OM HARMONISEREDE REGLER FOR KUNSTIG INTELLIGENS (AI ACT)

Kaj Grønbæk, Professor, formand for ATV's Digitale Vismandsråd

AI ACT - Formål og udfordringer

“Det er i Unionens interesse at bevare EU's teknologiske førerposition og sikre, at europæerne kan drage fordel af nye teknologier, der udvikles og som fungerer i overensstemmelse med Unionens værdier, grundlæggende rettigheder og principper. “

EU har desværre ikke en førerposition på AI….

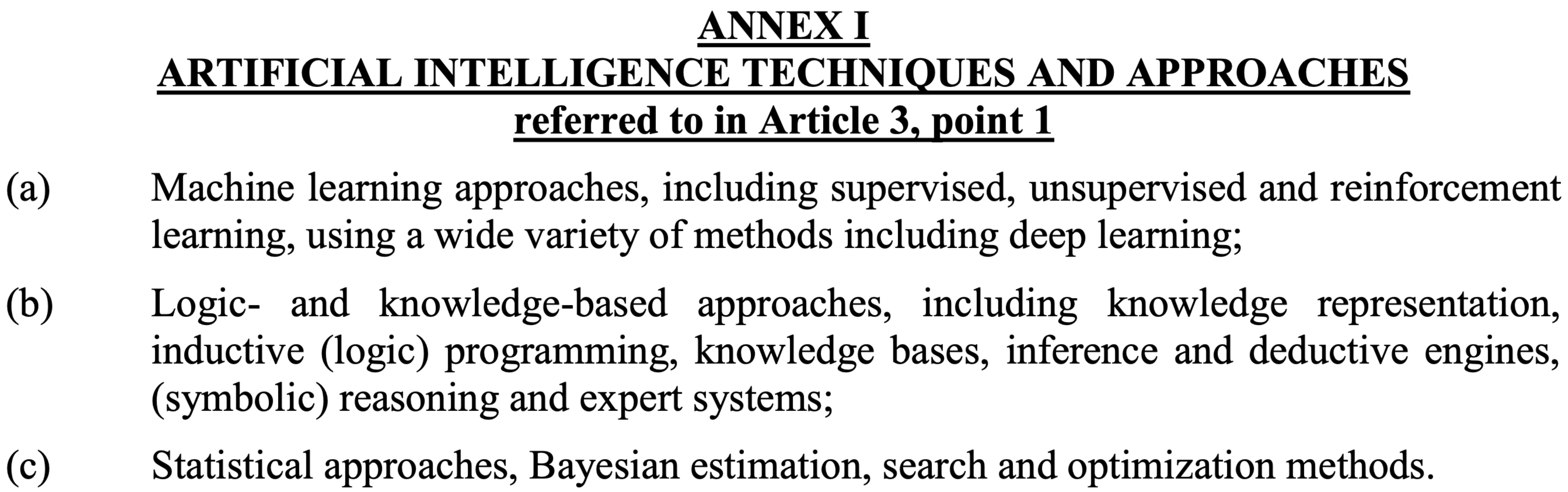
- Vi leverer solide forskningsbidrag – men vi er klart bagefter USA og Kina

Vismandsrådet, forskere og udviklere er bekymrede for, at AI ACT vil stille EU dårligere, derfor stiller vi spørgsmålene:

“Kan vi regulere AI uden at slå de europæiske Tech miljøer ihjel?”

“Og hvordan kan den nuværende formulering af AI Act blive en fordel for EU?”

