La légitimité de la standardisation de l'équité par le Règlement (UE) sur l'intelligence artificielle

Jerome De Cooman

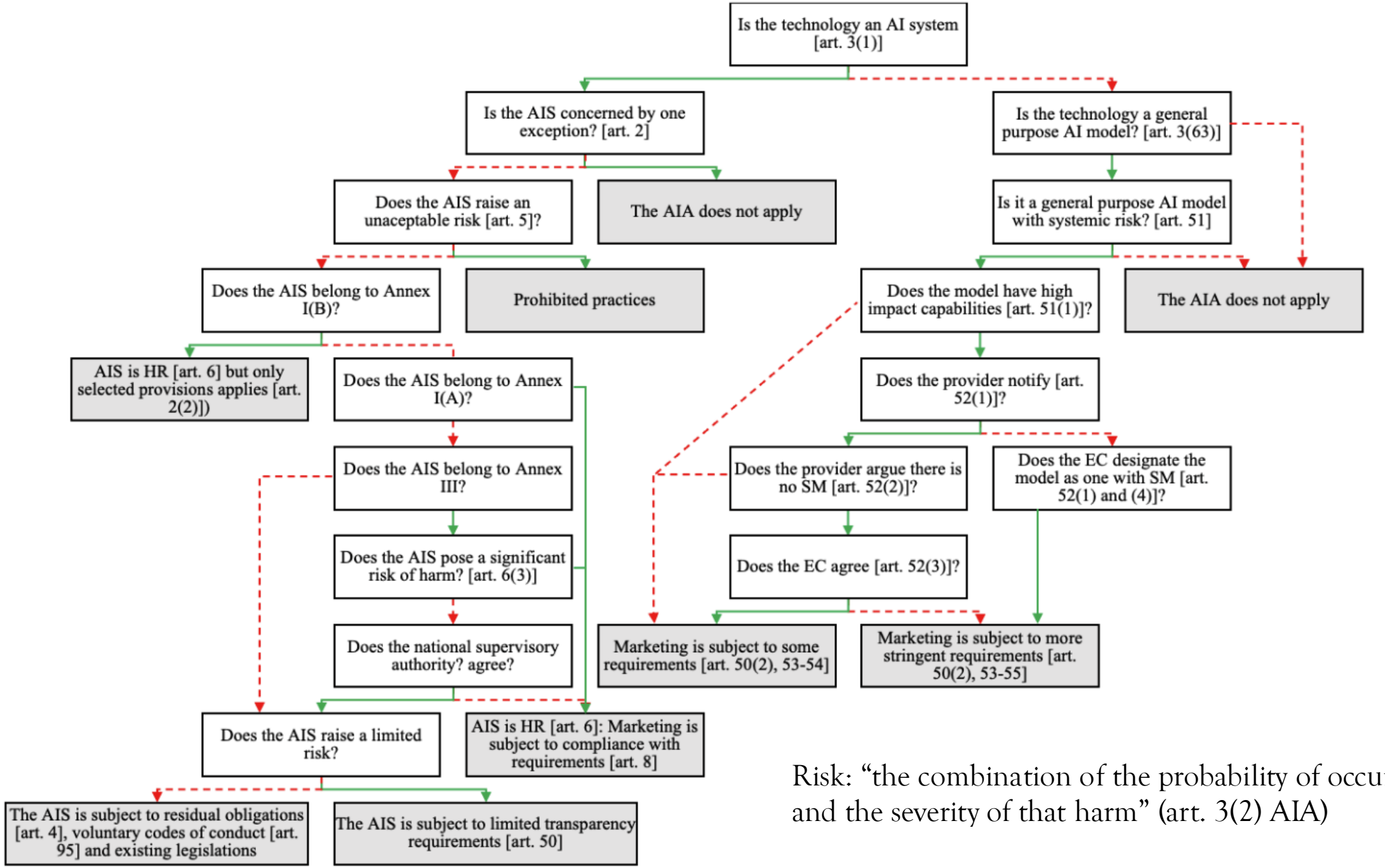
jerome.decooman@uliege.be

26 mai 2025, Paris

Content

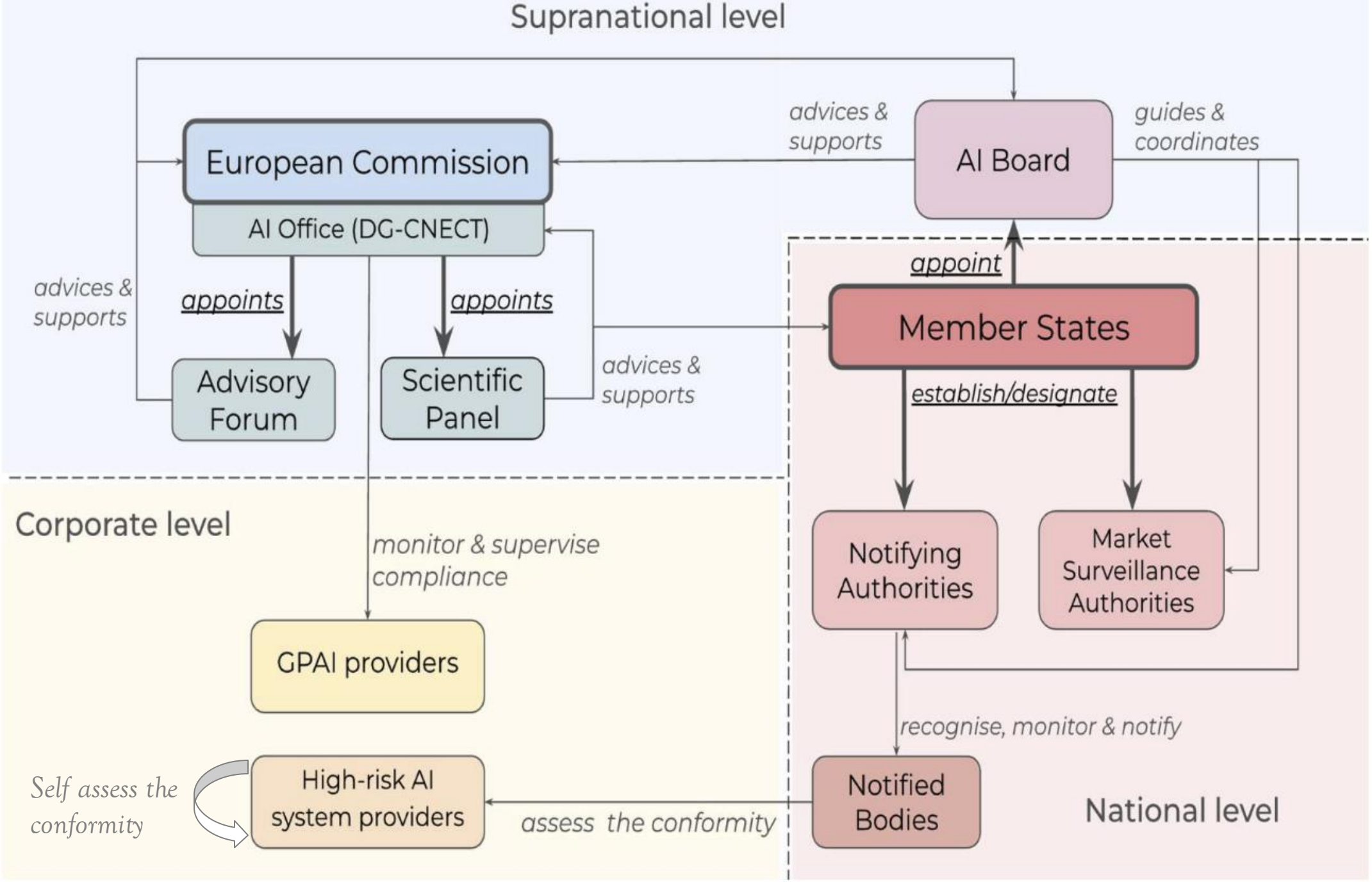
I. The AIA regulatory model

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards
II. Shift of paradigm or revolution grounded on traditions?
III. Is this surprising?
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy
V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts
VI. Conclusion



Risk: "the combination of the probability of occurrence of harm and the severity of that harm" (art. 3(2) AIA)

J. De Cooman, *The AI Act as a regulatory framework for technology: the rationales for regulation at the EU level revisited* (2024) https://hdl.handle.net/2268/319344



Adapted from C. Novelli et al., 'A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities' (2024) *European Journal of Risk Regulation* 1-25

A. The New Legislative Framework

- The AI Act is part of EU product safety
  - Based on so-called New Legislative Framework (NLF)
    - EU legislature does not define technical specifications
    - EU legislature define essential requirements a product has to meet
- For a NLF product to be marketed within the EU, it has to be CE marked
  - High-risk AI systems does not escape the rule
- CE marking follows a twofold objective
  - It allows product to traded in the internal market without restrictions
  - It informs consumers that the product is (supposed to be) safe

