Aandachtspunten inzake de bescherming van grondrechten bij de invoering van een digitale ‘slimme samenleving’

GBA Gegevensbeschermingsautoriteit

APD Autorité de protection des données

DSB Datenschutzbehörde

DPA Data Protection Authority

Hoorzitting over de
Digitale ‘slimme samenleving’

Frank Robben
frank.robben@mail.fgov.be
www.frankrobben.be

Senaat

Structuur van de uiteenzetting

- digitale ‘slimme samenleving’
  - voorbeelden
  - megatrends
  - enkele reflecties
- impact op grondrechten
- private initiatieven tot zelfregulering
- nood aan overheidsregulering ?
  - mogelijkheid en wenselijkheid
  - voorstel van basisprincipes
  - mogelijke aspecten van overheidsregulering
- besluit

Digitale ‘slimme samenleving’

- uitingen
  - robots
  - wearables
  - internet of things (IoT)
  - virtual & augmented reality
  - big data
  - artificial intelligence (AI)
  - …
- componenten
  - infrastructuur en devices
  - gegevens
  - processen
  - netwerken

Komt dichtbij

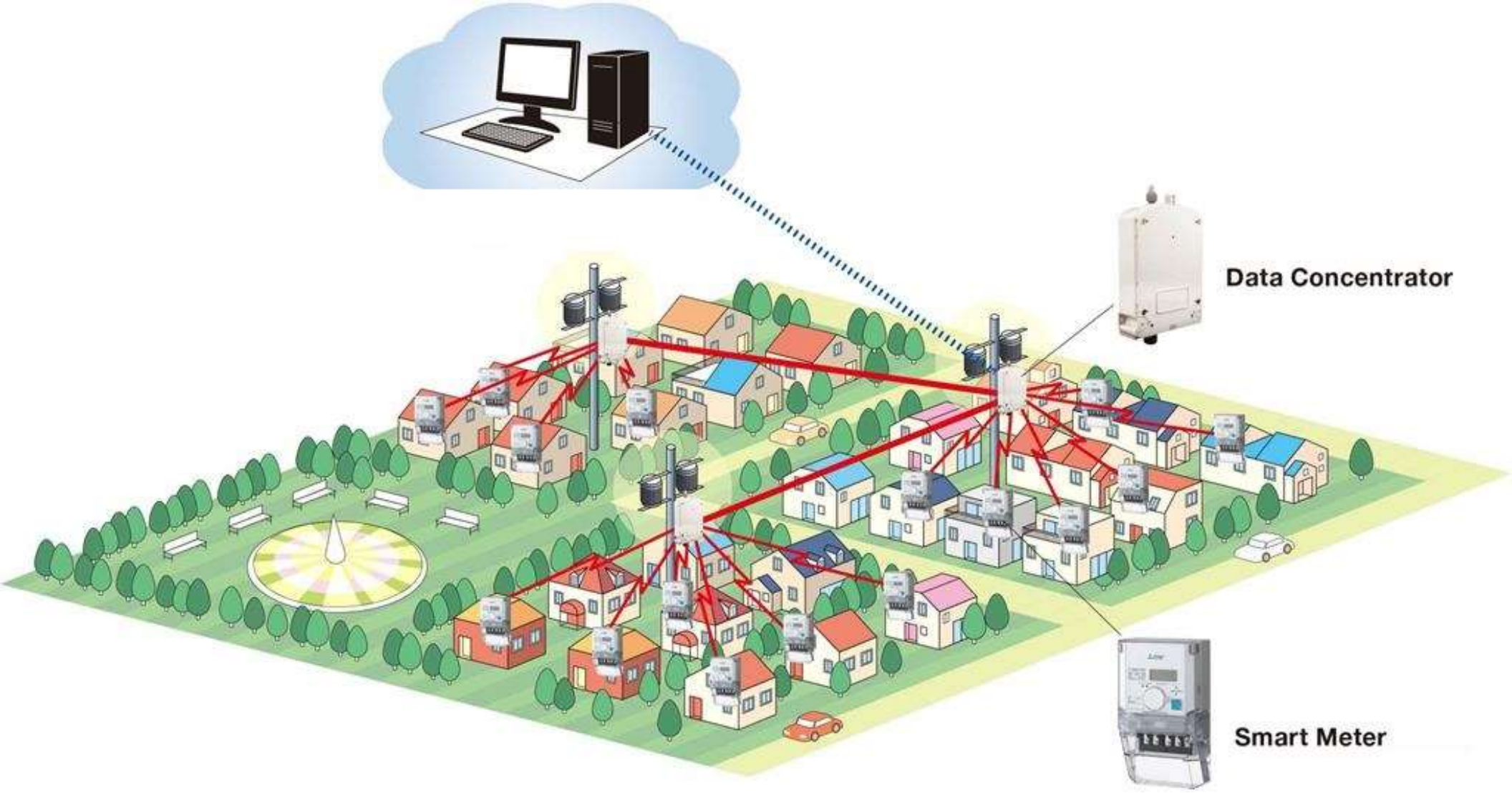
- zelfrijdende auto's
- zorgrobots
- e-coaches
- AI
  - in gezondheidszorg
  - in rechtspraak

