’ (‘social score’)
  - ‘Trustworthiness’ the quality of being good, honest, sincere, competent, committed etc. so that people can rely on you (Oxford dictionary)
  - The evaluation must encompass certain period of time, and
  - Be based on person’s social behaviour or known or predicted personal or personality characteristics
- Done ‘**by public authorities or on their behalf’**
  - Focus on public authorities due to power imbalances, public monopoly over certain tasks and services and inevitability of certain interactions of people with public authorities
  - Private actors might also be captured when entrusted with public task or acting on behalf of public authorities
- The score must **‘lead to’** (not necessarily on its own) **-** cause-effect relationship
- **‘Detrimental’** (harmful) **‘or unfavourable’** (not in favour of the person) **‘treatment’** (not only limited to access to services) <u>with 2 alternative or cumulative options:</u>

  i) with data from social contexts unrelated to the contexts in which the data was originally generated/collected, or

  ii) the treatment is unjustified or disproportionate to person’s social behaviour or its gravity

‘Social scoring’ by public authorities- art.5(1)c)

EXAMPLE FOR I) DATA FROM UNRELATED CONTEXTS

- A tax authority targets for fraud inspections people based on AI big data analytics that are, among other sources, also scraping from social networks private data about the daily life and behavior of people

EXAMPLE FOR II) UNJUSTIFIED OR DISPROPORTIONATE TREATMENT

- A social security service determines whether people who have obtained housing benefits are committing fraud based on an AI-enabled risk assessment with the determinant factor being their low water consumption during certain months

EXAMPLE FOR CUMULATIVE APPLICATION OF I) AND II)

- An AI system identifies at-risk children in need of social care based, among others, on factors such as insignificant or irrelevant social ‘misbehavior’ of parents from unrelated contexts (e.g. missing a doctor’s appointment or divorce)

Relation between the prohibition in art. 5(1)c) and other legislation

COMPLEMENTARITY WITH THE EU NON-DISCRIMINATION LAW

- EU non-discrimination law prohibits only unjustified discrimination (direct or indirect) based on an exhaustive list of protected characteristics (gender, race etc.) applicable only to social protection and public services
- The prohibition in art. 5(1) c) ii) prohibits *any* detrimental or unfavourable treatment that is unjustified or disproportionate to the social behavior - it is broader in scope and no need to demonstrate that members of other sex, age, race etc. groups have been treated more favourably

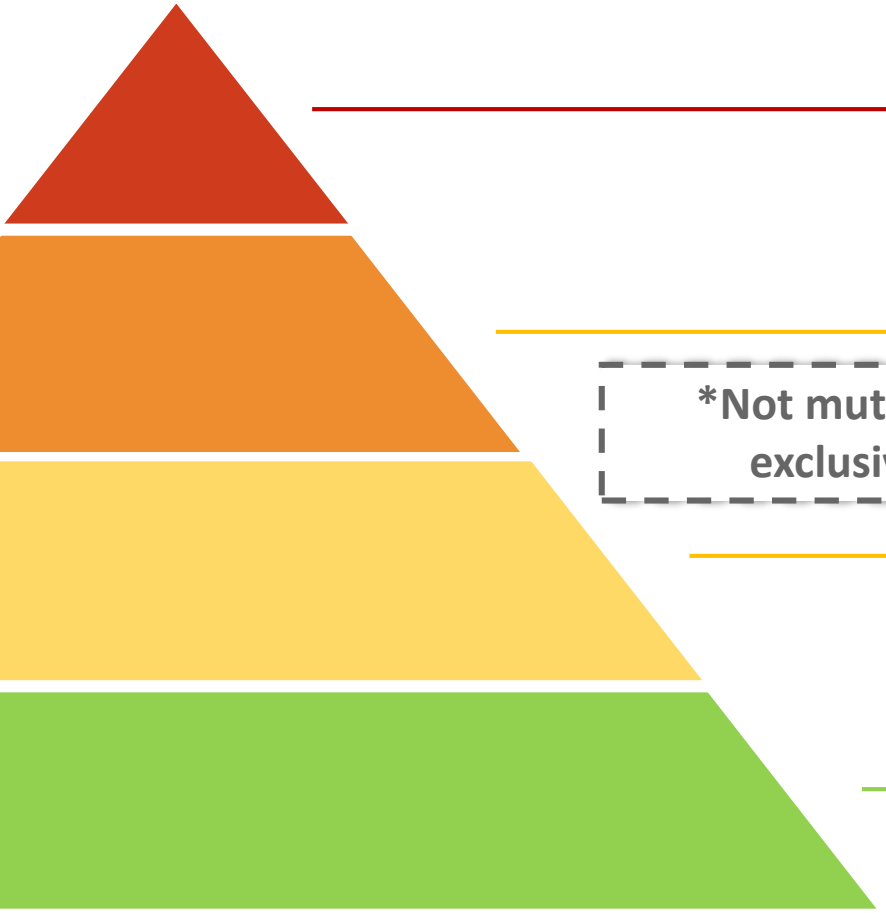
COMPLEMENTARITY WITH THE EU DATA PROTECTION LAW

- GDPR and the Law Enforcement Directive establish key principles for the processing of personal data, including by public authorities (e.g. lawfulness, fairness, purpose limitation, data minimisation)
- The prohibition in art. 5(1)c)i) and ii) complements these general principles with an explicit preventive prohibition of certain unfair social scoring practices violating person's rights to privacy and data protection

COMPLEMENTARITY WITH NATIONAL ADMINISTRATIVE LAW

- AI prohibited practice can involve conduct already covered by national administrative legislation
- The AIA adds an additional layer of protection at EU level which can be activated by public enforcement means regardless of whether national prohibitions exist