A. The New Legislative Framework

- CE marking is up to manufacturers
  - AI providers:
    - Ensure their systems meet essential requirements (Art. 16(a) AIA)
    - Carry out conformity assessment (Art. 43 AIA)
    - Issue technical file (Art. 11 and 18 AIA)
    - Issue EU declaration of conformity (Art. 47 AIA)
    - Affix the CE mark to the product (Art. 48 AIA)

A. The New Legislative Framework

- Conformity assessment:
  - Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products (recital 22):
    - "The manufacturer, having detailed knowledge of the design and production process, is best placed to carry out the complete conformity assessment procedure"
    - "Conformity assessment should remain the obligation of the manufacturer alone."
  - This sets the NLF apart from, *e.g.*, pharmaceutical regulations (Vos, 1999)
    - Where a public authority (EU medicine Agency) conducts an assessment before granting pre-marketing approval

A. The New Legislative Framework

- Conformity assessment in the AIA
- Providers of high-risk AI systems pursuant Annex I.A. have to follow the relevant conformity requirement under the relevant legal acts (art. 43(3) AIA)
  - *Caveat*: specific scope of application regarding high-risk AI systems pursuant Annex I.B (art. 2(2) AIA).
- Providers of high-risk AI systems pursuant Annex III have to follow the self-assessment procedure (art. 43(2) AIA)
  - Exception: providers of biometric AI systems (Annex III(1) AIA) who has applied either harmonised standards (art. 40) or common specifications (art. 41) can choose between self-assessment (Annex VI AIA) or the assessment by a notified body (Annex VIII)
    - Upshot: the assessment by a notified body is mandatory for biometric AI systems if harmonised standards or common specifications do not exist or were not applied by the provider (art. 43(1) AIA)
  - The provider may choose any notified body, except if it is put into service
    - by law enforcement, immigration, or asylum authorities --> data protection supervisory authority
    - by EU institutions, bodies, offices and agencies → European Data Protection Supervisor

B. Harmonised standards

- EU legislature sets the essential requirements
- Harmonised standards operationalise them
- In short:
  - The Commission issue a standardisation request to one of the three European Standardisation Organisations (ESOs)
    - European Committee for Standardisation (CEN)
    - European Committee for Electrotechnical Standardisation (CENELEC)
    - European Telecommunications Standards Institute (ETSI)
  - Chosen ESO drafts standards
  - Standards are published in the Official Journal of the EU

B. Harmonised standards

- In theory, harmonised standards are voluntary
- In practice, providers apply them
  - Presumption of conformity (art. 40(1) AIA)
- Upshot? The 'true' regulators of NLF products—including AI systems—are standardisation organisations (Vaele & Borgesius, 2021)

II. Shift of paradigm or revolution grounded on traditions?

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards

II. Shift of paradigm or revolution grounded on traditions?

III. Is this surprising?

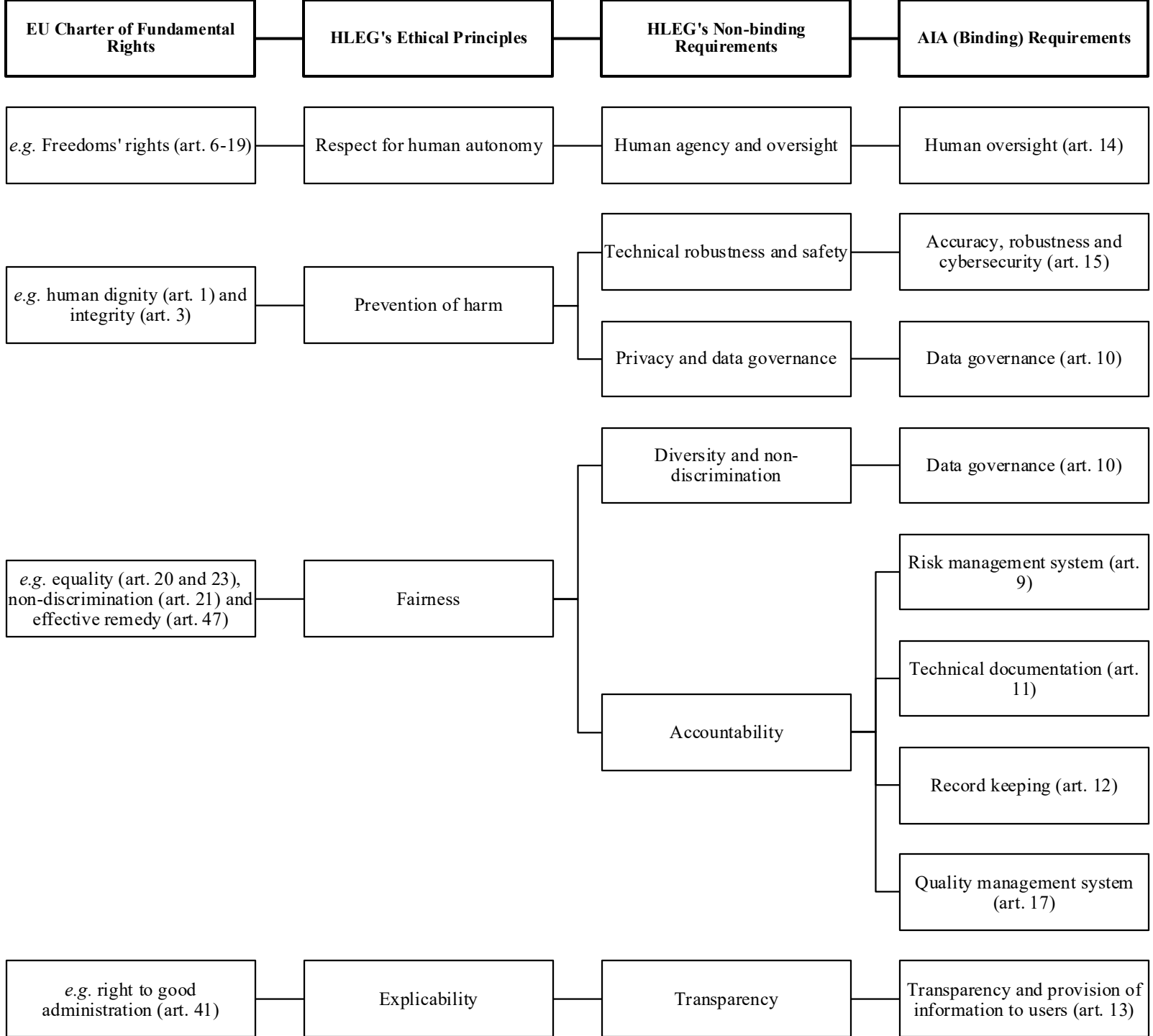
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy

V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts

VI. Conclusion

Is there something new under the sun in the NLF?

- Toy Safety - Directive 2009/48/EU
- Transportable pressure equipment - Directive 2010/35/EU
- Restriction of Hazardous Substances in Electrical and Electronic Equipment - Directive 2011/65/EU
- Construction products - Regulation (EU) No 305/2011Pyrotechnic Articles - Directive 2013/29/EU
- Recreational craft and personal watercraft - Directive 2013/53/EU
- Civil Explosives - Directive 2014/28/EU
- Simple Pressure Vessels - Directive 2014/29/EU
- Electromagnetic Compatibility - Directive 2014/30/EU
- Non-automatic Weighing Instruments - Directive 2014/31/EU
- Measuring Instruments - Directive 2014/32/EU
- Lifts - Directive 2014/33/EU
- ATEX - Directive 2014/34/EU
- Radio equipment - Directive 2014/53/EU
- Low Voltage - Directive 2014/35/EU
- Pressure equipment - Directive 2014/68/EU
- Marine Equipment - Directive 2014/90/EU
- Cableway installations - Regulation (EU) 2016/424
- Personal protective equipment - Regulation (EU) 2016/425
- Gas appliances - Regulation (EU) 2016/426
- Medical devices - Regulation (EU) 2017/745
- In vitro diagnostic medical devices - Regulation (EU) 2017/746
- EU fertilising products – Regulation (EU) 2019/1009
- Drones - Commission Delegated Regulation (EU) 2019/945 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems
- Batteries - Regulation (EU) 2023/1542
- Machinery - Regulation (EU) 2023/1230
- Ecodesign requirements for sustainable products - Regulation (EU) 2024/1781
- **Artificial Intelligence Act - Regulation (EU) 2024/1689**
- Cyber Resilience Act - Regulation (EU) 2024/2847
- Packaging and Packaging Waste - Regulation (EU) 2025/40