Komt dichtbij

Tot 3.000 camera's zullen uw nummerplaat filmen

21 augustus 2018 01:00

Komt dichtbij



Megatrends

- convergentie
  - nanotechnologie
  - biotechnologie
  - informatietechnologie
  - cognitieve technologie

- biologische organismen kunnen gemakkelijk technologisch worden gemanipuleerd

- technologische innovaties krijgen eigenschappen die we voorheen alleen associeerden met levende wezens

Zitten we nog aan het stuur ?

- decentrale autonome organisatie (DAO) en algoritmische bedrijven
  - organisatie die wordt gerund door regels gecodeerd als computerprogramma's, de zogenaamde smart contracts, die worden gehandhaafd op een blockchain
  - nadat een DAO is gestart, kan die zichzelf beheren omdat smart contracts mensen vervangen
  - algoritmische bedrijven maken snellere en slimmere beslissingen door gebruik te maken van adaptieve algoritmes in hun beheer van gebeurtenissen en in hun operaties

- programmeerbare economie
  - wereldwijde integratie van DAO’s en algoritmische bedrijven die de productie en consumptie van goederen en diensten ondersteunt

Zitten we nog aan het stuur ?

- informatieluchtbel (Eli Pariser)
  - resultaat van een gepersonaliseerde zoekopdracht, waarbij een algoritme selectief probeert te bepalen welke informatie de gebruiker zou willen zien, gebaseerd op informatie over die gebruiker (zoals locatie, eerder klikgedrag en zoekgeschiedenis)
  - systemen stemmen dus hun resultaten af op de gebruiker
  - gebruikers krijgen hierdoor geen informatie te zien die hun eigen standpunt tegenspreekt
  - zo worden gebruikers geïsoleerd in hun eigen culturele of ideologische luchtbel

Reflecties

- domein met veel potentieel van ‘dual-use’, zowel goedaardig als kwaadaardig
- complexiteit van algoritmes bemoeilijkt begrip van resultaat
- algoritmes en robots verhogen schaal qua inzetbaarheid => hogere impact (positief en negatief)
- de schaal en snelheid van de evolutie is dermate groot dat onvoorziene gevolgen een hoger risico vormen
- digitale reuzen (Google, Amazon, Facebook, Apple, Microsoft, Snap, Alibaba, Tencent, …) met enorm overwicht deinen uit
  - omzet groter dan BBP van veel landen
  - impact bij alle burgers
  - verwaarloosbare fiscale bijdrage
  - algemene attitude om lokale verzuchtingen naast zich neer te leggen
  - klassieke nationale regelgeving is absoluut niet effectief
- hoe onze burgers beschermen ?

Impact op verschillende grondrechten

- recht op persoonlijke levenssfeer
- recht op menselijke waardigheid
- recht op eigendom (op gegevens, op materiële goederen in virtuele wereld, …)
- vrijheid van meningsuiting
- toegang tot het recht en recht op een eerlijk proces
- bescherming tegen discriminatie
- heldere aansprakelijkheidsregelingen
- …

