KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

imec

From Jurassic Park to Westworld – Why Protecting Human Rights in the AI Era needs more than an Ethical Compass

Conference Artificial Intelligence and the Law

May 5, 2021

Prof. Dr. Peggy Valcke


JURASSIC
PARK

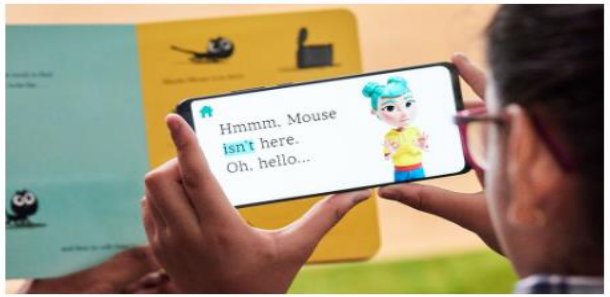

YOU WERE SO PREOCCUPIED WITH WHETHER OR NOT YOU COULD
YOU DIDN'T STOP TO THINK IF YOU SHOULD

From dinosaurs to AI…

AI is powerful

Vision, Tools

Empowering students who are blind with I-Assistant



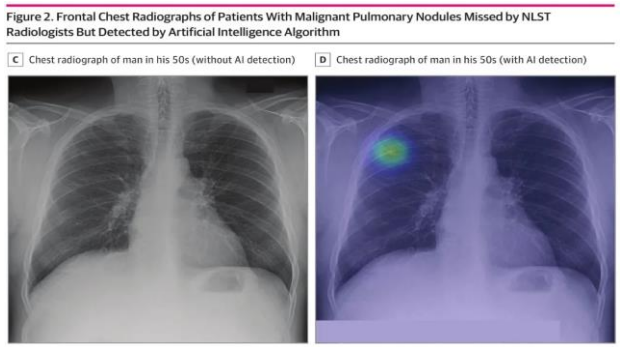
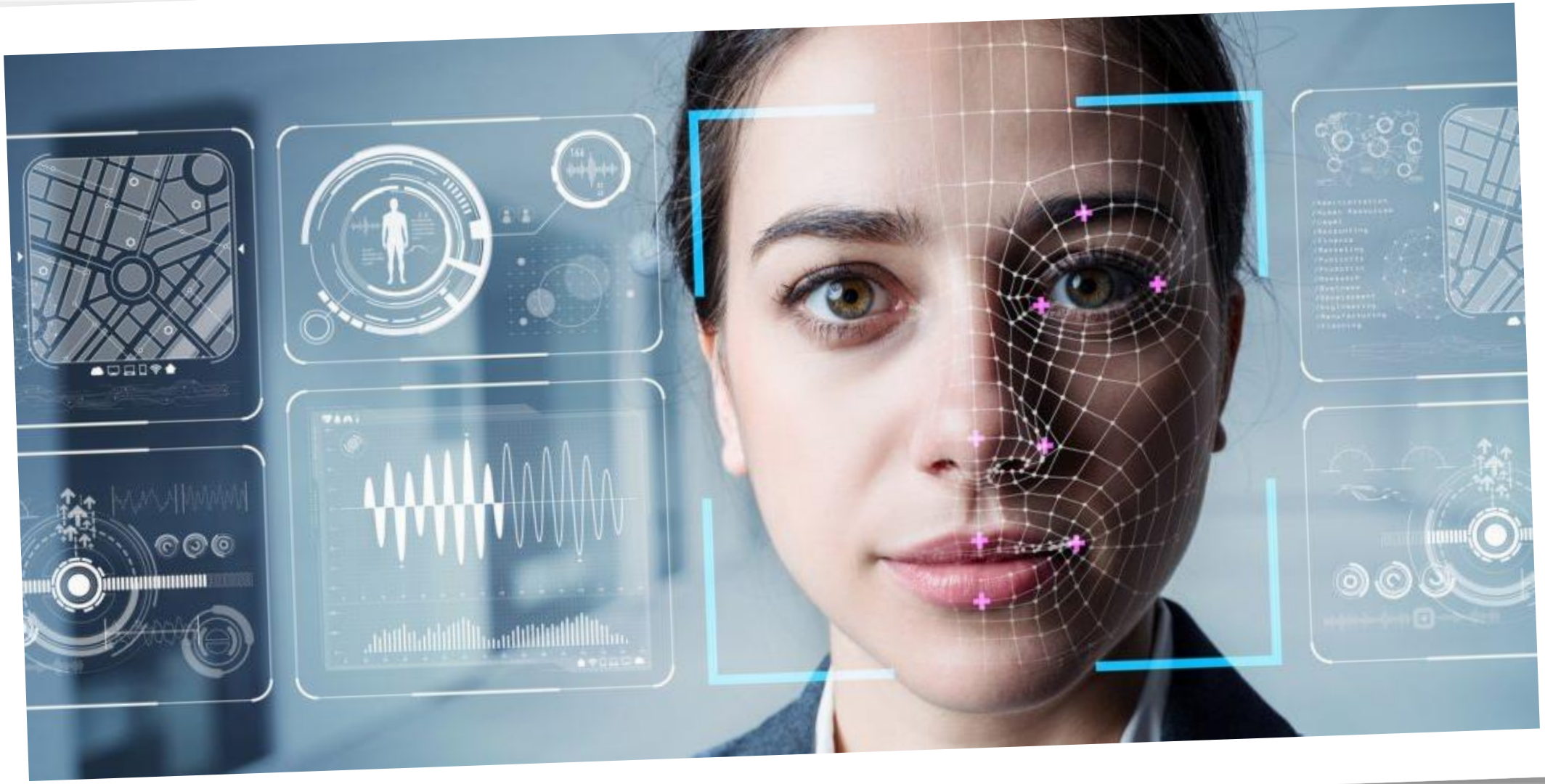