EU Charter of Fundamental Rights
HLEG's Ethical Principles
HLEG's Non-binding Requirements
AIA (Binding) Requirements
e.g. Freedoms' rights (art. 6-19)
Respect for human autonomy
Human agency and oversight
Human oversight (art. 14)
e.g. human dignity (art. 1) and integrity (art. 3)
Prevention of harm
Technical robustness and safety
Accuracy, robustness and cybersecurity (art. 15)
Privacy and data governance
Data governance (art. 10)
e.g. equality (art. 20 and 23), non-discrimination (art. 21) and effective remedy (art. 47)
Fairness
Diversity and non-discrimination
Data governance (art. 10)
Accountability
Risk management system (art. 9)
Technical documentation (art. 11)
Record keeping (art. 12)
Quality management system (art. 17)
e.g. right to good administration (art. 41)
Explicability
Transparency
Transparency and provision of information to users (art. 13)

Risk of fundamental rights violation

- "The Commission is empowered to adopt delegated acts in accordance with Article 97 to amend Annex III by adding or modifying use-cases of high-risk AI systems where (…) the ***AI systems pose a risk of harm to health and safety, or an adverse impact on fundamental rights***" (art. 7(1(b) AIA)
- "Deployers that are bodies governed by public law, or are private entities providing public services [except of they are deployers of 'critical infrastructure' AI system], and deployers of [AI systems intended to be used to evaluate the creditworthiness of natural persons] and [AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance], shall perform an ***assessment of the impact on fundamental rights*** that the use of such system may produce" (art. 27(1) AIA)

There is definitely something new under the sun

- This is the first time that the EU legislature tries to protect fundamental rights via harmonised standards
- Harmonised standards "incorporate core EU democratic values and interests, as well as green and social principles" (Commission, 2022)
- This raises three questions:
  - Is this surprising?
  - Is this problematic?
  - Is there a solution?

III. Is this surprising?

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards
II. Shift of paradigm or revolution grounded on traditions?
III. Is this surprising?
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy
V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts
VI. Conclusion

III. Is this surprising?

Merkel (2019)

"It will be the job of the next Commission to deliver something so that we have regulation similar to the ***General Data Protection Regulation*** that makes it clear that artificial intelligence serves humanity"

Von der Leyen (2020)

"AI that potentially interferes with people's rights have to be tested and certified before they reach our Single Market. This is a very simple question, because we do it just the same way with for example ***cars or chemicals, or cosmetics, or toys***"

III. Is this surprising?

- Regulation 1025/2012 organises harmonised standards:
  - "***Harmonised standard*** means a ***standard*** adopted by a ESOs" (art. 2(1)(c))
  - "***standards*** means a *technical specification*" (art. 2(1) in limine)
  - "*technical specification* means a document that prescribes technical requirements to be fulfilled by a product, process, service or system and which lays down (…) the characteristics required of a product [or service] including levels of quality, performance, interoperability, environmental protection, health or safety" (art. 2(4)(a)-(c))
- Fundamental rights are not mentioned in Regulation 1025/2012

III. Is this surprising?

- The rules applicable to products that are part of the NLF are explained in the Commission Blue Guide (last version: 2022)
- "The New Legislative Framework now constitutes a complete system bringing together all the different elements that need to be dealt with in product safety legislation in a coherent, comprehensive legislative instrument that can be used across the board in all industrial sectors, and even beyond (***environmental and health policies also have recourse to a number of these elements***), whenever EU legislation is required."
  - Fundamental rights are not mentioned in the Blue Guide

III. Is this surprising?

- Regulation 2019/1020 on market surveillance and compliance of products
  - Market surveillance authorities "shall suspend the release of a product for free circulation if (…) it is established that (…) it presents a serious risk to health, safety, the environment or ***any other public interest referred to in Article 1.***"
  - "The objective of this Regulation is to improve the functioning of the internal market by strengthening the market surveillance of products covered by the *Union harmonisation legislation referred to in Article 2*, with a view to ensuring that only compliant products that fulfil requirements providing a high level of protection of ***public interests***, such as ***health and safety in general***, health and safety in the workplace, ***the protection of consumers, the protection of the environment and public security and any other public interests protected by*** *that legislation*, are made available on the Union market." (Article 1)
- Again, no reference to fundamental rights

III. Is this surprising?

- "***Standardisation*** should play a ***key role*** to provide technical solutions to providers to ensure compliance with this regulation, in line with the state of the art, ***to promote innovation as well as competitiveness and growth in the single market*** (Recital 121 AIA)"
- However:
  - "The common specification should be an exceptional fall back solution to facilitate the provider's obligation to comply with the requirements of this Regulation, when the standardisation request has not been accepted by any of the [ESOs]European standardisation organisations, or ***when the relevant harmonised standards insufficiently address fundamental rights concerns***" (Recital 121 + Art. 41(1)(a)(iii) AIA)
  - "The participants in the standardisation process shall ***seek to promote investment and innovation in AI***, including through increasing legal certainty, ***as well as the competitiveness and growth of the Union market***, to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are ***consistent with Union values [and] fundamental rights***" (Art. 40(3) AIA)

IV. Is this problematic?

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards
II. Shift of paradigm or revolution grounded on traditions?
III. Is this surprising?
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy
V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts
VI. Conclusion

A. Smart regulation

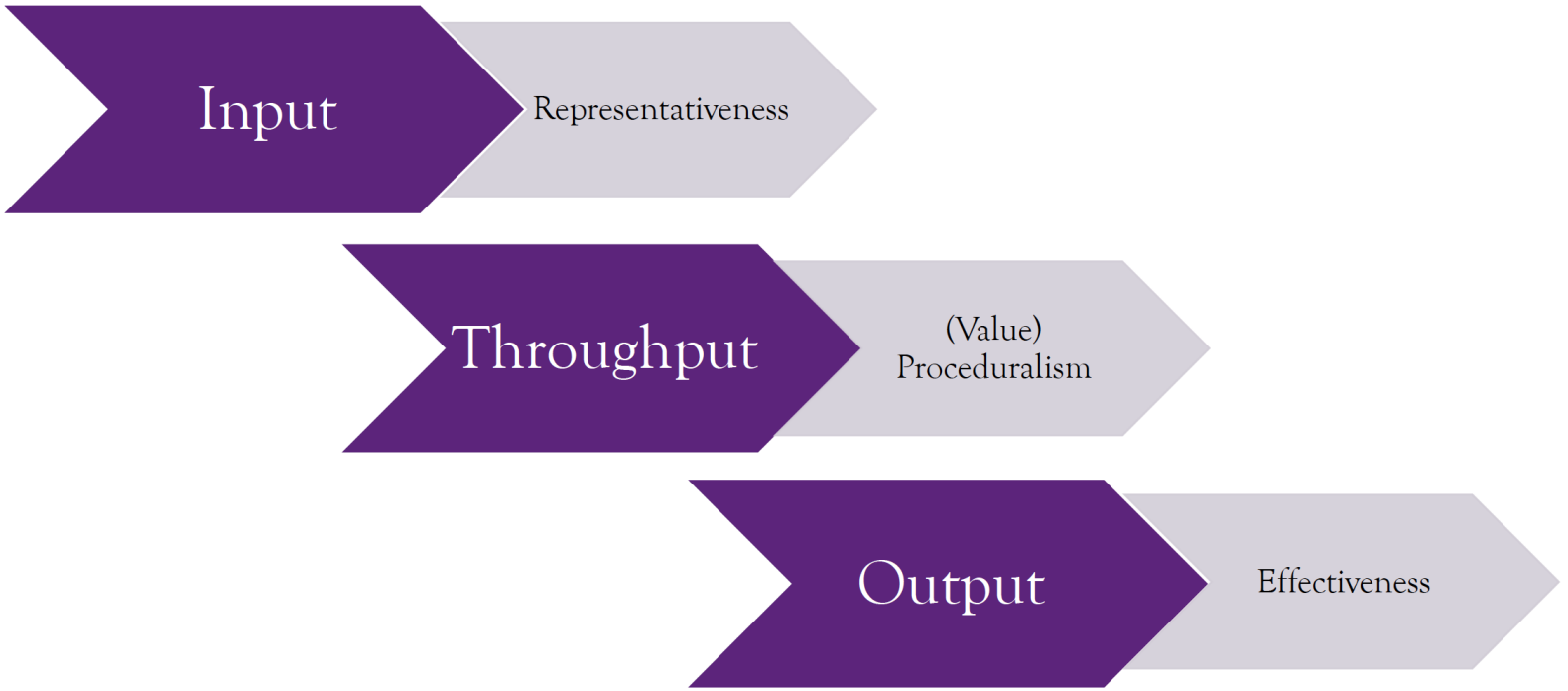
- Is this problematic? Not necessarily, according to regulatory theory
  - Regulation *tout court* is the product of state actors
  - *Self*-regulation is the product of non-State actors
  - *Co*-regulation is the product of their cooperation (co-operation; co-production; co-regulation), *i.e.*:
    - Self-regulation where there is a "regulatory 'gorilla in the closet' that secure[s] its ultimate success" (Gunningham & Sinclair, 2017)
    - "A model that combines both legislation and self-regulatory instruments in support of the law" (Kamara, 2017)
    - "A regulatory framework that involves both private parties and governmental actors in the setting, implementation, or enforcement of regulatory standards" (Van Cleynenbreugel, 2021)
    - Meta-regulation: "activities occurring in a wider regulatory space, under the auspices of a variety of institutions, including the state, the private sector and the public interest group" (Grabosky, 2017)
    - Multisource regulation (Drahos & Kryger, 2017)