Definition af Kunstig Intelligens i AI Act



ANNEX I

ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES

referred to in Article 3, point 1

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

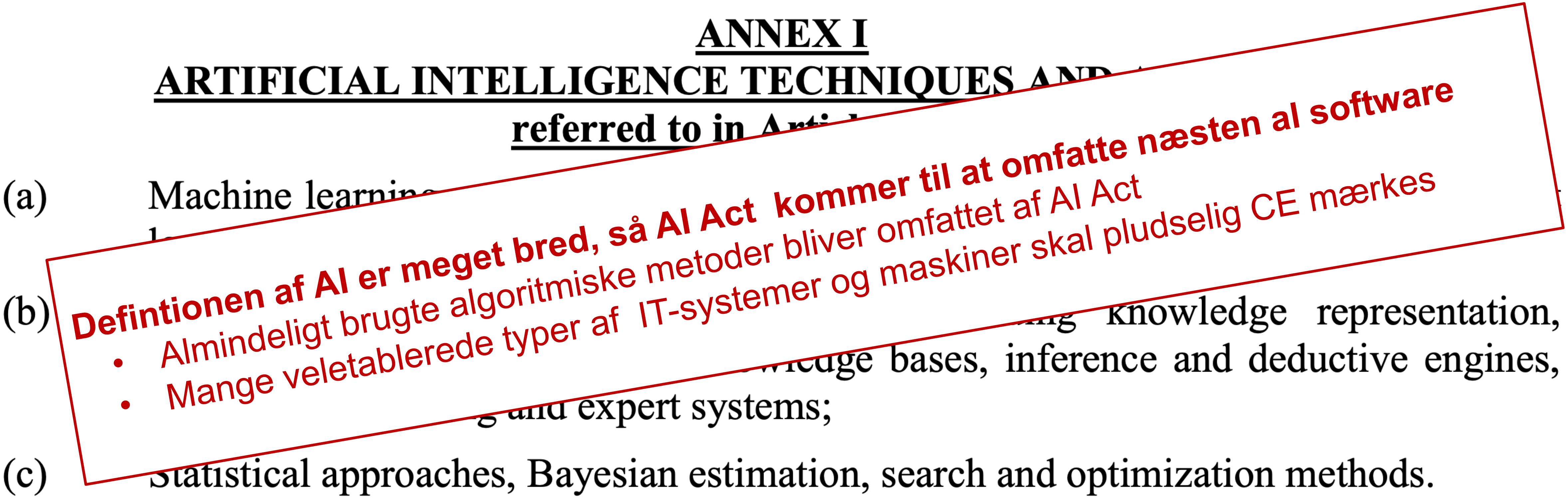
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Definition af Kunstig Intelligens i AI Act

ATV



ANNEX I

ARTIFICIAL INTELLIGENCE TECHNIQUES AND

referred to in Arti

(a) Machine learnin

(b) knowledge representation, wledge bases, inference and deductive engines, g and expert systems;

(c) tatistical approaches, Bayesian estimation, search and optimization methods.

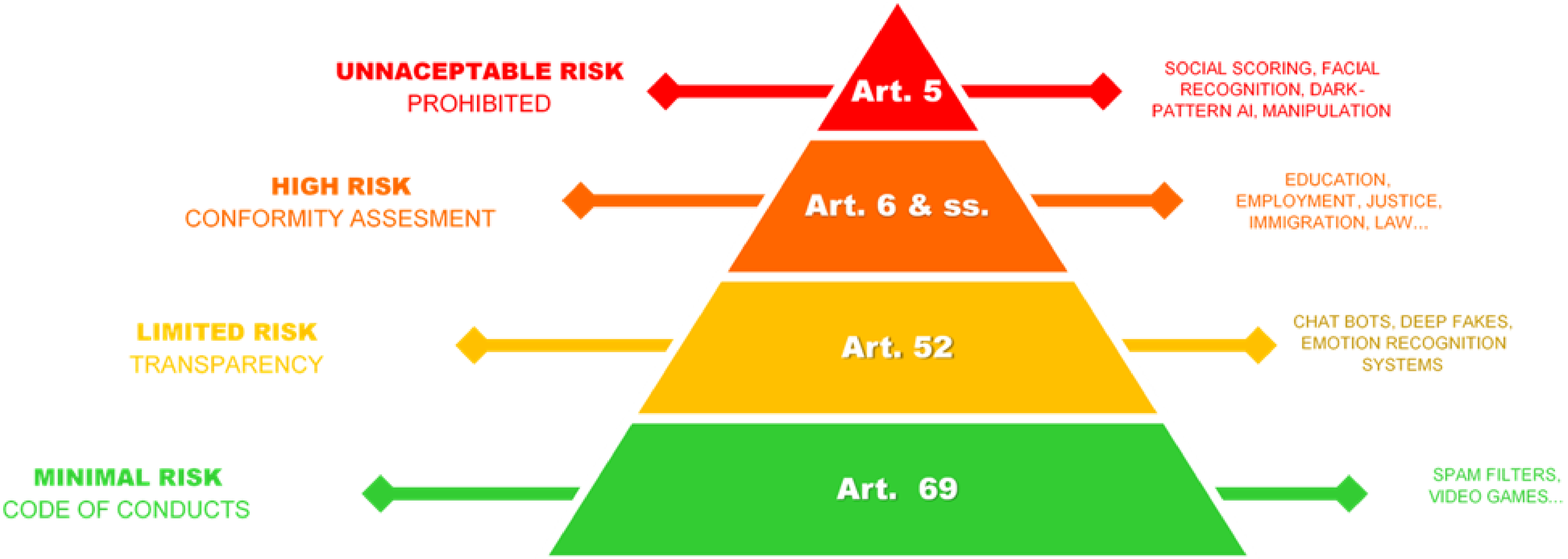
Defintionen af AI er meget bred, så AI Act kommer til at omfatte næsten al software