Impact op recht op persoonlijke levenssfeer

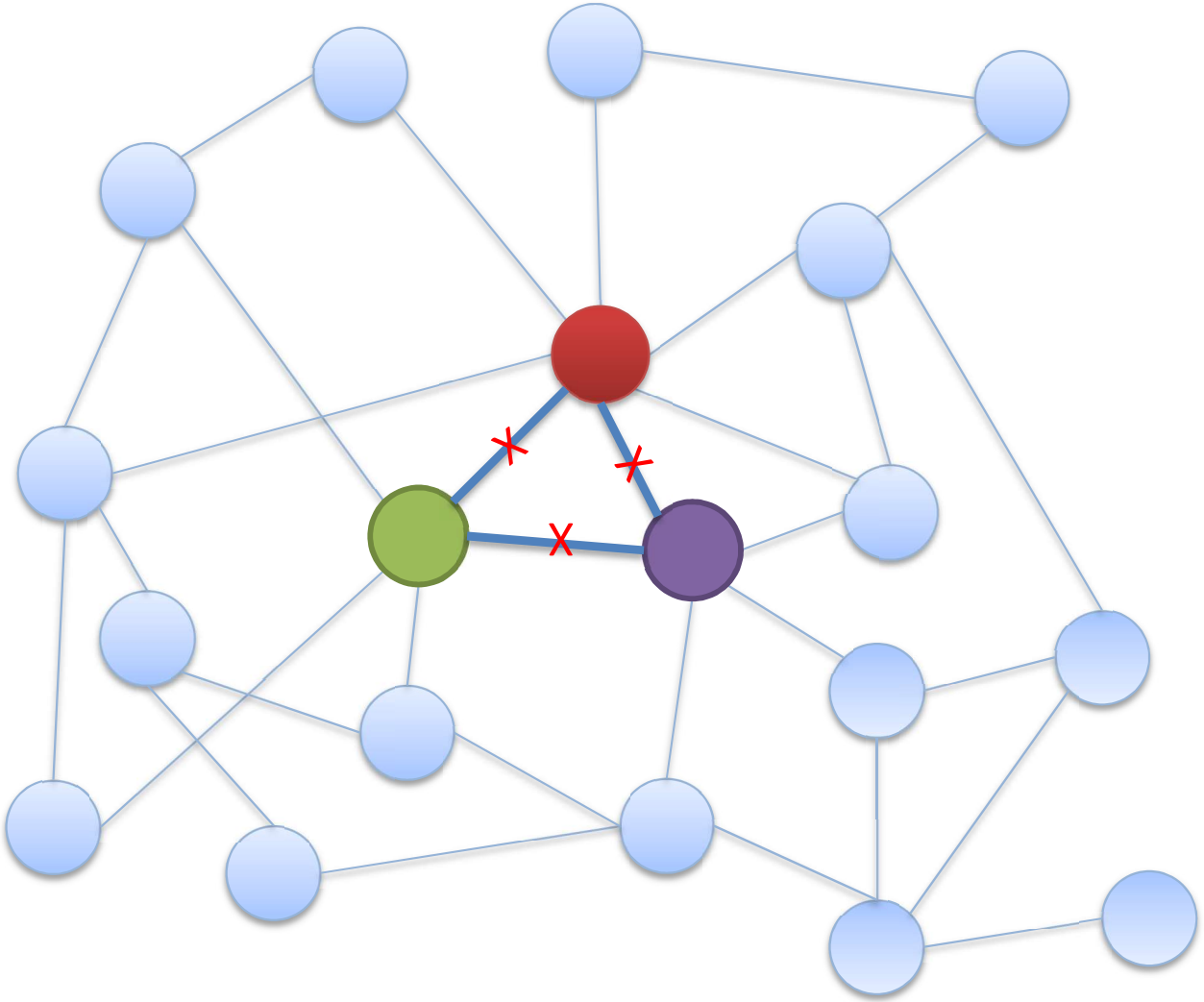
- aantal basisbeginselen blijven valabel
  - doelbinding
  - proportionaliteit
  - transparantie
  - informatieveiligheid

- gratis gebruiksmogelijkheden maakt gebruiker tot product door nood aan financiering uit hergebruik voor andere doeleinden