A. Smart regulation

- Is this problematic? Not necessarily, according to regulatory theory:
  - Meta-regulation and co-regulation are forms of ***smart regulation*** (Gunningham and Sinclair, 2017):
    - "A form of regulatory pluralism that embraces ***flexible, imaginative and innovative forms of social control***"
    - "***The use of*** multiple rather than single policy instruments, and ***a broader range of regulatory actors***, will produce ***better regulation***"
  - The AIA is co-regulation / meta-regulation / smart regulation
    - EU legislature sets up essential requirements
    - ESOs operationalise them

B. The legitimacy conundrum

- Is this problematic? Yes, according to:
  - The European Association for the Coordination of Consumer Representation in Standardisation AISBL (in short, ANEC)
  - The 'legitimacy' scholarship



1. Input legitimacy

- Input legitimacy: "the EU's responsiveness to citizen concerns as a result of participation by the people" (Schmidt, 2013).
- Input legitimacy of standardisation: all citizens' interests are sufficiently taken into account and "protected in privately set standards and therefore also on rules and procedures that allow for equal representation and balancing of all interests at stake" (Senden, 2020).

1. Input legitimacy

- Regulation 1025/2012
  - Works of standardisation bodies have to be transparent (art. 3-4)
  - All stakeholders must both be allowed to participate in the standardisation process (art. 5-7)
  - All stakeholders must have access to standards (art. 5-7)

- Pious wishes?
  - SMEs usually lack the financial resources to effectively participate in the standardisation process
  - Article 5 "does not entail any voting rights for these stakeholders unless such voting rights are prescribed by the internal rules of procedure of the European standardisation organisations" (recital 23)
  - Copyright issue (?)
    - *Public.Resource.org and Right to Know v. Commission and others*

1. Input legitimacy

- Ideally, *normative* questions should be left to EU legislature and *technical* question should be left to ESOs (Laux 2024)
- Actually, "standards have politics" (Solow-Niederman 2024 // Wiener 1980)
  - Kranzberg's first law: "technology is neither good nor bad; nor is it neutral" (1986)
  - So do harmonised standards
    - E.g., fairness (see next slide)
    - Choosing one particular definition of fairness signals a preference for a specific logic and set of priorities (Gornet and Maxwell, 2024)

2. Output legitimacy

- Output legitimacy: effectiveness and efficiency of the rule (Scharpf, 1999).
- Is standardisation effective?
  - "The purpose of this Regulation is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems" (recital 1 AIA)
- Legal basis: Art. 114 TFEU (and 16 TFEU)
- Method of harmonisation:
  - Full
  - Maximum
- *Common* market needs *common* rule, and harmonised standards fit this objective
  - But output legitimacy cannot be presumed a priori (De Bellis, 2020)

2. Output legitimacy

- Is standardisation effective?
  - Yes, when it comes to interoperability, cross-border services, and switch-costs reduction (e.g., telecommunication standards)
  - Not always, when it comes to health and safety standards
    - EU law requires medical grade silicon for breast implant
    - Breast implant (industrial silicon) CE certified by German notified body
    - 40,000 affected women in France; 400,000 worldwide
  - Probably not, when it comes to fundamental rights
    - Question: what makes providers the "best place to carry out complete conformity assessment procedure" when it comes to fundamental rights (*cf.* 768/2008/EC)?
    - Doubtful CEN (or ISO) and CENELEC (or IEC) have relevant expertise in the area of fundamental rights
  - Caveat: certification mainly concerns the verification that requested documentation is in place
    - "without scrutiny by an independent regulator, a safety case may not be worth the paper it is written on" (Hopkins, 2012)

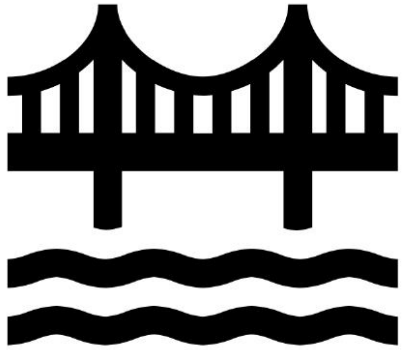
2. Output legitimacy

- Hypo:
  - An AI provider self-certifies its compliance with harmonised standards…
  - …but actually does not comply
  - Will be discover if market surveillance authority evaluate that AI system (post-market monitoring)
    - But this can only be done if the MSA has "sufficient reasons to consider an AI system present a risk" (Art. 77(3) and 79(2) AIA)
- Too soon to conclude AI harmonised standards are output (il)legitimate
  - But fair enough to assume they will not be the most effective way to protect fundamental rights

3. Throughput legitimacy

Input legitimacy

Output legitimacy



Representativeness

Effectiveness

Throughput legitimacy

Legitimate

Illegitimate

3. Throughput legitimacy

- Throughput legitimacy: regulation must be "judged in terms of the efficacy, accountability and transparency of the EU's governance processes along with their inclusiveness and openness to consultation with the people" (Schmidt, 2013)
- The ('deconstructed') *Meroni* doctrine (ECLI:EU:C:1958:7): The delegated powers must exclusively consist of well-defined executive powers (rather than delegating powers involving a wide margin of discretion to private actors) that can undergo an objective review by the delegating authority
- *ESMA* (Case C-270/12): delegation of discretionary power is legal if there is judicial review
- Judicial review
  - the institutional bedrock of most rule of law models (Möllers, 2016)
  - the EU is "a community based on the rule of law." (ECJ, *Les Verts*, § 23)
  - Harmonised standards are part of EU law (ECJ, *James Elliott Construction*, § 40)
  - *Ergo*: harmonised standards *should* be subject to judicial scrutiny

3. Throughput legitimacy

- But:
  - Harmonised standards are part of EU law but are not "acts of EU institutions, bodies, offices or agencies of the Union"
    - CJEU has jurisdiction to interpret harmonised standards in preliminary rulings (art. 267 TFEU)
    - CJEU has no jurisdiction in annulment action oriented towards harmonised standards (Art. 263 TFEU)
      - Even if it was the case, the CJEU would be limited to assessing whether the Commission made a mistake when publishing the reference to the harmonised standards in the Official Journal (Ebers, 2021)
- Problem:
  - No ex ante control (*Meroni* doctrine) and no ex-post judicial scrutiny (*ESMA*)
    - Shaky constitutional grounds

V. Case studies

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards
II. Shift of paradigm or revolution grounded on traditions?
III. Is this surprising?
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy

V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts
VI. Conclusion

A. Fairness in EU (digital) law

- Abuse of a dominant position (Art. 102(a) TFEU): is incompatible with the internal market and prohibited to "directly or indirectly imposing ***unfair*** purchase or selling prices or other ***unfair*** trading conditions."
- However, competition authorities are reluctant to impose sanctions solely on the basis of the unfairness of trading conditions (Colangelo, 2023).
  - "Endemic uncertainty"

A. Fairness in EU (digital) law

- Yet, the European legislator increasingly refers to fairness as a foundational concept in digital regulation.
  - "The principles of ***fair*** and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes. The controller should provide the data subject with any further information necessary to ensure ***fair*** and transparent processing" (Recital 60 GDPR)
- Regulation (EU) 2022/1925 on contestable and ***fair*** markets in the digital sector (DMA):
  - "The market processes are often incapable of ensuring ***fair*** economic outcomes with regard to core platform services" (Recital 5 DMA)

A. Fairness in EU (digital) law

- Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (DA)
  - A contractual term is ***unfair*** if it is of such a nature that its use grossly deviates from good commercial practice in data access and use, contrary to ***good faith and fair dealing*** (Art. 13(3) DA).
    - Directive 93/13/CEE on ***unfair terms*** in consumer contracts
    - Directive 2005/29/CE concerning ***unfair business-to-consumer commercial practices*** in the internal market
- Regulation (UE) 2019/1150 on promoting ***fairness*** and transparency for business users of online intermediation services
  - "Providers of those services often have superior bargaining power, which enables them to, in effect, behave unilaterally in a way that can be ***unfair*** and that can be harmful to the legitimate interests of their businesses users and, indirectly, also of consumers in the Union. For instance, they might unilaterally impose on business users practices which grossly deviate from good commercial conduct, or are contrary to ***good faith and fair dealing***" (recital 2)