- Almindeligt brugte algoritmiske metoder bliver omfattet af AI Act
- Mange veletablerede typer af IT-systemer og maskiner skal pludselig CE mærkes

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Risiko-baseret model for regulering

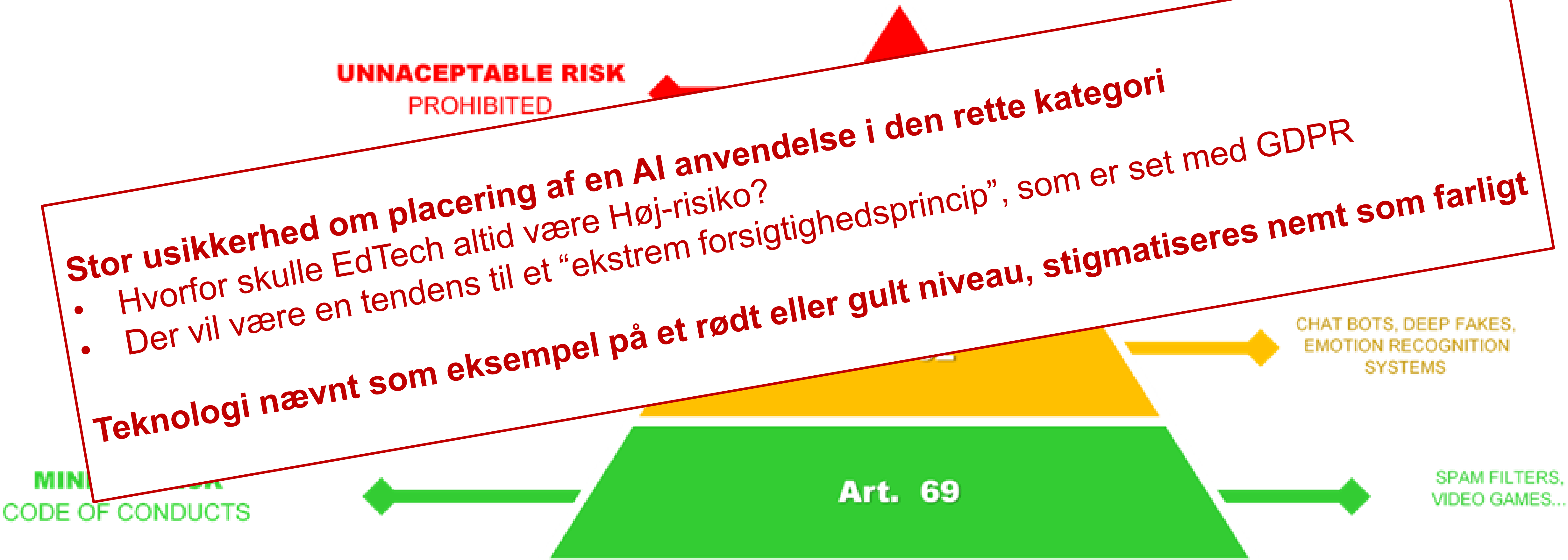
ATV


UNNACEPTABLE RISK
PROHIBITED
Art. 5
SOCIAL SCORING, FACIAL RECOGNITION, DARK-PATTERN AI, MANIPULATION
HIGH RISK
CONFORMITY ASSESMENT
Art. 6 & ss.
EDUCATION, EMPLOYMENT, JUSTICE, IMMIGRATION, LAW...
LIMITED RISK
TRANSPARENCY
Art. 52
CHAT BOTS, DEEP FAKES, EMOTION RECOGNITION SYSTEMS
MINIMAL RISK
CODE OF CONDUCTS
Art. 69
SPAM FILTERS, VIDEO GAMES...

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Risiko-baseret model for regulering

ATV


UNNACEPTABLE RISK
PROHIBITED
CHAT BOTS, DEEP FAKES, EMOTION RECOGNITION SYSTEMS
MINI
CODE OF CONDUCTS
Art. 69
SPAM FILTERS, VIDEO GAMES...