- informed consent niet meer werkzaam als mogelijke rechtsbasis

Big data

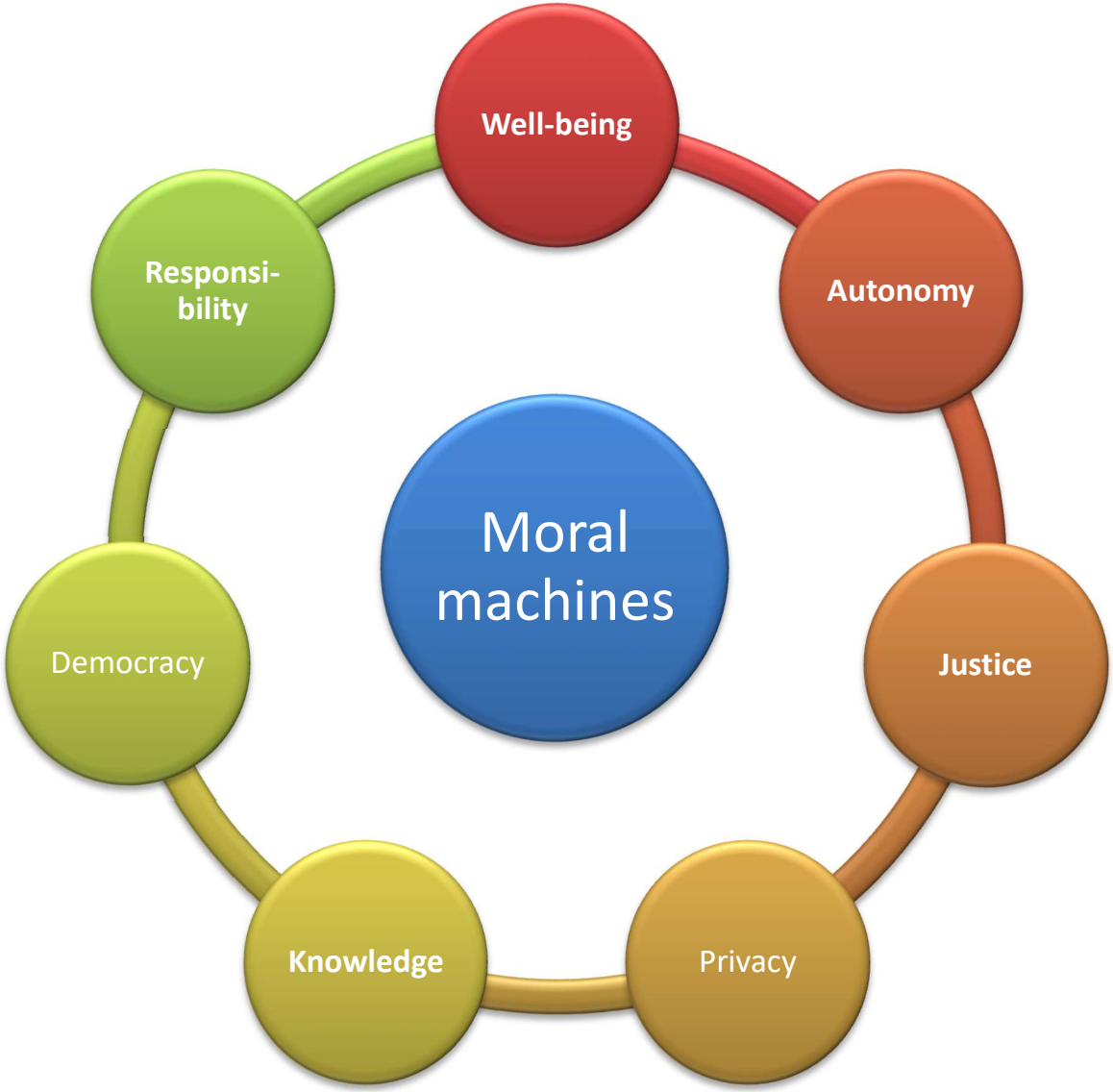
- aantal methoden van risicobeheersing zijn niet meer werkzaam, bv. pseudonimisering en anonimisering



Private initiatieven tot zelfregulering

- voorbeelden
  - Asilomar AI principles: https://futureoflife.org/ai-principles/
  - Montréal Declaration: https://www.montrealdeclaration-responsibleai.com/the-declaration
  - Ethical OS: https://ethicalos.org
- zonder meer positief dat private initiatieven ontstaan met betrekking tot ethische vragen rond nieuwe technologieën
- vertrouwen in de nieuwe technologieën wordt beschouwd als een vereiste voor hun succes
- zal de industrie voldoende zelfregulerend werken ? niet afdoende aangetoond op basis van recent verleden

Ethische AI principes



https://www.montrealdeclaration-responsibleai.com/the-declaration

Voorgestelde ethische AI principes

- **Well-being:** the development of AI should ultimately promote the well-being of all sentient creatures