A. Fairness in EU (digital) law

- Regulation (UE) 2022/2065 on a Single Market For Digital Services (DSA)
  - "it is appropriate to set certain rules on the content, application and enforcement of the terms and conditions of those providers in the interests of transparency, the protection of recipients of the service and the avoidance of ***unfair*** or arbitrary outcomes." (Recital 45)
  - "the out-of-court dispute settlement that it offers takes place in accordance with clear and ***fair rules of procedure*** that are easily and publicly accessible, and that comply with applicable law, including this Article" (art. 21(3)(f) DSA)

- No cited legislation defined fairness
  - RGPD: fairness = protection of data subject (Malgieri 2020)
  - DMA: fairness = promoting competition (Bostoen 2024)
  - DA & DSA: fairness = good faith and fair dealing (Husovec, 2023)
  - DSA: fairness = clear rule of procedure (e.g., right to a fair trial)

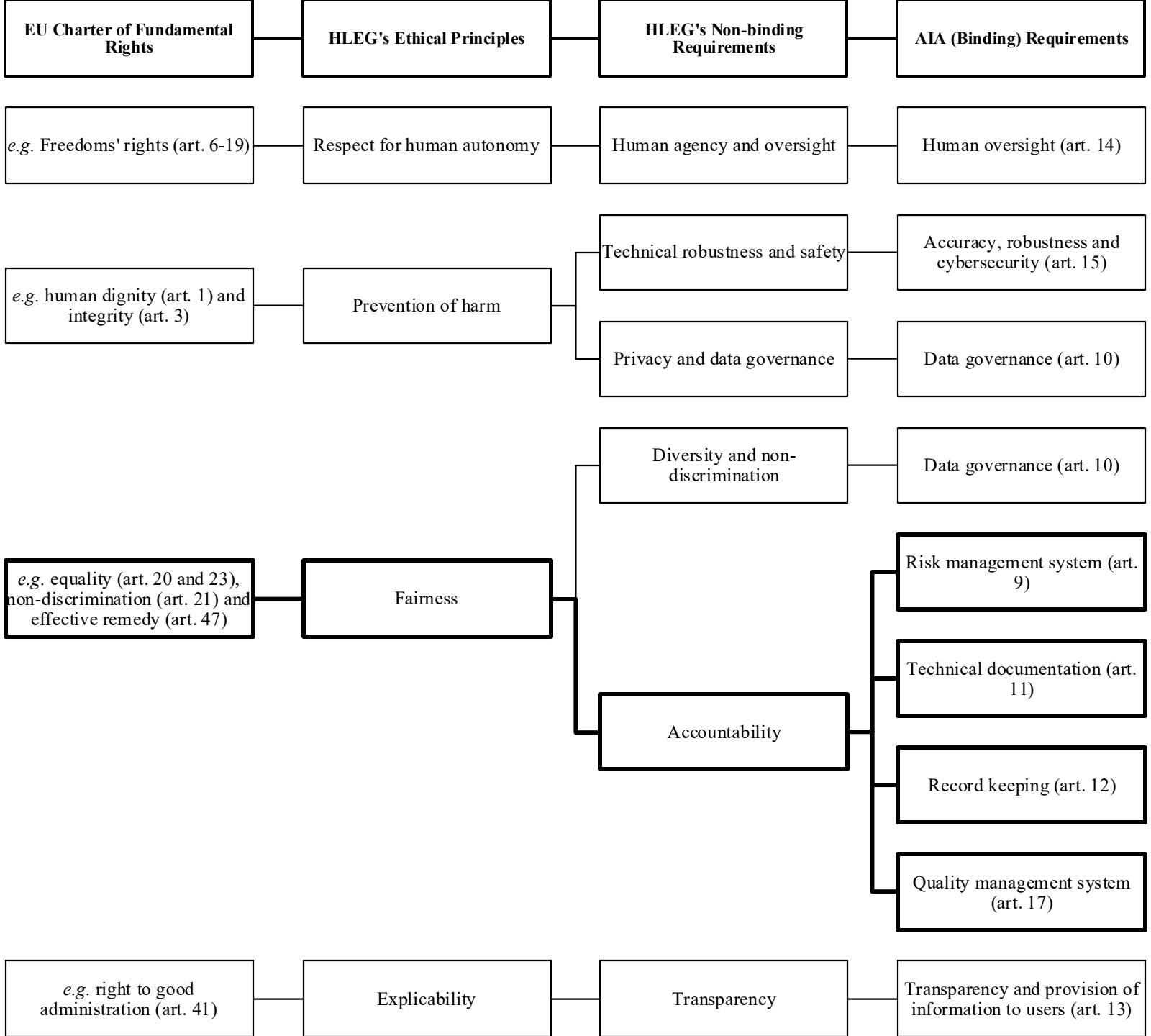
A. Fairness in EU (digital) law

- The EU AI Act (2014/1689) seems to deviate from this 'fairness crazyness'
- Surprising, given the AI HLEG's Ethics Guidelines (2019)
  - HLEG: "The development, deployment and use of AI systems must be ***fair***. While we acknowledge that there are many different interpretations of fairness, we believe that fairness has ***both a substantive and a procedural dimension***"

B. The case of 'standardised' procedural fairness

- "Le volet ***procédural*** de l'équité suppose la **capacité de contester les décisions** prises par des systèmes d'IA et par les êtres humains qui les utilisent, ainsi que celle d'introduire un recours efficace à l'encontre de ces decisions. Pour ce faire, l'entité ***responsable*** de la décision doit pouvoir être identifiée."
  - Voies de recours (*redress mechanisms*) : pas de détail
  - "Responsabilité" (*accountability*) :
    - "**Auditabilité**" : "possibilité d'"valuer [par des auditeurs internes et externes] les algotithmes, les données et les processus de conception" (*caveat*: ⊄ informations sur le modèle économique et la propriété intellectuelle)
    - "**Documentation des incidences negatives**" : "Il convient de garantir la capacité aussi bien de ***documenter les actions*** ou décisions contribuant à un certain résultat du système que de ***répondre aux conséquences*** d'un tel résultat (…). Le recours aux ***analyses d'impact*** (…) peut contribuer à réduire le plus possible les effets négatifs"

Ref.: Grozdanovski and De Cooman (2023)



1. Redevabilité (*accountability*)

- Pour le HLEG, accountability = responsabilité
- Problème de traduction : responsabilité = liability, responsibility, accountability
- Accountability → account–ability → rendre des comptes (Capron, 2016)
  - "La société a le droit de demander compte à tout agent public de son administration" (Art. 15 Déclaration des Droits de l'Homme et du Citoyen de 1789)
  - To be accountable = être comptable de son action
  - Accountability = "redevabilité", "responsabilité" *au sens large*
    - HLEG: accountability is 'an ethical standard' which might '*fall short of legal consequence*.'
- Responsabilité civile (AILD) ou du fait des produits défectueux (R-PLD) : *renvoi*