Stor usikkerhed om placering af en AI anvendelse i den rette kategori

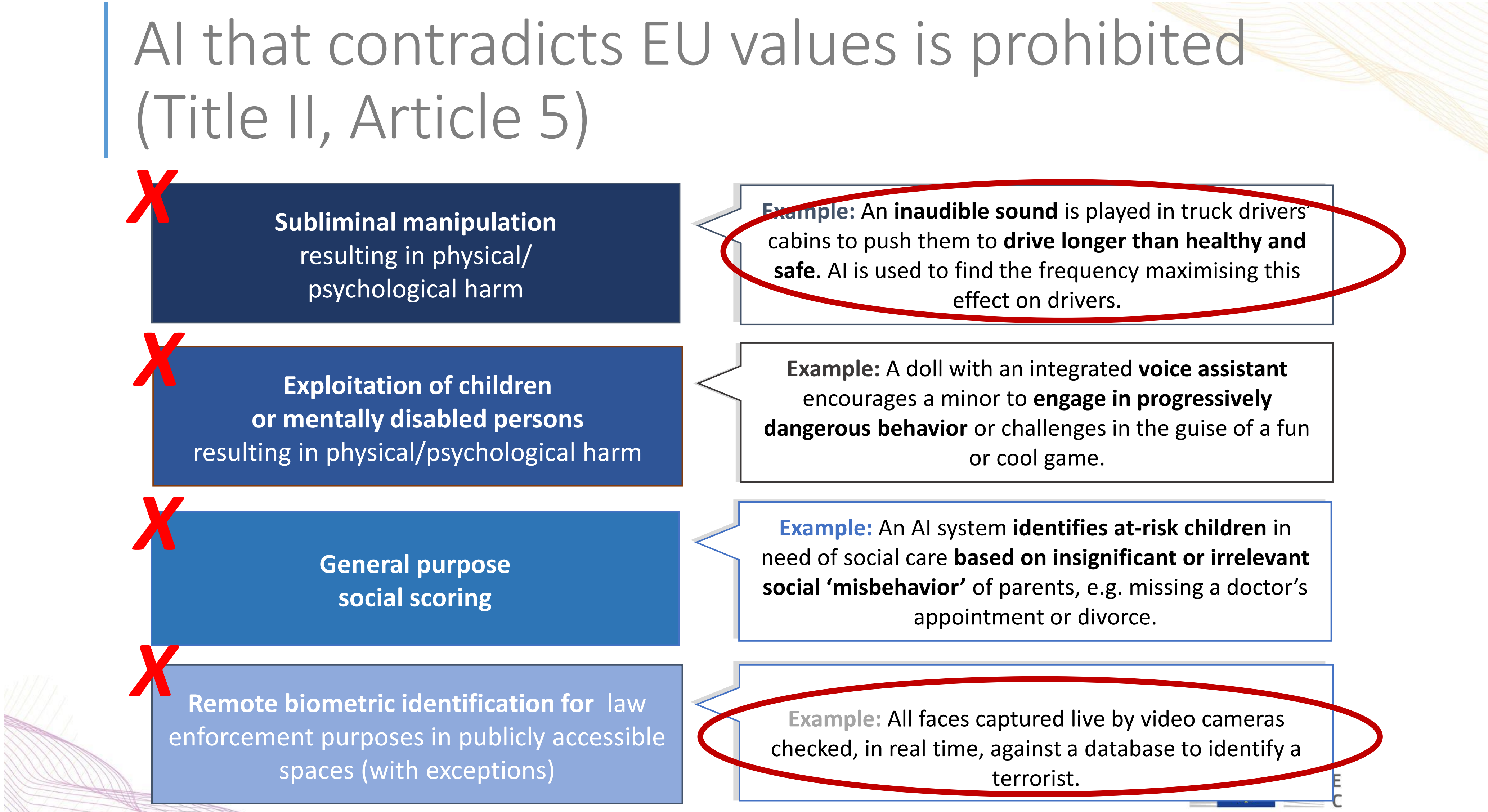
- Hvorfor skulle EdTech altid være Høj-risiko?
- Der vil være en tendens til et “ekstrem forsigtighedsprincip”, som er set med GDPR

Teknologi nævnt som eksempel på et rødt eller gult niveau, stigmatiseres nemt som farligt

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Eksempler på uacceptabel AI I EU materialet

ATV



AI that contradicts EU values is prohibited (Title II, Article 5)

✗ **Subliminal manipulation** resulting in physical/ psychological harm

Example: An **inaudible sound** is played in truck drivers' cabins to push them to **drive longer than healthy and safe**. AI is used to find the frequency maximising this effect on drivers.