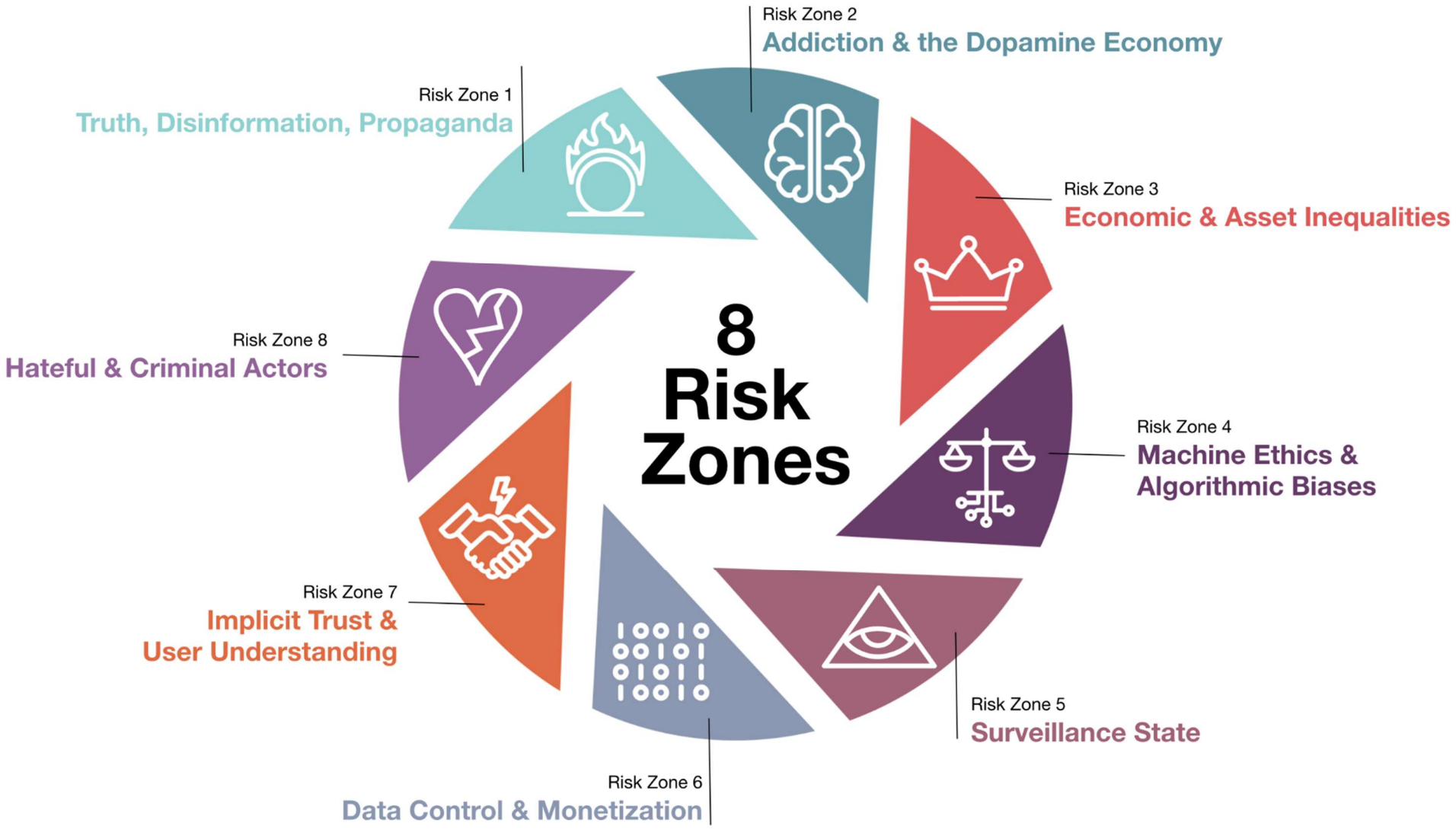
- **Autonomy:** the development of AI should promote the autonomy of all human beings and control, in a responsible way, the autonomy of computer systems

- **Justice:** the development of AI should promote justice and seek to eliminate all types of discrimination, notably those linked to gender, age, mental/physical abilities, sexual orientation, ethnic/social origins and religious beliefs

Voorgestelde ethische AI principes

- **Privacy:** the development of AI should offer guarantees respecting personal privacy and allowing people who use it to access their personal data as well as the kinds of information that any algorithm might use
- **Knowledge:** the development of AI should promote critical thinking and protect us from propaganda and manipulation
- **Democracy:** the development of AI should promote informed participation in public life, cooperation and democratic debate
- **Responsibility:** the various players in the development of AI should assume their responsibility by working against the risks arising from their technological innovations

Reflectie rond 8 risicozones



https://ethicalos.org

Truth, disinformation, propaganda

- what type of data do users expect you to accurately share, measure or collect ?
- how could bad actors use your tech to subvert or attack the truth? what could potentially become the equivalent of fake news, bots or deepfake videos on your platform ?
- how could someone use this technology to undermine trust in established social institutions, like media, medicine, democracy, science ?
- could your tech be used to generate or spread misinformation to create political distrust or social unrest ?
- imagine the form such misinformation might take on your platform; even if your tech is meant to be apolitical in nature, how could it be co-opted to destabilize a government ?

Addiction & dopamine economy

- does the business model behind your chosen technology benefit from maximizing user attention and engagement - i.e., the more, the better ? if so, is that good for the mental, physical or social health of the people who use it? what might not be good about it ?
- what does ‘extreme’ use, addiction or unhealthy engagement with your tech look like ? what does ‘moderate’ use or healthy engagement look like ?
- how could you design a system that encourages moderate use ? can you imagine a business model where moderate use is more sustainable or profitable than always seeking to increase or maximize engagement?
- if there is potential for toxic materials like conspiracy theories and propaganda to drive high levels of engagement, what steps are being taken to reduce the prevalence of that content ? is it enough ?