1. Redevabilité (*accountability*)

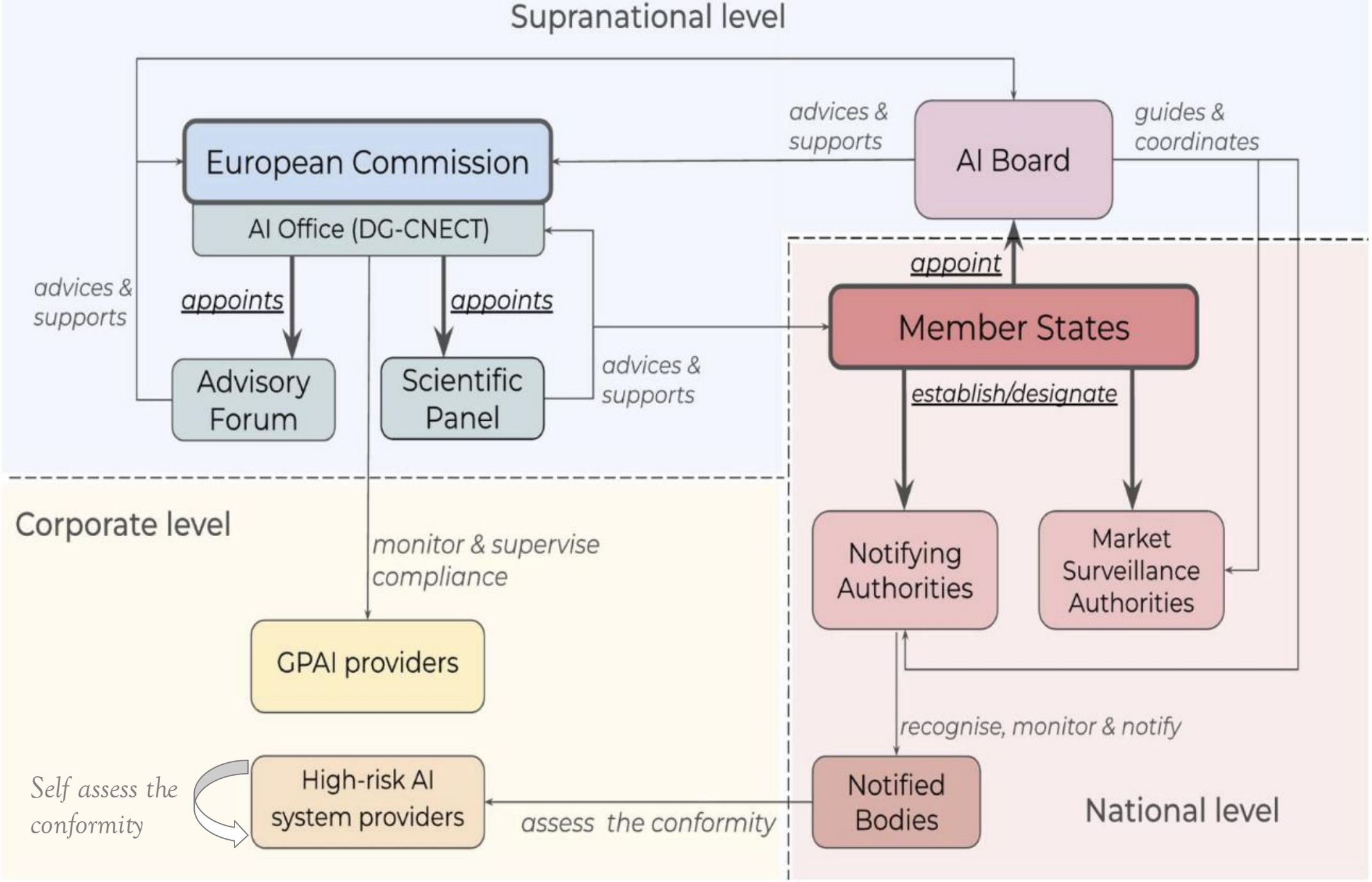
- La redevabilité a un double sens (Bovens 2010) :
  - Définition matérielle : concept normatif composé de standards utilisés pour évaluer le comportement des agents.
    - Accountability = responsiveness = bonne gouvernance → "volonté d'agir de façon transparence, equitable et juste" (Bovens 2010)
  - Définition procédurale : la redevabilité signifie (littéralement) qu'une personne doit rendre des comptes à une autre
- HLEG : "accountability" = "auditabilité"
  - Choix du volet procedural
    - Redevabilité = *accountability as answerability*

1. Redevabilité (*accountability*)

- La redevabilité est un concept relationnel (Bovens 2007; Lindberg, 2013) :
  - délégation à un agent (*A*) d'une tâche (*T*) par une source (*P* pour principal), au nom de laquelle l'agent agit.
  - L'agent doit ensuite justifier sa conduite auprès du forum (*F*) qui supervise l'agent et évalue sa conduite
- La redevabilité est un système de contrôle au sens cybernétique du terme (Olsen, 2014) :
  - Définition des normes (standard-settings) : *A* respecte les règles éthiques (HLEG) ou juridiques (AIA)
  - Collecte d'information (information-gathering) : supervision de *A* par *F* (dimension dialogique)
  - Modification du comportement (Behaviour-modification) : approbation ou sanction du comportement de *A* par *F*
- La combinaison de ces éléments permet la construction d'une *matrice* de redevabilité (Mulgan, 2000; Novelli *et al.*, 2024a)

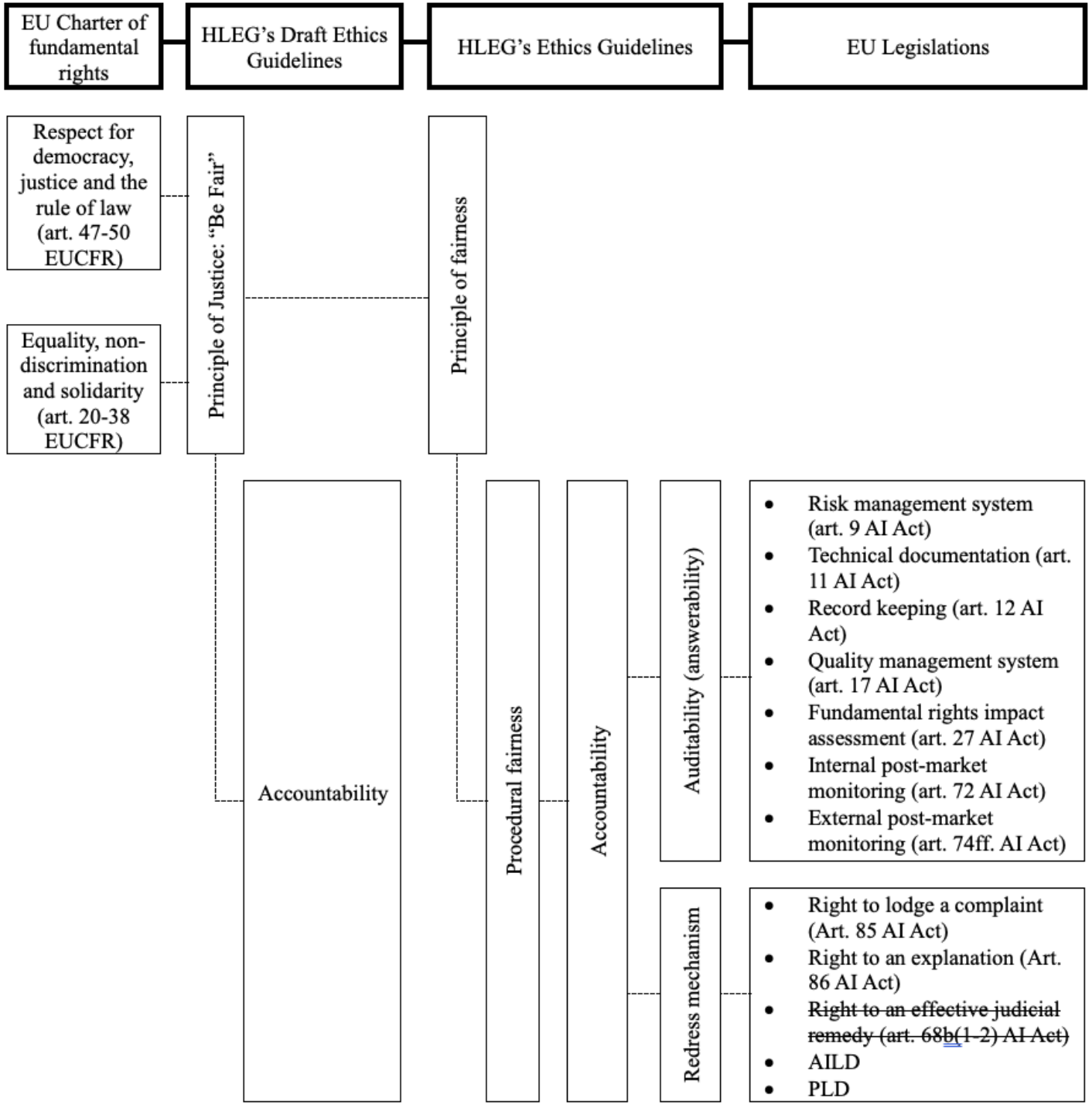
1. Redevabilité (*accountability*)

Requirements for accountability	Features of accountability
Actors	Agent (who?)
	Forum (to whom?)
	Principal (by whom?)
Standard-setting	Context (what for?)
	Range (about what?)
	Standard (according to what?)
Information-gathering	Process (how?)
Behaviour-modification	Consequence (what follows?)



Adapted from C. Novelli et al., 'A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities' (2024) *European Journal of Risk Regulation* 1-25

Ref.: De Cooman (2025)



1. Redevabilité (*accountability*)

- Actors :
  - Agent? AI operators
  - Forum? Market surveillance authority
  - Principal? EU Legislature
- Standard-settings :
  - Context? High-risk AI system
  - Range? "The risk management system shall be understood as a continuous iterative process ***planned and run throughout the entire lifecycle*** of a high-risk AI system, requiring regular systematic review and updating." (art 9(2) AIA)
  - Standard? EU AI Act

Requirements for accountability	Features of accountability	Holding AI operator accountable
Actors	Agent (who?)	AI operators (art. 3(8) AIA)
	Forum (to whom?)	Market surveillance authority
	Principal (by whom?)	EU legislature
Standard-setting	Context (what for?)	High-risk AI systems (art. 8 AIA)
	Range (about what?)	Compliance throughout the AI lifecycle: • Design; • Development; and • Deployment.
	Standard (according to what?)	EU AIA
Information-gathering	Process (how?)	Self-assessment prior marketing: • Risk management system (art. 9 AIA) • Technical documentation (art. 11 AIA) • Logs keeping (art. 12 AIA) • Quality management system (art. 17 AIA) • Fundamental rights impact assessment (art. 28 AIA) Post-market monitoring (art. 3(25) and 72 AIA): • Internal (art. 73 AIA) • External (art. 74ff AIA)
Behaviour-modification	Consequence (what follows?)	• Achieving system conformity, withdrawal, recall, disablement (art. 20 and 79(2) and (5) AIA); and • Fines (art. 99(4) AI Act)

Is there a solution?