✗ **Exploitation of children or mentally disabled persons** resulting in physical/psychological harm

Example: A doll with an integrated **voice assistant** encourages a minor to **engage in progressively dangerous behavior** or challenges in the guise of a fun or cool game.

✗ **General purpose social scoring**

Example: An AI system **identifies at-risk children** in need of social care **based on insignificant or irrelevant social 'misbehavior'** of parents, e.g. missing a doctor's appointment or divorce.

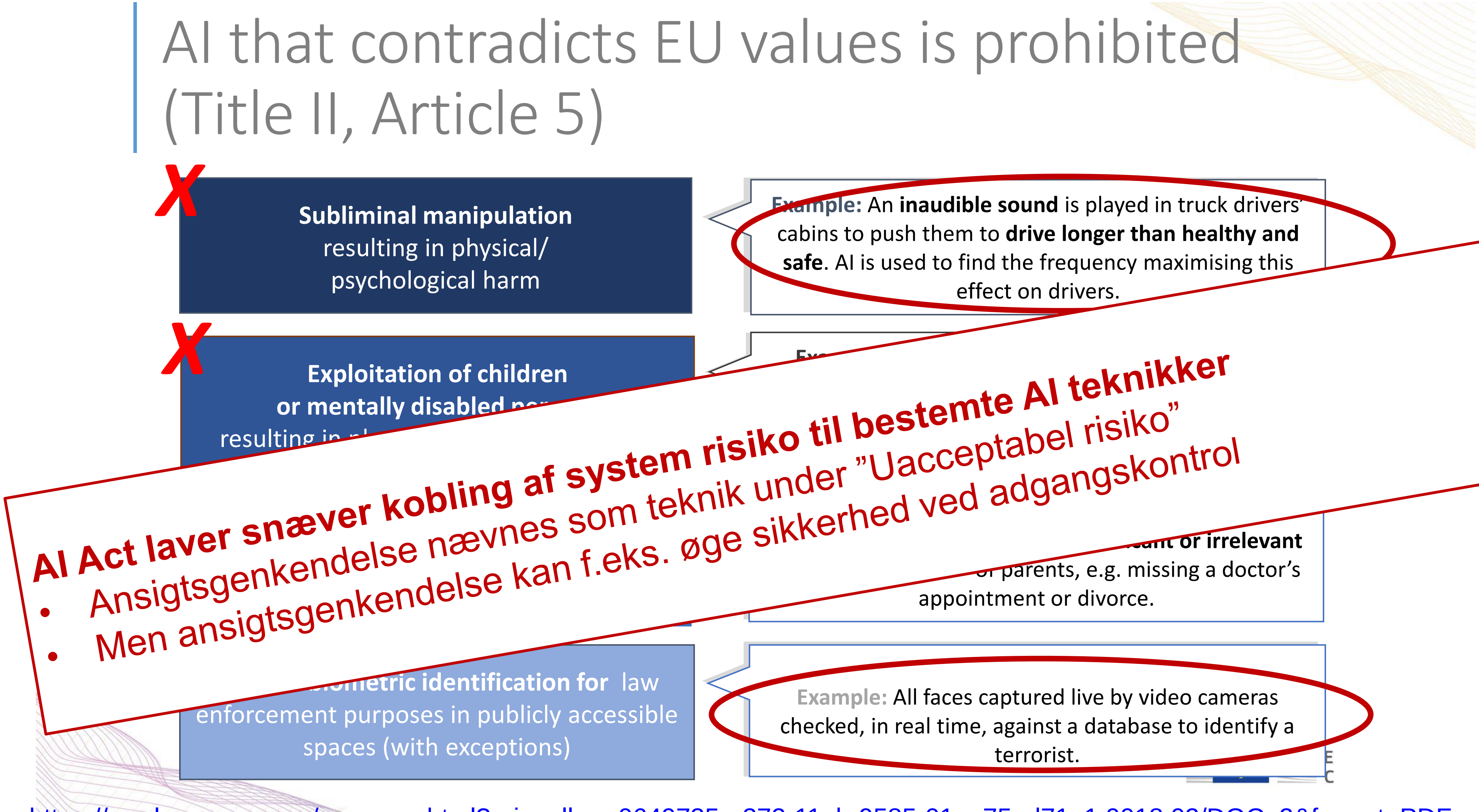
✗ **Remote biometric identification for** law enforcement purposes in publicly accessible spaces (with exceptions)

Example: All faces captured live by video cameras checked, in real time, against a database to identify a terrorist.