Regulatory approach to biometrics



Unacceptable risk
Real-time RBI systems for law enforcement purposes in publicly accessible spaces

Prohibited (with limited exceptions)

High risk
All RBI systems and

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

AI with specific transparency obligations
Emotional recognition and categorisation systems

Permitted but subject to information/transparency Obligations

Minimal or no risk
Biometric authentication/ verification
Closed set identification/ controlled environment

Permitted with no restrictions

European Commission

Real-time biometric identification in law enforcement - art.5(1)d)-(4)

RATIONALE:

Real-time RBI in publicly accessible places may be considered **particularly intrusive** in the rights and freedoms of people because of the constant feeling of surveillance and the immediate impact

AIMS TO PROTECT:

- Right to human dignity – Art. 1 of the Charter
- Rights to privacy and data protections – Art. 7 and Art. 8 of the Charter
- Rights to equality and non-discrimination – Art. 20 and Art. 21 of the Charter
- Rights to assembly and association – Art. 12 of the Charter

Real-time biometric identification in law enforcement - art.5(1)d)-(4)

RATIONALE AND LEGAL BASIS

- Art. 5(1)d) **complements existing data protection law**
  - Article 9 GDPR in principle prohibits the use of biometric systems for identification purposes unless limited exceptions apply
  - GDPR does not apply for law enforcement purposes, which are covered by the Law Enforcement Directive (LED)
  - Article 10 LED allows biometric identification systems (when necessary and subject to safeguards) where authorised by Union or Member State law.

- **Art. 16 TFEU** is the legal basis for Art. 5(1)(d) forbidding certain use of biometric data which applies as lex specialis to Art. 10 of LED.

- Member States are to define detailed national rules for the authorization within the limits of Art 5.

Real-time biometric identification in law enforcement - art.5(1)d)-(4)

ART. 5(1)D) – PROHIBITION/EXCEPTION

- Prohibition of **real-time remote biometric identification** in **publicly accessible spaces** for the purpose of **law enforcement**
- Defined and narrow **exceptions** from the prohibition:

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 the European Arrest Warrant and punished in the Member State concerned for a maximum period of at least three years.

ART. 5(2) – Any use shall

- Take into account **nature of situation** and **consequences on fundamental rights**
- Subject to **appropriate safeguards and limits** in time and space

Real-time biometric identification in law enforcement - art.5(1)d)-(4)

ART. 5(3) – AUTHORISATION OF EXCEPTION

- **Ex-ante authorisation** by judicial authority or independent administrative body
- Only where **necessary and proportionate**
- Exceptional ex-post authorisation in cases of urgency

ART. 5(4) – LEGAL BASIS FOR AUTHORISATION

- Member States' **discretion to fully or partially authorise** use of exceptions
- Member States are to define the necessary detailed rules

Real-time biometric identification in law enforcement - art.5(1)d) - Prohibition

5 CUMULATIVE ELEMENTS FOR THE PROHIBITION

- The deployment must serve **purpose of law enforcement**
  - Prevention, investigation, detection or prosecution of criminal offences: does not include national security or other police work
- Applies to AI systems for the purpose of **biometric identification**
  - Identification vs. authentication/verification: purpose is to identify an individual out of a group of many
- AI system must allow **remote** biometric identification
  - AI systems that operate at a distance in an uncontrolled environment
- AI system must enable **real-time** remote biometric identification
  - Real-time vs. post processing: real-time processing occurs without a significant delay. There are limited opportunities for further checks or corrections
- AI system must be deployed **in a publicly accessible space**
  - Refers to any physical place that is accessible to the public, irrespective of whether privately or publicly owned

Real-time biometric identification in law enforcement - art.5(1)d) - Prohibition

FOR EXAMPLE: COVERED BY ART. 5(1)(D)

- **Real-time identification by law enforcement authority in public spaces**. All faces in the town square captured live by video-protection cameras are cross-checked, in real time, against a database held by the law enforcement agencies.

FOR EXAMPLE: NOT COVERED BY ART. 5(1)(D)

- **Post identification by law enforcement in public spaces**. Analysis of selected video footage (e.g. of a past incident) to identify offenders, for example after the G20 summit in Hamburg.
- **Real-time identification by private users in public spaces.** Since July 2019, Danish football club Brondby IF has been using real-time facial recognition outside its stadium to identify persons that have been banned from attending before reaching the entrance. To ensure approval by the Danish Data Protection Authority, the club has committed to a number of safeguards. For example, the system is prohibited from internet connectivity.

Both use cases would be considered **high-risk** subject to requirements under Title III, and existing **data protection rules** apply.

Real-time biometric identification in law enforcement - art.5(1)d) - Exceptions

FOR EXAMPLE:

- **Search for potential victims of crime**. A child has been kidnapped. Law enforcement authorities may use real-time RBI to quickly find it.

FOR EXAMPLE:

- **Serious threat to life, of injury or terrorism**. There are credible indications (e.g. that at a given football match, there may be a terror attack. With the help of RBI activated in strategic places some hours before the event, possible perpetrators might be identified and stopped before entering the stadium.

FOR EXAMPLE:

- **Serious crime**. There is a credible information that a searched-for murderer was on the train from Madrid to Barcelona. RBI can be triggered for limited period of time at the train station and/or other public spaces the murder is likely to be present at the given time.

Further Biometrics Provisions

Other RBI systems (real-time and post)

Considered **high-risk** requiring an ex ante third party conformity assessment

Emotion Recognition and Biometric Categorisation

Specific **transparency obligations** under Art. 52

Biometric systems can fall under high risk under annex III

~~Biometric authentication and verification technologies permitted with no restrictions.~~

Data protection rules apply