- What standards **can** do:
  - **Procedure-based standards** (*i.e.*, standards that guide how a company should structure its internal systems)
    - *E.g.*, CEN JT021019 "competence requirements for AI ethicists professionals" (under drafting):
    - "This document provides a ***systematized framework for the competencies of AI ethicists***, categorizing them into knowledge, skills, and attitudes related to the specific activities and tasks of the role. It ***identifies requirements and recommendations necessary for individuals to effectively perform as AI ethicists***. These competencies encompass a strong understanding of European values and fundamental rights, further enhancing the knowledge, skills, and attitudes required for this profession. ***The document aims to foster a shared understanding of the essential concepts and principles inherent to the AI ethicist role***. It illustrates a clear, uniform approach to the integral components of this profession. Moreover, the document outlines how the role of AI ethicists can be seamlessly integrated into a wide variety of organizations. These include, but are not limited to, commercial enterprises, government agencies, and non-profit organizations."

2. Standardisation "do's and don'ts"

- What standards **cannot** do:
  - EC's request concerns Chapter III, Section 2, AIA
    - Includes article 9 (risk management system)
      - Standards can define how to conduct a risk assessment system (i.e., details what elements it should contain)
      - Standards cannot determine what "residual risk (…) is judged to be acceptable" (art. 9(5) AIA)
        - A standard cannot determine what type of risk is acceptable or not
      - Standards cannot organise trade-off between, *e.g.*, fairness and performance (but can provide different different definitions and ways of measuring fairness)
        - "The technical documentation referred to in Article 11(1) shall contain (…) a detailed description of the elements of the AI system and of the process for its development, including (…) the decisions about possible trade-off made regarding the technical solutions adopted to comply with the requirements set out in Chapter III, Section 2" (Annex IV(2)(b)).

C. The case of 'standardised' substantive fairness

- "Le volet ***matériel*** suppose l'engagement de veiller à une ***répartition égale et juste des bénéfices et des coûts***, et de veiller à ce que les individus et les groupes ne fassent pas l'objet de biais injustes, de ***discrimination*** et de stigmatisation"
  - L'équité (*equity*) : justice distributive :
    - l'IA peut s'avérer être un excellent outil pour combattre les inégalités en matière d'éducation et créer des programmes de formation personnalisés et adaptables qui pourraient aider chacun à obtenir de nouvelles qualifications, aptitudes et compétences en fonction de ses propres capacités d'apprentissage. Cela pourrait augmenter la vitesse d'apprentissage et améliorer la qualité de l'enseignement – de l'école primaire à l'université."
    - "Il convient également d'encourager l'égalité des chances en ce qui concerne l'accès à l'éducation, aux biens, aux services et à la technologie"

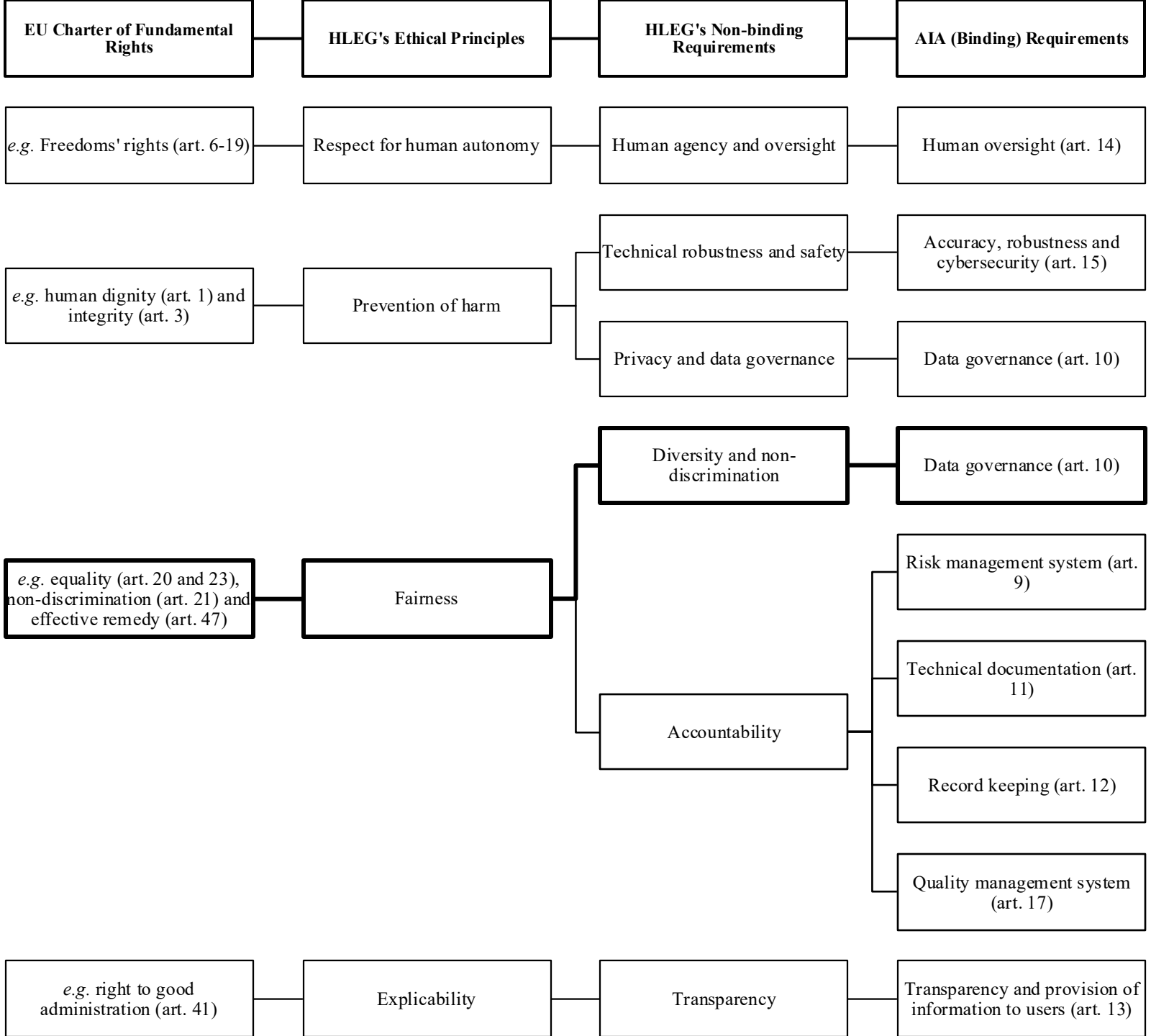
C. The case of 'standardised' substantive fairness

- "Le volet matériel suppose l'engagement de veiller à une ***répartition égale et juste des bénéfices et des coûts***, et de veiller à ce que les individus et les groupes ne fassent pas l'objet de biais injustes, de ***discrimination*** et de stigmatisation"
  - L'égalité : (*equality*) diversité et non-discrimination
    - "**Absence de biais injuste**" : "les biais (…) discriminatoires devraient être supprimés lors de la phase de collecte. La manière dont les systèmes d'IA sont mis au point (par exemple la programmation des algorithmes) peut également être entachée de biais (…) Le recrutement de personnes issues de contextes, de cultures et de disciplines différents peut garantir la diversité des opinions et devrait être encouragé.
    - "**Accessibilité et conception universelle**" : "les systèmes devraient être centrés sur l'utilisateur et conçus de manière à permettre à toute personne d'utiliser des produits ou services d'IA, quels que soient son âge, son sexe, ses capacités ou ses caractéristiques"
    - "**Participation des parties prenantes**" : "Pour mettre au point des systèmes d'IA dignes de confiance, il est souhaitable de consulter les parties prenantes sur lesquelles le système est susceptible d'avoir des effets"

