DEFICIENCY OF "REDLINE/GREENLINE" APPROACH TO RISK MANAGEMENT IN AI APPLICATIONS

Andrey Kuleshov, CFA,
Moscow Institute of Physics and Technology
(a.kuleshov@phystech.ai)

Andrey Ignatiev,
MGIMO University
Center for Global IT Cooperation

Anna Abramova, PhD,
MGIMO University
(anna.vl.abramova@gmail.com)

REGULATORY CHALLENGE OF AI

- Failure to meet moral expectations is an impediment to technology: protecting human rights is a must
- Challenge for public authorities – no effective models for regulation of AI technologies
- Transborder impact of AI: need coordinated international approach
- Considerable momentum to replicate the "red line - green line " (RGL) regulations also for AI
- No common definition of AI – identify technologies based on feature that pose specific risks
- RGL not effective – risks depend on
  - Domain
  - Context
  - Actors, etc.

INTERNATIONAL DEBATE ON RISK MITIGATION FOR AI

- Council of Europe: CAHAI
  - Risk based approach reflecting specific features of AI, scale, connectedness, and reach
  - All agents in AI value chain play a role as risk factors
  - Re-calibration and learning by AI leading to emergent properties
- ISO/IEC
  - Trustworthiness – assess resilience, reliability, accuracy, security of AI systems
  - Bias – address at source
  - Risk management: no quick fixes! Identitfy, evaluate, prioritize
- OSCE
  - #SAIFE project – emphasis on ethical behaviour to protect free speech

INTERNATIONAL DEBATE ON RISK MITIGATION FOR AI (CONTD)

- OECD
  - Lifecycle approach to AI risk management
  - Part of framework for risk management in digital technologies
- European Commission
  - Risk based approach
  - Depends on uses of AI, where risk is
    - Unacceptable
    - High
    - Low
  - Codes of conduct for high risk applications

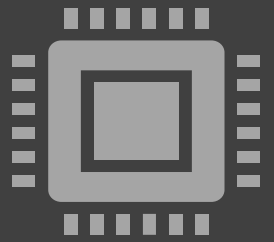
CHALLENGES AND DEFICIENCIES IN DEFINING RED AND GREEN LINES

Inherent risks of AI:

Risks due to autonomous behavior of AI: risk of scaling cascading failures

Risks due to superior learning capacity: in general, cannot address without losing substance of AI

Risks due to lack of transparency: data centricity and "dumbing down" of AI



Application risks of AI:

Risks due to domain dependence of narrow AI: performance of weak AI is only defined in its application domain

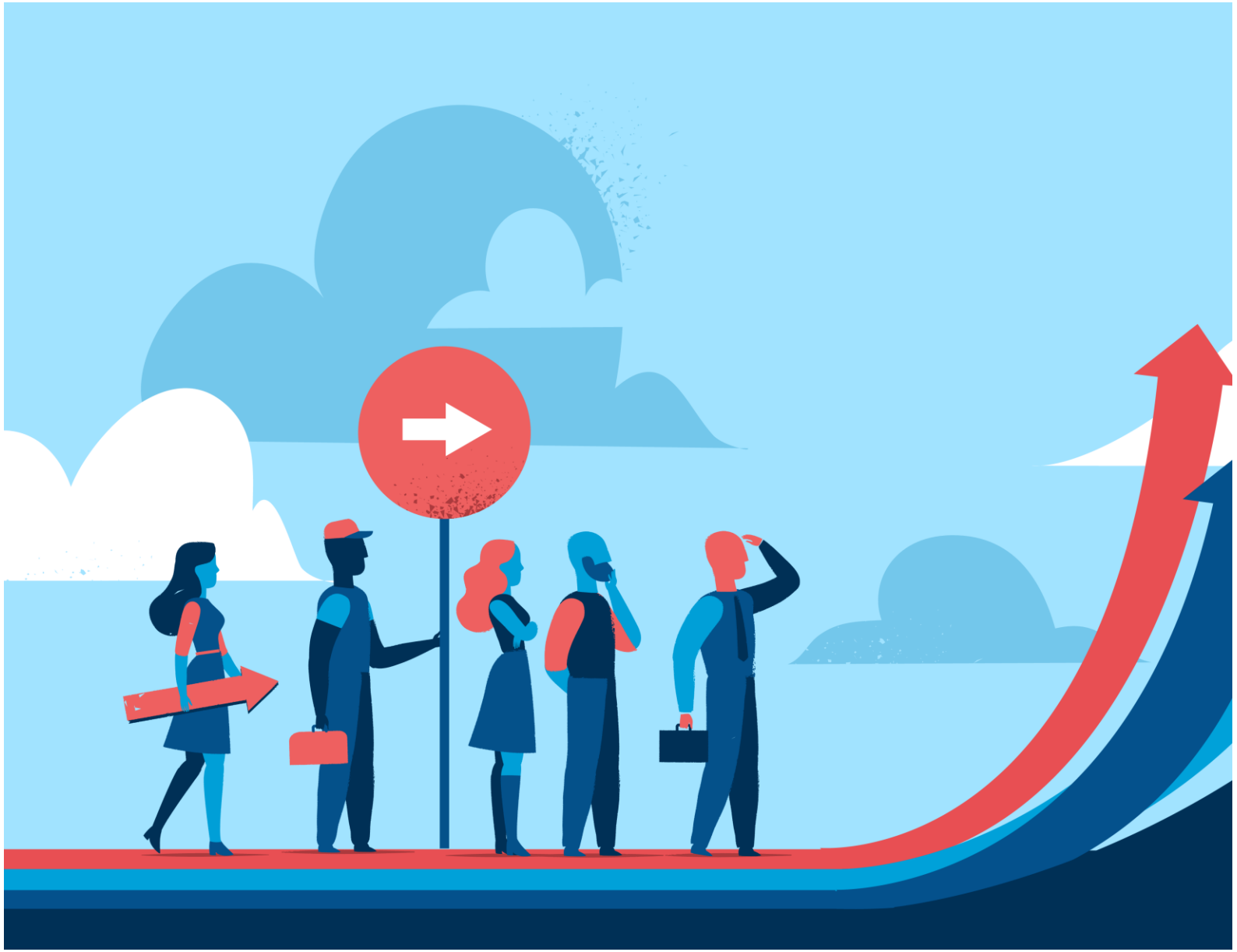
Malicious intent by actors of AI value chain: responsibility of actors

SEEKING A MORE NUANCED APPROACH

- "Redlining" AI *technologies* makes AI virtually unusable in most practical applications with meaningful positive contribution to society
  - unsurmountable legal barriers
  - Competitive pressures and regulatory arbitrage
  - Non-verifiable compliance
- A "nuanced" approach targets applications, not technologies
- Regulatory instruments implementable at application level:
  - Standardization
  - Certification, licensing and regulatory sandboxes
  - Using code of ethics
  - AI risk management guidelines

CONCLUSIONS

- The simplicity of RGL approach illusory, practical use complicated to the point of ineffectiveness
- Transparent liability of actors, standards of responsible behaviour
- New regulations targeting malicious uses of unique features of AI
- Has to be implemented by all actors:
  - Individuals and civil society: observe the law and ethical standards relevant to application
  - Corporations: further, provide accurate and complete information
  - Government: support the regulatory infrastructure



THANK YOU!

A.KULESHOV@PHYSTECH.AI

ANNA.VL.ABRAMOVA@GMAIL.COM