

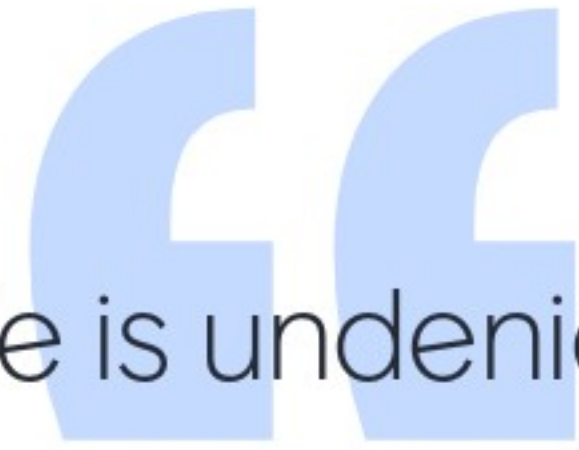
AI policy and in practice - concerns and opportunities

Tina Đaković

organisational coordinator of Human Rights House Zagreb

tina.djakovic@kucaljudskihprava.hr





The power of AI to serve people is undeniable, but so is AI's ability to feed human rights violations at an enormous scale with virtually no visibility. Action is needed now to put human rights guardrails on the use of AI, for the good of all of us

– Michelle Bachelet, United Nations High Commissioner for Human Rights (15th September 2021)

What is new and unique?

- lack of transparency and explainability
- lack of quality data
- privacy and autonomy risks
- inequalities, discrimination and bias
- over-use of AI
- large impact
- and more!!!

Impact on Human Rights:

- Liberty and Security; Fair Trial; No Punishment without Law; Effective remedy (Art. 5, 6, 7, 13 ECHR)
- Private and Family Life; Physical, Psychological and Moral Integrity (Art. 8 ECHR)
- Freedom of expression; Freedom of assembly and association (Art. 10, 11 ECHR)
- Equality and Non-Discrimination (Art. 14 ECHR, Protocol 12)
- Social and Economic Rights (Art. 2, 3, 5, 11, 12, 13 and 20 ESC)
- democracy, rule of law

Transparency Center

Facebook

Download (CSV)

Overview

Adult Nudity and Sexual Activity

Bullying and Harassment

Child Endangerment: Nudity and Physical Abuse and Sexual Exploitation

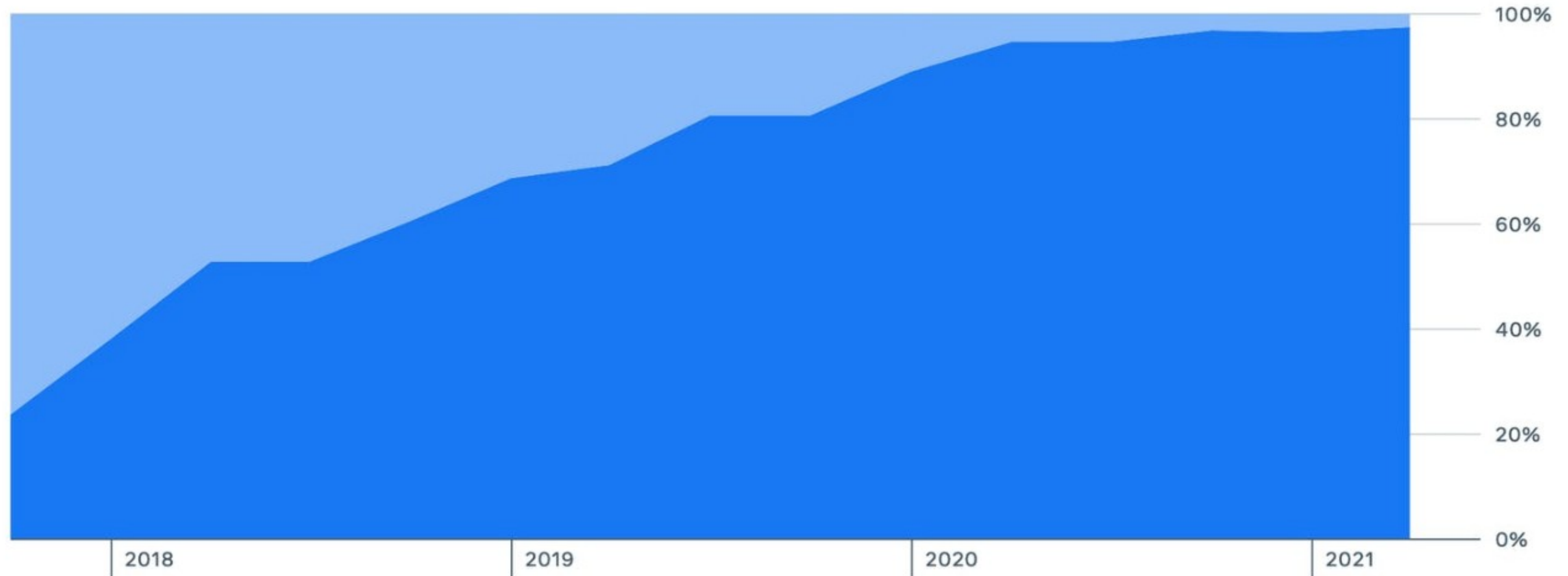
Dangerous Organizations: Terrorism and Organized Hate