Figure 2. Frontal Chest Radiographs of Patients With Malignant Pulmonary Nodules Missed by NLST Radiologists But Detected by Artificial Intelligence Algorithm

C Chest radiograph of man in his 50s (without AI detection)

D Chest radiograph of man in his 50s (with AI detection)




ID 646300628981
ID 324267334567
ID : 254876592
MALE
BROWN HAIR
CAUCASIAN
STRESSED
ID : 548765942
MALE
GREY HAIR
CAUCASIAN
RELAXED
BAG
SYSTEM
RECOGNITION
IN PROGRESS ...
27%
ID : 258654892
FEMALE
CAUCASIAN
RUNNING
BAG
ID : 458625
MALE
CAUCASIAN
RELAXED
BAG
Visitor
Erika Mustermann
Staff
John Doe
Contractor
Jane Doe

AI can now read emotions

January 8, 2020 12.18pm GMT

Emotion recognition technology, an outgrowth of facial recognition technology, continues to advance quickly. Steve Jurvetson/flickr, CC BY-SA

FACEPTION
Facial Personality Analytics

Our Technology Verticals About us News & Events Blog Contact us

FACEPTION IS A FACIAL PERSONALITY ANALYTICS TECHNOLOGY COMPANY

We reveal personality from facial images at scale to revolutionize how companies, organizations and even robots understand people and dramatically improve public safety, communications, decision-making, and experiences.

OUR CLASSIFIERS


High IQ


Academic Researcher


Professional Poker Player


Terrorist

Utilizing advanced machine learning techniques we developed and continue to evolve an array of classifiers. These classifiers represent a certain persona, with a unique personality type, a collection of personality traits or behaviors. Our algorithms can score an individual according to their fit to these classifiers.

Personal

ng your feelings, at the touch of a button.