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Eksempler på uacceptabel AI I EU materialet

ATV



AI that contradicts EU values is prohibited (Title II, Article 5)

✗

Subliminal manipulation resulting in physical/ psychological harm

Example: An **inaudible sound** is played in truck drivers' cabins to push them to **drive longer than healthy and safe**. AI is used to find the frequency maximising this effect on drivers.

✗

Exploitation of children or mentally disabled per

resulting in pl

Ex

cant or irrelevant

of parents, e.g. missing a doctor's appointment or divorce.

ometric identification for law enforcement purposes in publicly accessible spaces (with exceptions)

Example: All faces captured live by video cameras checked, in real time, against a database to identify a terrorist.

AI Act laver snæver kobling af system risiko til bestemte AI teknikker

- Ansigtsgenkendelse nævnes som teknik under "Uacceptabel risiko"
- Men ansigtsgenkendelse kan f.eks. øge sikkerhed ved adgangskontrol

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Høj-risiko AI

ATV

Kræver CE mærkning

High-risk Artificial Intelligence Systems (Title III, Annexes II and III)

Certain applications in the following fields:

1 **SAFETY COMPONENTS OF REGULATED PRODUCTS**

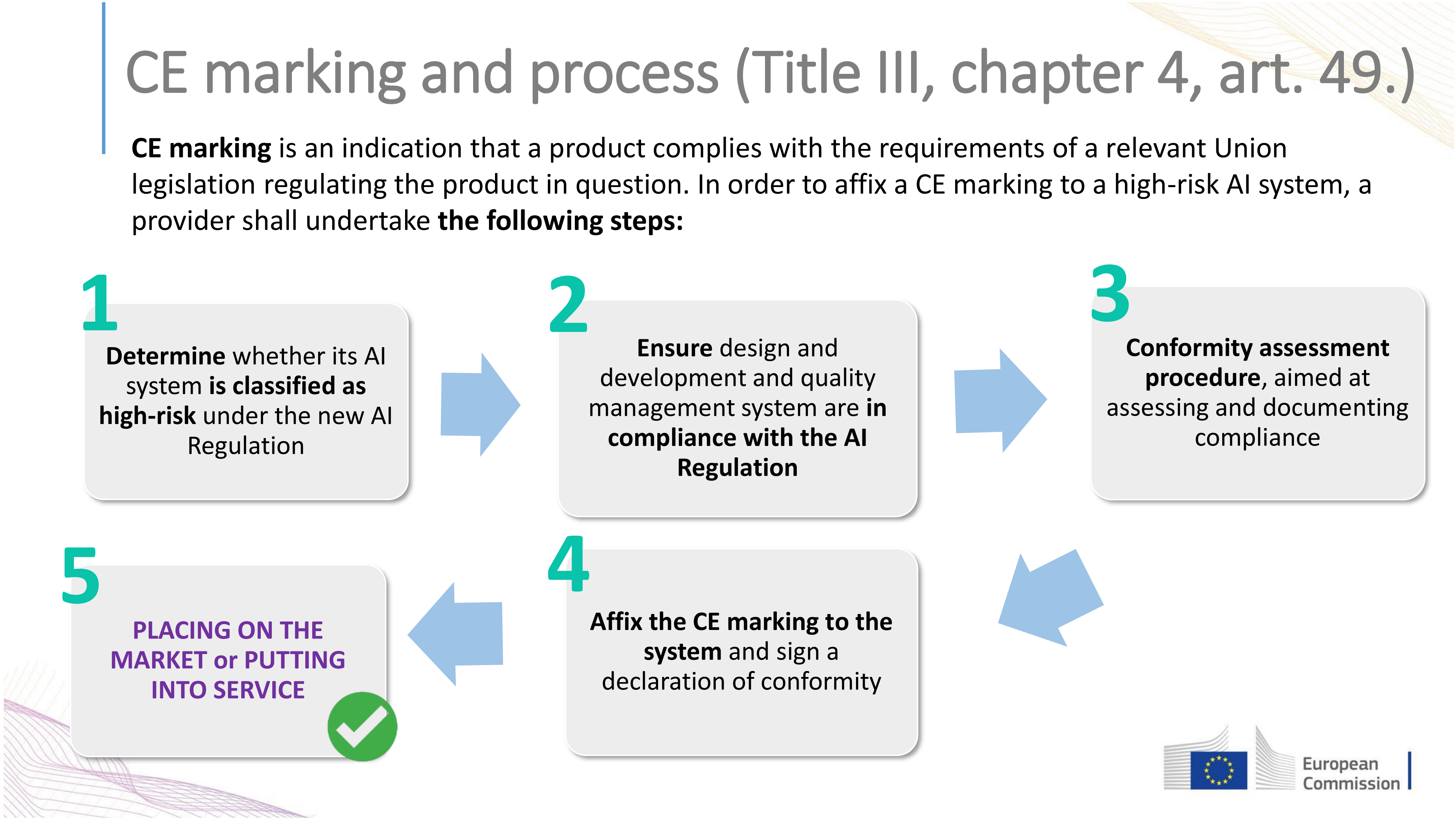
(e.g. medical devices, machinery) which are subject to third-party assessment under the relevant sectorial legislation