Fake Accounts

Hate Speech

Regulated Goods: Drugs and

Of the violating content we actioned for hate speech, how much did we find before people reported it?



Found and flagged by us Reported by users



THE WALL STREET JOURNAL.

the facebook files

Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts.

AI has only minimal success in removing hate speech, violent images and other problem content, according to internal company reports



AI hate speech content moderation

- accuracy
- “platforms have reached a scale where only AI solutions seem viable; AI solutions allow platforms to grow further.” (Gillespie 2020)
- transparency and accountability
- bias - data quality, "neutrality"
- de-politicisation of the politics of content moderation?
- overblocking



What to do?

- engage in policy processes - engage underrepresented groups
- regulatory limits on deployments of AI that unduly restrict human rights
- transparency, human review
- independent human rights impact assessments
- rather than the individuating aim of removing discrete pieces of content, moderation technologies should aim to repair, educate and sustain communities
- (AI Content Moderation, Racism and (de)Coloniality, E. Siapera: <https://link.springer.com/article/10.1007/s42380-021-00105-7>)

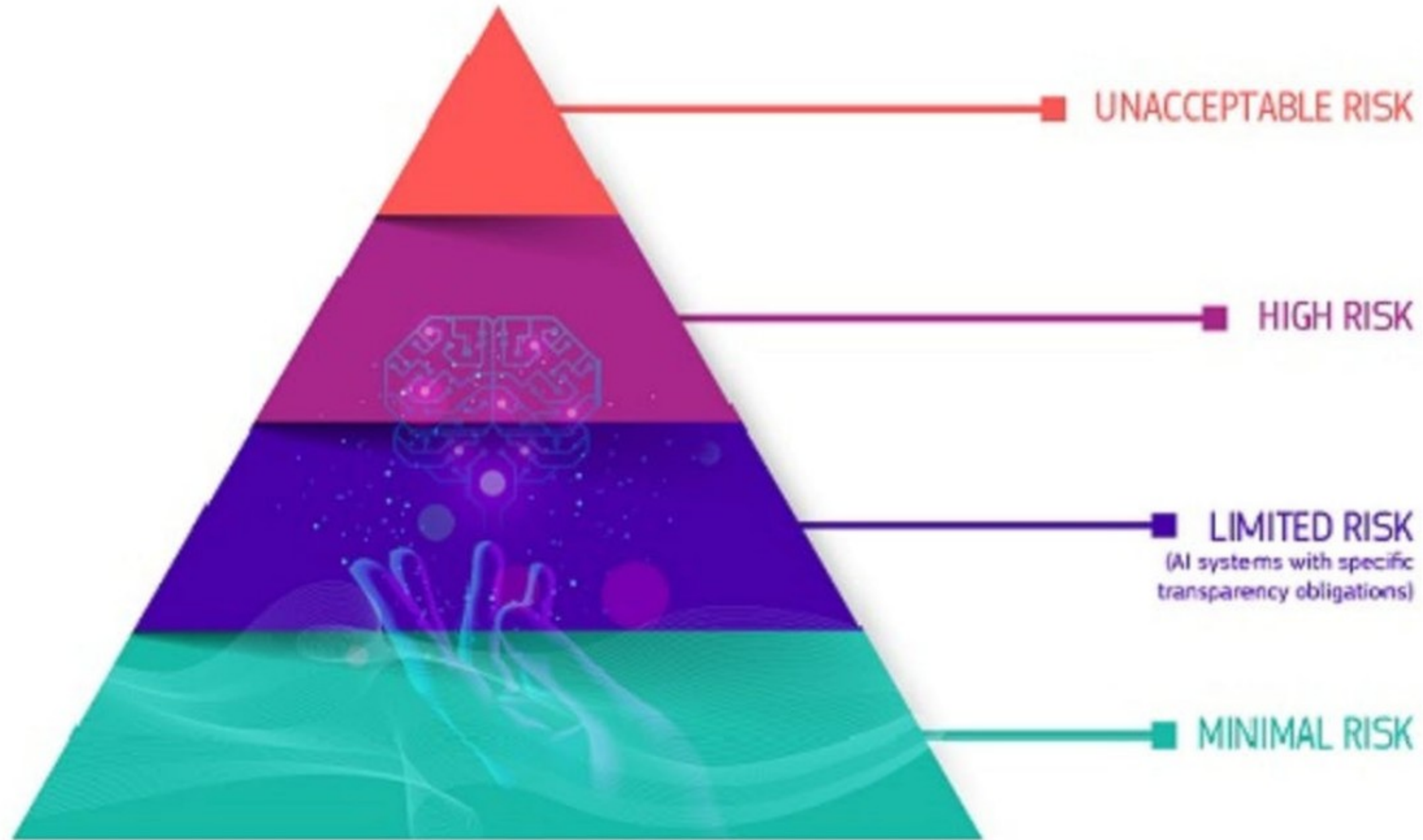


Council of Europe