Economic & asset inequalities

- who will have access to this technology and who won't ? will people or communities who don't have access to this technology suffer a setback compared to those who do ? what does that setback look like ? what new differences will there be between the 'haves' and 'have-nots' of this technology ?
- what asset does your technology create, collect, or disseminate (eg gigs, a virtual currency, health data, deep AI) ? who has access to this asset ? who has the ability to monetize it ? is the asset (or profits from it) fairly shared or distributed with other parties who help create or collect it ?
- are you using machine learning and robots to create wealth, rather than human labor ? if you are reducing human employment, how might that impact overall economic well-being and social stability ?
- are there other ways your company or product can contribute to our collective economic security, if not through employment of people ?

Machine ethics & algorythmic biases

- does this technology make use of deep data sets and machine learning ? if so, are there gaps or historical biases in the data that might bias the technology ?
- have you seen instances of personal or individual bias enter into your product's algorithms ? how could these have been prevented or mitigated ?
- is the technology reinforcing or amplifying existing bias ?
- who is responsible for developing the algorithm ? is there a lack of diversity in the people responsible for the design of the technology ?
- how will you push back against a blind preference for automation (the assumption that AI-based systems and decisions are correct, and don't need to be verified or audited) ?
- are your algorithms transparent to the people impacted by them ? is there any recourse for people who feel they have been incorrectly or unfairly assessed ?

Surveillance state

- how might a government or military body utilize this technology to increase its capacity to surveil or otherwise infringe upon the rights of its citizens ?
- what could governments do with the data you're collecting about users if they were granted access to it, or if they legally required or subpoenaed access to it ?
- who, besides government or military, might use the tools and data you're creating to increase surveillance of targeted individuals ? Whom would they track, why - and do you want your tech to be used in this way ?
- are you creating data that could follow users throughout their lifetimes, affect their reputations, and impact their future opportunities ? will the data your tech is generating have long-term consequences for the freedoms and reputation of individuals ?
- whom would you not want to use your data to surveil and make decisions about individuals, and why not? what can you do to proactively protect this data from being accessible to them ?

Data control & monetization

- do your users have the right and ability to access the data you have collected about them ? how can you support users in easily and transparently knowing about themselves what you know about them ?
- if you profit from the use or sale of user data, do your users share in that profit ? what options would you consider for giving users the right to share profits on their own data ?
- could you build ways to give users the right to share and monetize their own data independently ?
- what could bad actors do with this data if they had access to it ? what is the worst thing someone could do with this data if it were stolen or leaked ?
- do you have a policy in place of what happens to customer data if your company is bought, sold or shut down?

Implicit trust & user understanding

- does your technology do anything your users don't know about, or would probably be surprised to find out about? if so, why are you not sharing this information explicitly - and what kind of backlash might you face if users found out ?
- if users object to the idea of their actions being monetized, or data being sold to specific types of groups or organizations, though still want to use the platform, what options do they have ? is it possible to create alternative models that build trust and allows users to opt-in or opt-out of different aspects of your business model moving forward ?
- are all users treated equally ? if not - and your algorithms and predictive technologies prioritize certain information or sets prices or access differently for different users - how would you handle consumer demands or government regulations that require all users be treated equally, or at least transparently unequally ?

Hateful & criminal actors

- how could someone abuse your technology to bully, stalk, or harass other people ?
- what new kinds of ransomware, theft, financial crimes, fraud, or other illegal activity could potentially arise in or around your tech ?
- do technology makers have an ethical responsibility to make it harder for bad actors to act ?
- how could organized hate groups use your technology to spread hate, recruit, or discriminate against others ? what does organized hate look like on your platform or community or users ?
- what are the risks of your technology being weaponized ? what responsibility do you have to prevent this ? how do you work to create regulations or international treaties to prevent the weaponizing of technology ?

Zijn private initiatieven voldoende ?

- aantal internationale of supranationale organisaties en landen nemen initiatieven
  - EU adviesorganen
    - European Group on Ethics and Science: brede algemene scope waarvan ICT er maar één is, geen permanente aandacht voor ICT

      zie http://ec.europa.eu/research/ege/index.cfm
    - European AI Alliance

      zie https://ec.europa.eu/digital-single-market/en/european-ai-alliance
  - UK: Centre for data ethics and innovation

    zie https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation

- soortgelijk initiatief te overwegen ? of impact verwerven in EU adviesorganen ?