For You

For Wor

Measuring the happi people and teams, e

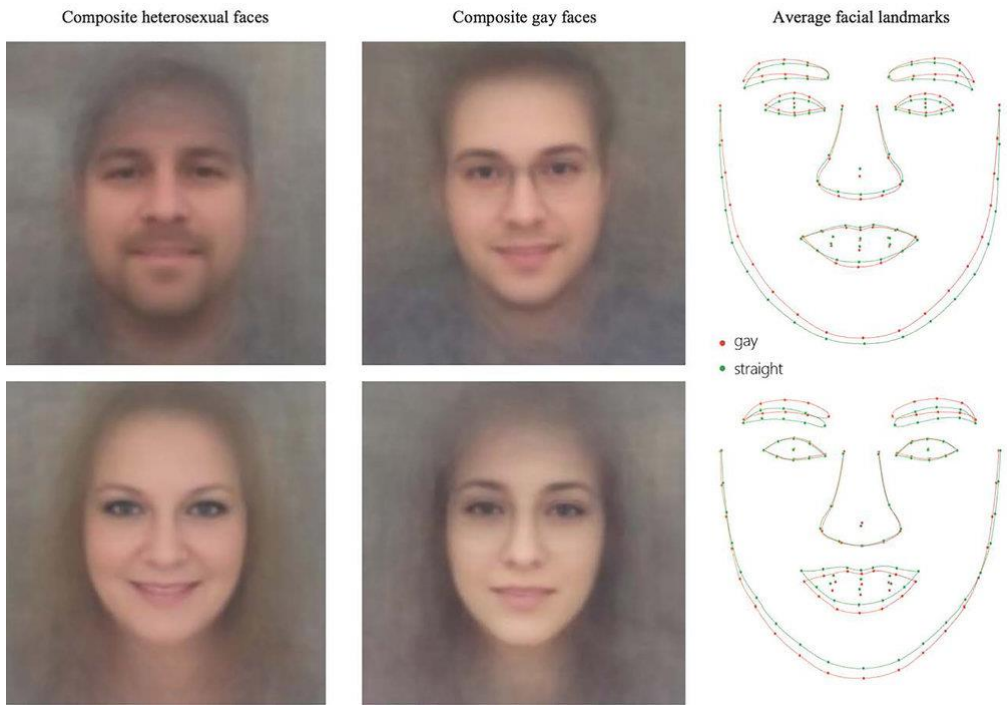
For Work

http://gendershades.org/

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

The Stanford gaydar AI is hogwash

Stanford team behind BS gaydar AI says facial recognition can expose political orientation



Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

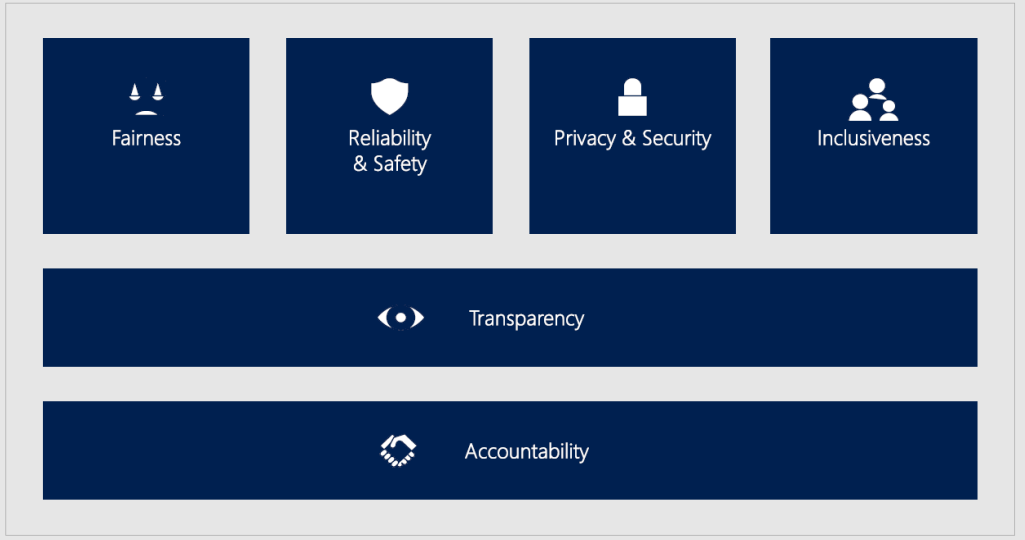
Contributors: Yilun Wang, Michal Kosinski