It is crucial that the Council of Europe's standards on human rights, democracy and the rule of law are effectively anchored in appropriate legislative frameworks by member States. While the existing general international and regional human rights instruments, including the European Convention on Human Rights (ECHR), remain applicable in all areas of life, a Council of Europe legal response, aimed at filling legal gaps in existing legislation and tailored to the specific challenges raised by AI systems should be developed, based on broad multi-stakeholder consultations.

<https://www.coe.int/en/web/artificial-intelligence/cahai>

A risk-based approach



The EU's approach to artificial intelligence centres on excellence and trust, aiming to boost research and industrial capacity and ensure fundamental

AI Watch

National strategies on Artificial Intelligence

A European perspective

2021 EDITION — a JRC-OECD report

National policy / strategy

By June 2021, 20 EU Member States and Norway have published their national AI strategies, while 7 Member States are in the final drafting phase:

policy initiatives across various policy areas, including human capital (i.e. educational development), from the lab to the market (i.e. research & development, innovation, business and public sector development), networking (i.e. collaboration and dissemination), regulation (i.e. ethical guidelines, legislation and standardisation) and infrastructure (i.e. data and telecommunication infrastructure)



“
EU’s AI law needs major changes to prevent
discrimination and mass surveillance -
fundamental changes are needed for a truly
people-centred AI law.

– EDRI - European Digital Rights Initiative: <https://edri.org/our-work/eus-ai-law-needs-majo>

