A RISK-BASED APPROACH TO AI REGULATION
A risk-based approach to regulation

- **Unacceptable risk**: e.g. social scoring
- **High risk**: e.g. recruitment, medical devices
- **Minimal or no risk**:
  - Permitted with no restrictions
  - Permitted subject to information/transparency Obligations
  - Permitted subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive*
THE BALANCING OF INTERESTS
AI is good …
• For citizens
• For business
• For the public interest

… but creates some risks
• For the safety of consumers and users
• For fundamental rights
EXPLANATORY MEMORANDUM

AI (...) can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities (...) However, the same elements and techniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society (...)

this proposal presents a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market".
1) NO RULES FOR MOST AI SYSTEMS
Most AI systems will not be high-risk (Titles IV, IX)

New transparency obligations for certain AI systems (Art. 52)

- Notify humans that they are interacting with an AI system unless this is evident
- Notify humans that emotional recognition or biometric categorisation systems are applied to them
- Apply label to deep fakes (unless necessary for the exercise of a fundamental right or freedom or for reasons of public interests)

Possible voluntary codes of conduct for AI with specific transparency requirements (Art. 69)

- No mandatory obligations
- Commission and Board to encourage drawing up of codes of conduct intended to foster the voluntary application of requirements to low-risk AI systems
2) OVERWHELMING BALANCE
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.
• SUBLIMINAL TECHNIQUES ONLY IF PHYSICAL/PSYCHOLOGICAL HARM

• SOCIAL SCORING ONLY IF BY PUBLIC AUTHORITIES

• EXPLOITATION OF VULNERABILITIES ONLY FOR SPECIFIC CATEGORIES

• BIOMETRIC IDENTIFICATION ONLY FOR REAL TIME, PUBLIC AUTHORITIES, LAW ENFORCEMENT, AND WITH EXCEPTION

• NO REFERENCE TO OTHER USES THAT CONTRADICT EU VALUES, DISCRIMINATION, STIGMATIZATION, ETC.
3) DEFINITIONAL BALANCE
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
Art. 7 - Amendments to Annex III

A. the intended **purpose** of the AI system;

B. the extent to which an AI system has been **used** or is likely to be used;

C. the extent to which AI system has **already caused harm** to the **health and safety** or adverse impact on the **fundamental rights** OR has given rise to significant **concerns** in relation to the materialisation of such harm or adverse impact;

D. the **potential extent of such harm or such adverse impact** (intensity and its ability to affect a plurality of persons);

E. the extent to which potentially harmed or adversely impacted persons are **dependent on the outcome** produced with an AI system, in particular because for practical or legal reasons it is **not reasonably possible to opt-out** from that outcome;

F. the extent to which potentially harmed or adversely impacted persons are in a **vulnerable position** in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age;

G. the extent to which the outcome produced with an AI system is easily **reversible**, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible;

H. the extent to which legislation provides for: (i) effective measures of **redress** in relation to the risks posed by an AI system, with the exclusion of claims for damages; (ii) effective measures to **prevent** or substantially **minimise** those risks.
<table>
<thead>
<tr>
<th>Requirement</th>
<th>Details</th>
</tr>
</thead>
</table>
| Establish and implement **risk management processes** &                   | In light of the **intended purpose** of the AI system \  
| Establish **documentation** and design logging features (traceability & auditability) | Use high-quality **training, validation and testing data** (relevant, representative etc.)  
| Ensure appropriate certain degree of **transparency** and provide users with **information** (on how to use the system) | Ensure **human oversight** (measures built into the system and/or to be implemented by users)  
| Ensure **robustness, accuracy** and **cybersecurity**                       |                                                                                                                                          |
4) AD HOC BALANCE
Art. 9 - Risks Management System
§ 4. “The risk management measures referred to in paragraph 2, point (d) shall be such that any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable (…)”

Art. 17 - Quality Management System
§ 2. “The implementation of aspects referred to in paragraph 1 shall be proportionate to the size of the provider’s organisation”

Art. 61 - Post-Market Monitoring System
§ 1. “Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the artificial intelligence technologies and the risks of the high-risk AI system”
• SELF-VERIFIED QUALITY MANAGEMENT SYSTEM (art. 17)

• SELF-ASSESSED TECHNICAL DOCUMENTATION (art. 18)

• SELF-VERIFIED POST-MARKET MONITORING SYSTEM (art. 61)

• THIRD-PARTY CONFORMITY ASSESSMENT ONLY FOR BIOMETRIDI ID SYSTEMS
“By setting the standards, we can pave the way to ethical technology worldwide and ensure that the EU remains competitive along the way.”

*Margrethe Vestager, Executive Vice-President for a Europe fit for the Digital Age*

“Today’s proposals aim to strengthen Europe's position as a global hub of excellence in AI from the lab to the market, ensure that AI in Europe respects our values and rules, and harness the potential of AI for industrial use.”

*Thierry Breton, Commissioner for Internal Market*
Thanks

guido.smorto@unipa.it