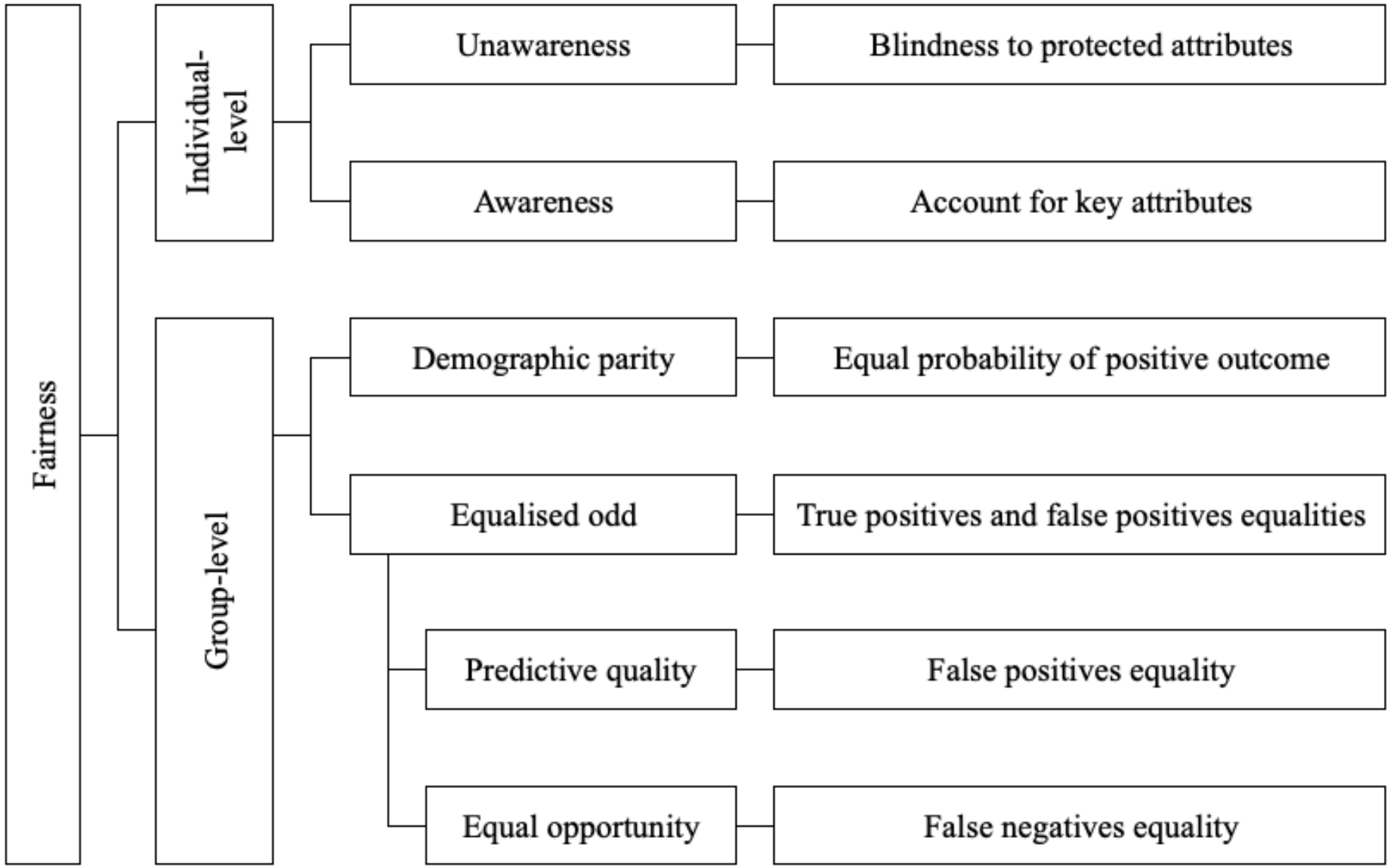
C. The case of 'standardised' substantive fairness

- "AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law" (Recital 27 AI Act).
  - Recital 27 AI Act was once article 4(a)(1)(e) in the version amended by the European Parliament.
  - The shift, from a (mandatory) provision to a (non-mandatory) recital reveals the EU legislature's willingness to 'dilute' the importance of fairness in the final version of the text.
  - This suggestion is in line with the verbatim of Recital 27, which stresses 'the application of those principles [including fairness] should be translated, when possible, in the design and use of AI models' and 'should in any case serve as a basis for the drafting of codes of conduct under this Regulation'.
- This has been lost in translation
  - Fairness as equity (distributive justice): An even distribution of costs and benefits is no longer on the agenda
  - Fairness as diversity is a non-mandatory requirement.
    - AI operators should 'be encouraged to apply on a voluntary basis additional requirements related, for example, to (…) inclusive and diverse design and development of AI systems, including attention to vulnerable persons and accessibility to persons with disability' (Recital 165 AI Act).
    - Codes of conduct should facilitate the 'inclusive and diverse design of AI systems, including through the establishment of inclusive and diverse development teams and the promotion of stakeholders' participation in that process' (art. 95(2)(d) AI Act).
- Fairness in the AI Act translates as non-discrimination and heavily focuses on discrimination because of data.

Ref.: Grozdanovski and De Cooman (2023)



1. Fairness as non-discrimination



1. Fairness as non-discrimination

- In the *Seymour-Smith* case (C-167/97), the CJEU explained that, when statistics are available, the 'best approach' to compare them is to consider the respective proportions of advantaged group members who receive positive outcomes and those with negative ones compare those proportions with those of disadvantaged groups.
- Demographic disparity: a system would in theory be fair if the difference in the probability of favourable results between the disadvantaged and the advantaged groups is equal to 0.
- In practice, a small (positive or negative) difference can still be considered fair (the closer to 0, the greater the fairness).
- No threshold set by the CJEU ('battle of the figures')
  - A difference of 8.5% is considered insignificant (*Seymour-Smith*)
  - A difference of 50% is significantly high (*Villar Láiz*)

2. Standardisation "do's and don'ts"

- What standards can do:
  - **Product-based standards**: define and identify relevant and workable tools or measures known in the literature
    - E.g., ISO/IEC TR 24027:2021 "provides mitigation techniques that can be applied throughout the AI system life cycle in order to treat unwanted bias"

ISO/IEC TR 24027:2021(en) Information technology

Table of contents

2. Standardisation "do's and don'ts"

- **Quality** standards (aka **performance** standards; Blind, 2004):
  - Usual in product safety (*e.g.*, what materials are to be used)
    - ISO 3506-5:2022(en)—Fasteners: "The properties of stainless steel and nickel alloy fasteners for high temperature applications result from the chemical composition of the material, from the heat treatment process and from the manufacturing process of the fasteners. Static or dynamic properties at room temperature like tensile strength, hardness or fatigue resistance are not sufficient enough to design fasteners for high temperature applications properly.
    - In fact, at high temperatures e.g. above 300 °C, additional phenomena occur, for instance:
      - decrease in tensile properties and hardness,
      - hot oxidation and scaling,
      - stress relaxation,
      - creep.
    - All these phenomena significantly affect the durability and service life of fasteners. Therefore:
      - a proper choice of material grade is essential to avoid heavy hot oxidation,
      - qualification of fasteners through dedicated tests should be performed."

2. Standardisation “do’s and don’ts”

- **Quality** standards (aka **performance** standards; Blind, 2004):
  - Usual in product safety
  - But
    - “nearly impossible to establish for AI systems due to their probabilistic nature, which makes their reaction to certain tests ***highly dependent on the situation***, the data on which the system has been trained, etc.” (Gornet and Maxwell, 2024)
    - Such standards would require setting a threshold for *e.g.*, acceptable level of fairness and, hence, acceptable level of discrimination
      - Above (or below) that threshold, the AI system would be ‘fair enough’
    - Harmonised standards should neither make value-laden judgement nor answer normative questions (Laux et al. 2014)

VI. Conclusion

I. The AIA regulatory model
   A. The New Legislative Framework
   B. Harmonised standards
II. Shift of paradigm or revolution grounded on traditions?
III. Is this surprising?
IV. Is this problematic?
   A. Smart regulation
   B. The legitimacy conundrum
      1. Input legitimacy
      2. Output legitimacy
      3. Throughput legitimacy
V. Case studies
   A. Fairness in EU (digital) law
   B. The case of 'standardised' procedural fairness
      1. Accountability
      2. Standardisation 'do's and don'ts'
   C. The case of 'standardised' substantive fairness
      1. Fairness as non-discrimination
      2. Standardisation 'do's and don'ts
VI. Conclusion

Conclusion

- The EU AI Regulation is part of the NLF
  - It is surprising
  - It is not necessarily problematic
    - Harmonised threshold-based standards cannot be set for the protection of fundamental rights
      - It is not up to standardisation organisations to define what an acceptable level of protection to fundamental rights is (i.e., to define which system is 'fair enough')
    - ESOs should be invited to develop harmonised standards which disseminate good practices (how to design risk management pursuant art. 9 AIA), or means of disclosures (what the technical disclosure pursuant art. 11 AIA should contain), but not define the acceptable residual risk or the appropriate trade-off to be made between, e.g., fairness and performance
  - Under these conditions, the AIA—as part of the NLF—will deserve to be labelled a ***smart*** regulation

LIÈGE université
Cité