Date created: 2017-02-15 05:37 PM | Last Updated: 2020-05-26 12:11 AM

Identifier: DOI 10.17605/OSF.IO/ZN79K


YOU WERE SO PREOCCUPIED WITH WHETHER OR NOT YOU COULD
YOU DIDN'T STOP TO THINK IF YOU SHOULD

Microsoft's ethical design framework for AI



European Commission

INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION

AI

ETHICS GUIDELINES FOR TRUSTWORTHY AI

IEEE

Advancing Techno for Huma

ETHICALLY ALIGNED DESIGN

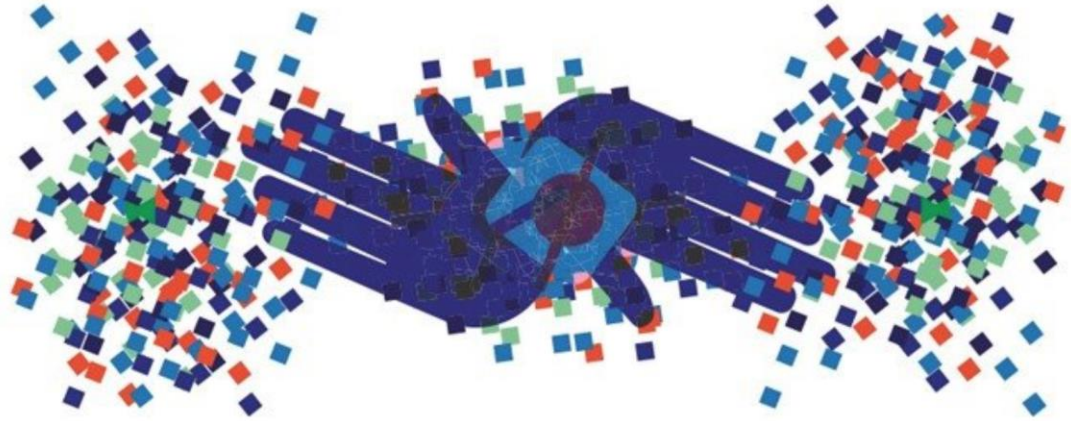
First Edition Overview

A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems

United Nations Educational, Scientific and Cultural Organization

Recommendation on the Ethics of Artificial Intelligence

ICC-UNESCO Roundtable



Systematic review & analysis of 84 AI ethics guidelines

published until 4/23/2019

https://www.nature.com/articles/s42256-019-0088-2

Documents issued by:

Public sector ~31%
26 documents from governmental org. & IGOs

Private sector ~27%
23 documents from companies & private sector alliances.

Academic
/research institutions, NPOs, professional assoc./scientific societies, etc.