Is ‘slim’ reguleerbaar ?

Is 'slim' reguleerbaar ?

The Regulatory Problems of Artificial Intelligence

The Discreetness Problem
AI projects could be developed without the large-scale, integrated institutional frameworks needed by most 20th century industrial institutions.

The Diffuseness Problem
AI projects can be developed by a diffuse set of actors operating in a diffuse set of locations and jurisdictions.

The Discreteness Problem
AI projects will capitalise on or make use of discrete technologies and components, the full potential of which will not be apparent until the components come together.

The Opacity Problem
The technologies underlying AI will tend to be opaque to most potential regulators.

The Foreseeability Problem
AI can be autonomous and operate in ways that are unforeseeable by the original programmers. This will give rise to a potential 'liability gap'.

The Narrow Control Problem
An AI could operate in ways that are no longer under the control of those who are legally responsible for it.

The General Control Problem
An AI could elude the control of all human beings. This is the problem alluded to by the likes of Nick Bostrom in *Superintelligence*.

The Definitional Problem

What is Artificial Intelligence? You need to define the object of regulation, but AI admits of no easy definition. Modern preference for defining AI in terms of 'thinking' and 'acting' rationally tends toward under and over inclusivity

Ex Ante Challenges
Problems with the research and development of Artificial Intelligence

Ex Post Challenges
Problems with the creation and implementation of Artificial Intelligence

Matthew Scherer - **REGULATING ARTIFICIAL INTELLIGENCE SYSTEMS: RISKS, CHALLENGES, COMPETENCIES, AND STRATEGIES** -*Harvard Journal of Law & Technology*

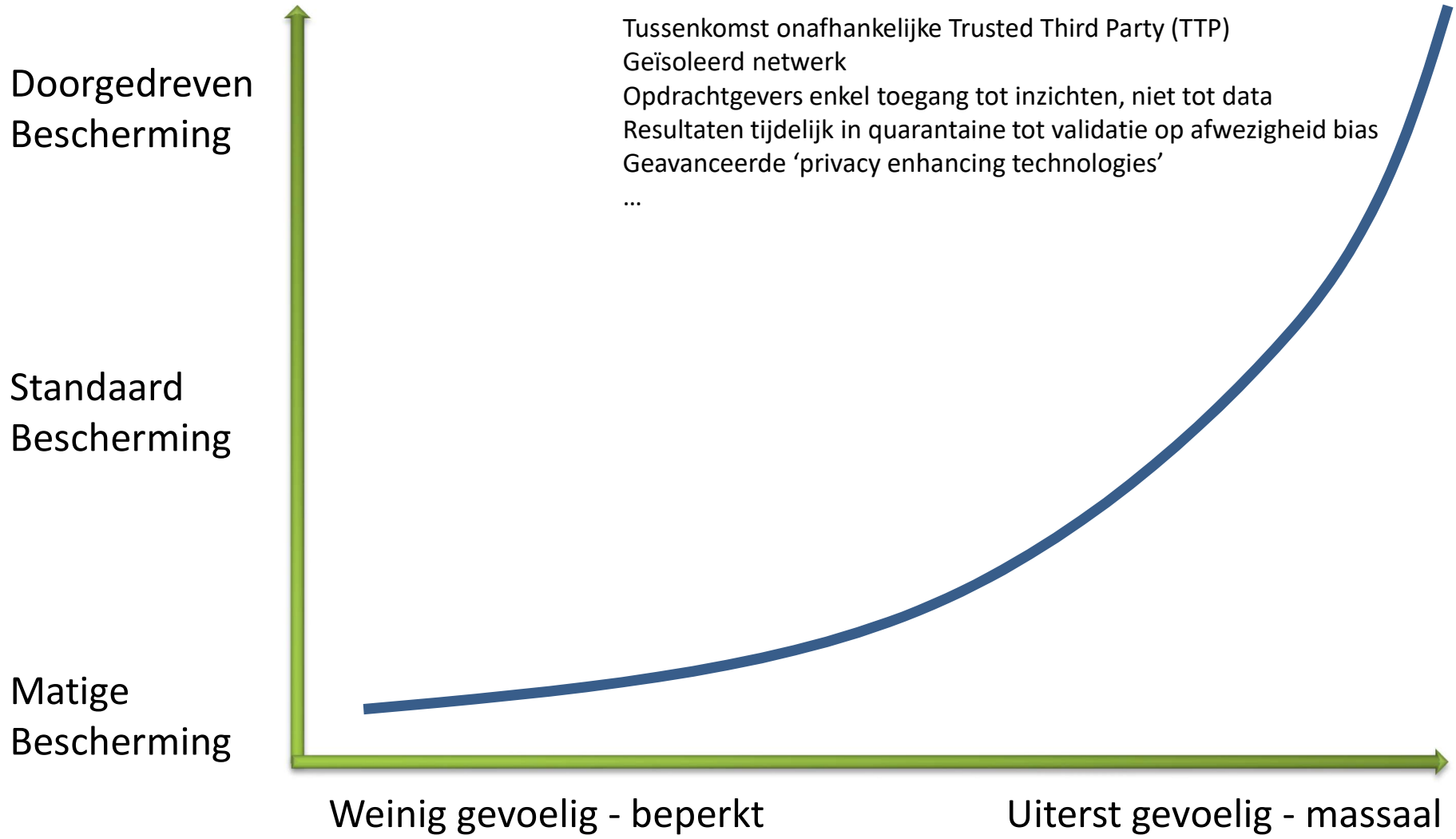
Is ‘slim’ reguleerbaar ?