2 **CERTAIN (STAND-ALONE) AI SYSTEMS IN THE FOLLOWING FIELDS**

- ✓ Biometric identification and categorisation of natural persons
- ✓ Management and operation of critical infrastructure
- ✓ Education and vocational training
- ✓ Employment and workers management, access to self-employment
- ✓ Access to and enjoyment of essential private services and public services and benefits
- ✓ Law enforcement
- ✓ Migration, asylum and border control management
- ✓ Administration of justice and democratic processes

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

CE mærkningsproces



CE marking and process (Title III, chapter 4, art. 49.)

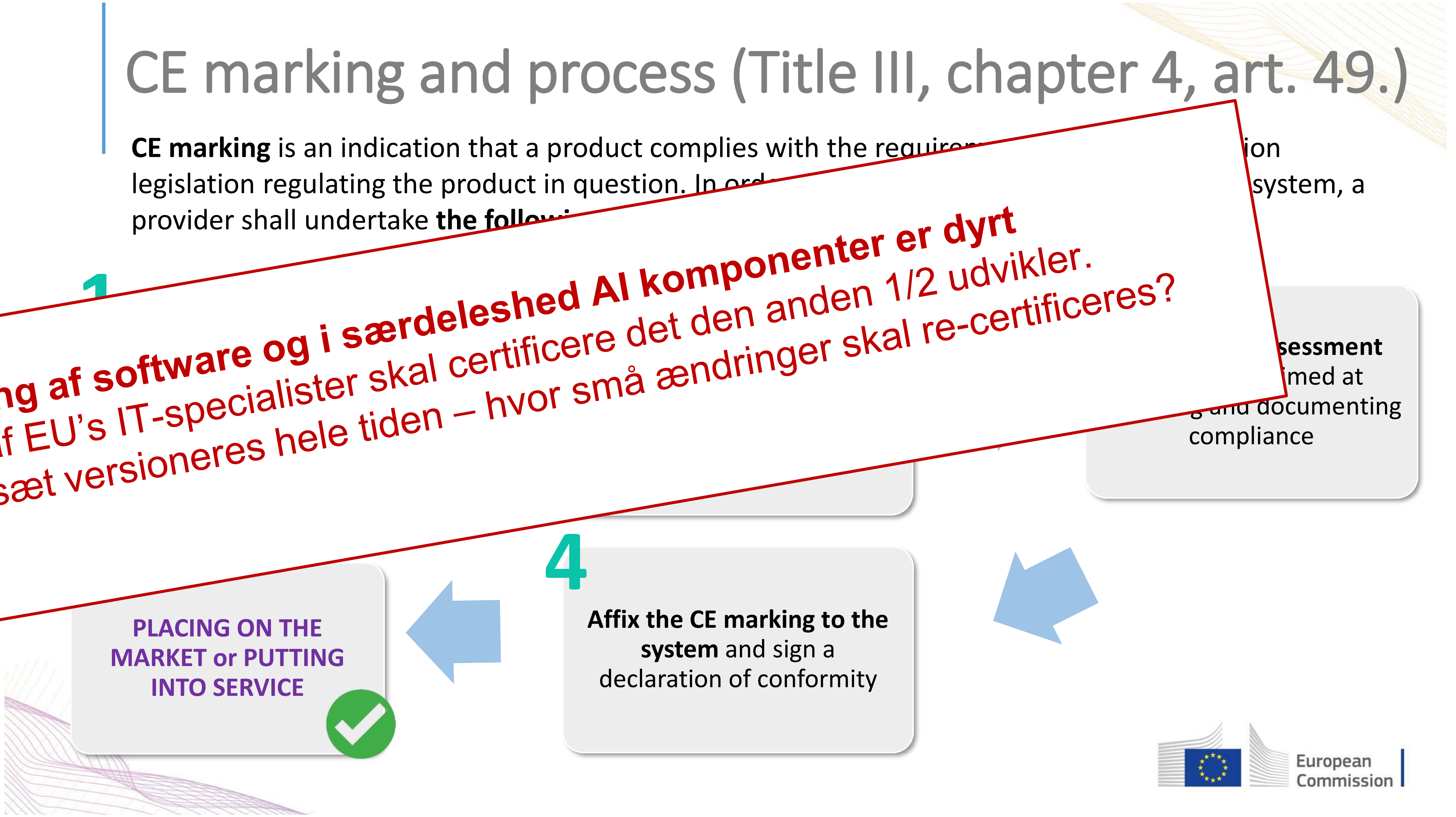
CE marking is an indication that a product complies with the requirements of a relevant Union legislation regulating the product in question. In order to affix a CE marking to a high-risk AI system, a provider shall undertake **the following steps:**

1. **Determine** whether its AI system **is classified as high-risk** under the new AI Regulation
2. **Ensure** design and development and quality management system are **in compliance with the AI Regulation**
3. **Conformity assessment procedure**, aimed at assessing and documenting compliance
4. **Affix the CE marking to the system** and sign a declaration of conformity
5. PLACING ON THE MARKET or PUTTING INTO SERVICE ✓

European Commission

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

CE mærkningsproces



CE marking and process (Title III, chapter 4, art. 49.)

CE marking is an indication that a product complies with the requirem ion legislation regulating the product in question. In ord system, a provider shall undertake **the followi**

Affix the CE marking to the system and sign a declaration of conformity

PLACING ON THE MARKET or PUTTING INTO SERVICE

Detaljeret certificering af software og i særdeleshed AI komponenter er dyrt

- Vi riskerer at 1/2 af EU's IT-specialister skal certificere det den anden 1/2 udvikler.
- Software og datasæt versioneres hele tiden – hvor små ændringer skal re-certificeres?

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0018.02/DOC_2&format=PDF

Det er vigtigt…

- **At værne om etikken i alle *anvendelser* af digitale teknologier - også AI!**
- **At huske, at det er mennesker, der beslutter hvilke *anvendelser*, der skal udvikles**
  - Algoritmer, Data og AI **er ikke I sig selv farlige eller uetiske**
  - Men *anvendelser* kan være farlige eller uetiske
- **At der reguleres på *praksis/anvendelser* uafhængigt af den brugte teknologi**
  - Overvågning, social scoring, forsyningsinfrastruktur mv
- **At bemærke at GDPR faktisk er teknologiuafhængig!**
  - GDPR siger **ikke** noget om type af databasesystemer, der må benyttes
  - Man skal “bare” overholde regler om ***anvendelsen* af persondata**!!
- **At europæiske forskere ikke begrænses unødigt i deres legitime forskning i AI**
  - GDPR reglerne har f.eks. forsinket meget sundhedsdataforskning i Europa og DK

Hvad kunne der gøres i stedet?

- **AI Act bør formuleres teknologiuafhængigt, så man ikke stigmatiserer bestemte AI eller algoritmiske metoder med potentialer i legitime og etisk forsvarlige anvendelser**
  - Nuværende form er en barriere for Europæisk AI R&D, pga detailregulering på teknologi

- **Fokuser på færdselsloven istedet for detailregulering af bilens tekniske dele!**
  - Det er forbudt at køre overfor rødt uafhængigt af om man bruger en tromlebremse eller en skivebremse til at standse bilen med

- **Lav separate Acts for Surveillance, Social scoring mm**
  - **I stedet** for at hægte det op på AI teknologi, fordi der pt findes eksempler på brug af det
  - Ville overvågning med injicerede RFID chips være mere etisk til unik genkendelse af folk end AI baseret biometri?
  - Nej vel? Skal vi så definere RFID chips, som AI for at få denne anvendelse udelukket??

TAK! JEG SER FREM TIL EN GOD DEBAT!