- moeilijkheid om AI te definiëren
- AI is vaak autonoom en werkt op creatieve wijze (bv. reinforcement learning) => mogelijke negatieve gevolgen zijn moeilijk tot niet voorzienbaar
- kan AI in bepaalde omstandigheden zo gaan werken dat er geen controle meer op bestaat of kan bestaan ?
- AI-initiatieven kunnen quasi onopgemerkt starten, kunnen verspreid zijn over meerdere staten, of zijn gebeurlijk onmogelijk in hun werking te doorgronden

Basisprincipes van eventuele regulering

- opportuniteiten benutten, risico's vermijden
- internationale regeling en handhaving
- duurzame, technologieneutrale regeling
- eerder organisatorische systeembenadering dan snel verouderende inhoudelijke principes ('moral machines')
- nadruk op by design aanpak

Doorgedreven gebruik vergt doorgedreven beschermingsmaatregelen



Mogelijke aspecten van regulering

- verplichtingen voor aanbieders
  - by design maatregelen, zoals
    - middelen van regie door gebruikers
    - motivering bij beïnvloeding van gedrag
    - aanmoediging van menselijk contact
  - verantwoordelijkheid voor algoritmes/resultaat
  - transparantie over algoritmes en open source
    - mogelijkheid tot controle
    - mogelijkheid tot aanpassing
    - quid complexiteit ?
  - kwaliteit van de gegevens
  - data destruction policies
  - bewaking door onafhankelijk ethicus

Mogelijke aspecten van regulering

- rechten van betrokkenen
  - permanente sensibilisering en vorming tot verantwoord gebruik
  - recht op transparantie
  - nieuwe rechten (zie voorstel Rathenau Instituut in rapport aan Raad van Europa: https://www.rathenau.nl/nl/digitale-samenleving/mensenrechten-het-robottijdperk)
    - recht niet gemeten, geanalyseerd of gecoacht te worden
    - recht op betekenisvol menselijk contact

Mogelijke aspecten van regulering

- informatieveiligheid
  - gebruikers- en toegangsbeheer
  - (dynamische her)vercijfering
    - in motion
    - at rest
    - in use
  - bescherming van integriteit van infrastructuur, netwerken, gegevens en processen
  - auditeerbaarheid
  - vermijden van onrechtmatig gebruik (intentioneel of niet, door externen en internen)
    - periferiebeveiliging
    - immuniteit van systemen
  - beroep op trusted third party

Besluit

- naast de enorme potentiële voordelen van een digitale ‘slimme samenleving’ is er aandacht nodig voor ongewenste of onbedoelde negatieve effecten
- door private initiatieven uitgewerkte ethische kaders kunnen inspireren
- het leeuwendeel van de toepassingen zal multinationaal zijn
  - best ageren binnen een internationaal of supranationaal kader
  - beleid uitwerken om invloed te hebben op de multinationele actoren
- een gedetailleerde inhoudelijke juridische regeling is noch mogelijk, noch wenselijk

Besluit

- ‘sandbox’ waar zowel technologische, ethische als juridische impact van innovatie kan worden getoetst ?
- officieel observatorium dat kan helpen de risico’s en de potentiële negatieve gevolgen van de digitale ‘slimme’ samenleving op te vangen ?
- nieuw beleidsdomein om dit fenomeen samenhangend en effectief op te vangen (Europees Commissaris van ‘digitale multinationals’) ?
- parallel met farma-industrie ?
  - goedkeuringsproces van algoritmes
  - ‘testing’ fase

frank.robben@mail.fgov.be

@FrRobben

https://www.ksz.fgov.be
https://www.ehealth.fgov.be
https://www.frankrobben.be