nature machine intelligence

PERSPECTIVE

https://doi.org/10.1038/s42256-019-0088-2

The global landscape of AI ethics guidelines

Anna Jobin, Marcello Ienca and Effy Vayena*

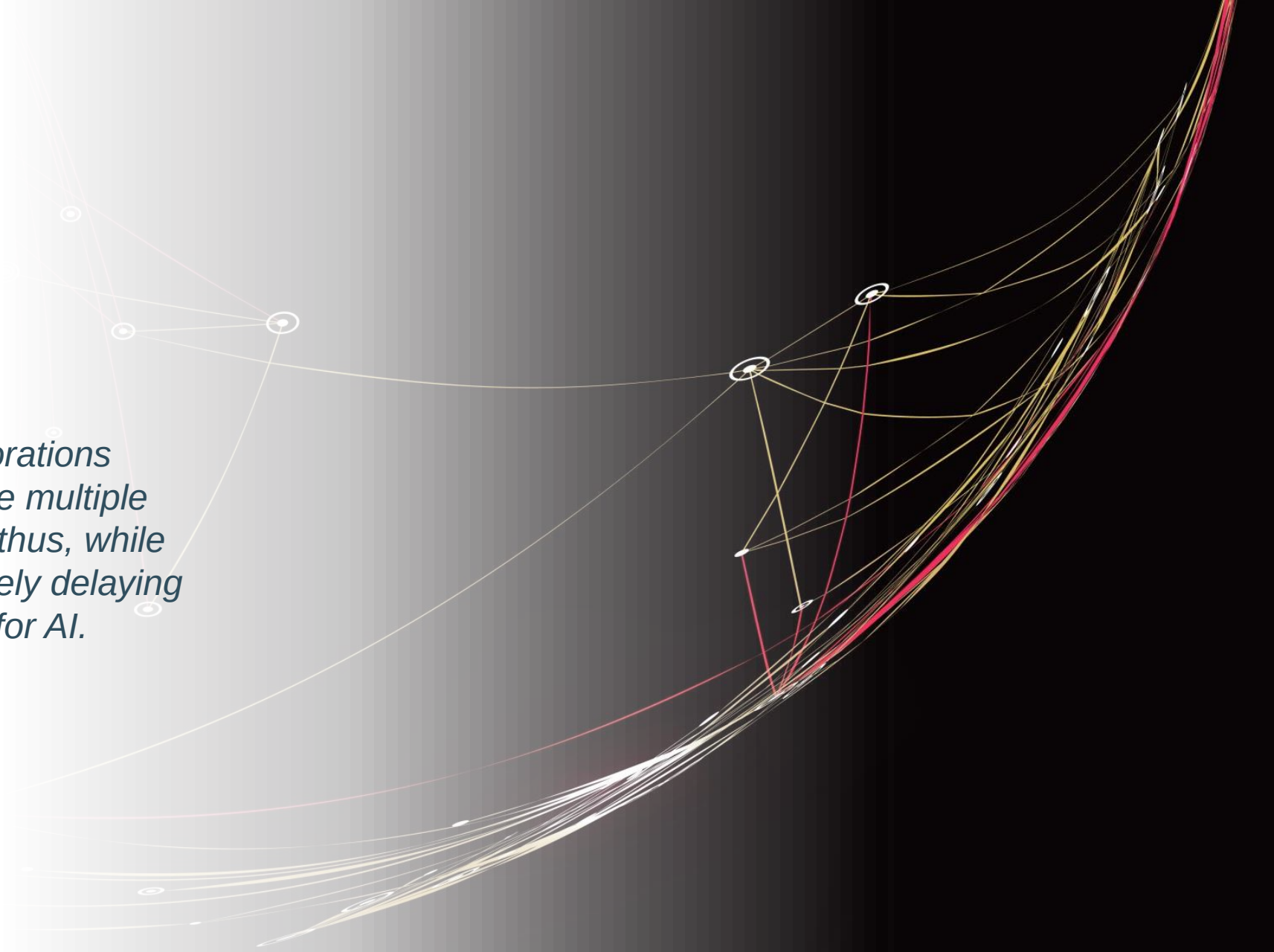
In the past five years, private companies, research i guidelines for ethical artificial intelligence (AI). How is debate about both what constitutes 'ethical AI' and needed for its realization. To investigate whether a glo the current corpus of principles and guidelines on ethi cal principles (transparency, justice and fairness, non relation to how these principles are interpreted, why t and how they should be implemented. Our findings hig substantive ethical analysis and adequate implementa

No single ethical principle common to all documents, but emerging convergence around transparency, justice/fairness, non-maleficence, responsibility, & privacy.

Ethical principle	Number of documents	Included codes
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Justice and fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	60/84	Responsibility, accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good
Freedom and autonomy	34/84	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment
Trust	28/84	Trust
Sustainability	14/84	Sustainability, environment (nature), energy, resources (energy)
Dignity	13/84	Dignity
Solidarity	6/84	Solidarity, social security, cohesion

In a move of genius, the corporations interested have started to finance multiple initiatives to work on ethics of AI, thus, while pretending best intentions, effectively delaying the debate and work on law for AI.

(Paul Nemitz, 2018)



The law will stifle innovation and make life of start-ups very difficult!

The law is not able to develop as fast as technology and business models!

The law is not precise enough to regulate complex technology!

Experience with internet regulation

- Californian ideology: ‘better ask forgiveness than permission’ attitude (‘Google Books’: conflict with copyright laws and ‘Uber’: with labour law and regulation of public transport).
- Disruptive innovation
- Common denominator: evade responsibility
- Google took away a ‘lesson’, namely to consider ‘lobbyists and lawyers’ earlier to be able to play the ‘sometimes’ necessary game of politics.

A Declaration of the Independence of Cyberspace

by John Perry Barlow

Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.

We have no elected government, nor are we likely to have one, so I address you with no greater authority than that with which liberty itself always speaks. I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear.

Governments derive their just powers from the consent of the governed. You have neither solicited nor received ours. We did not invite you. You do not know us, nor do you know our world. Cyberspace does not lie within your borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions.

The law will stifle innovation and make life of start-ups very difficult!

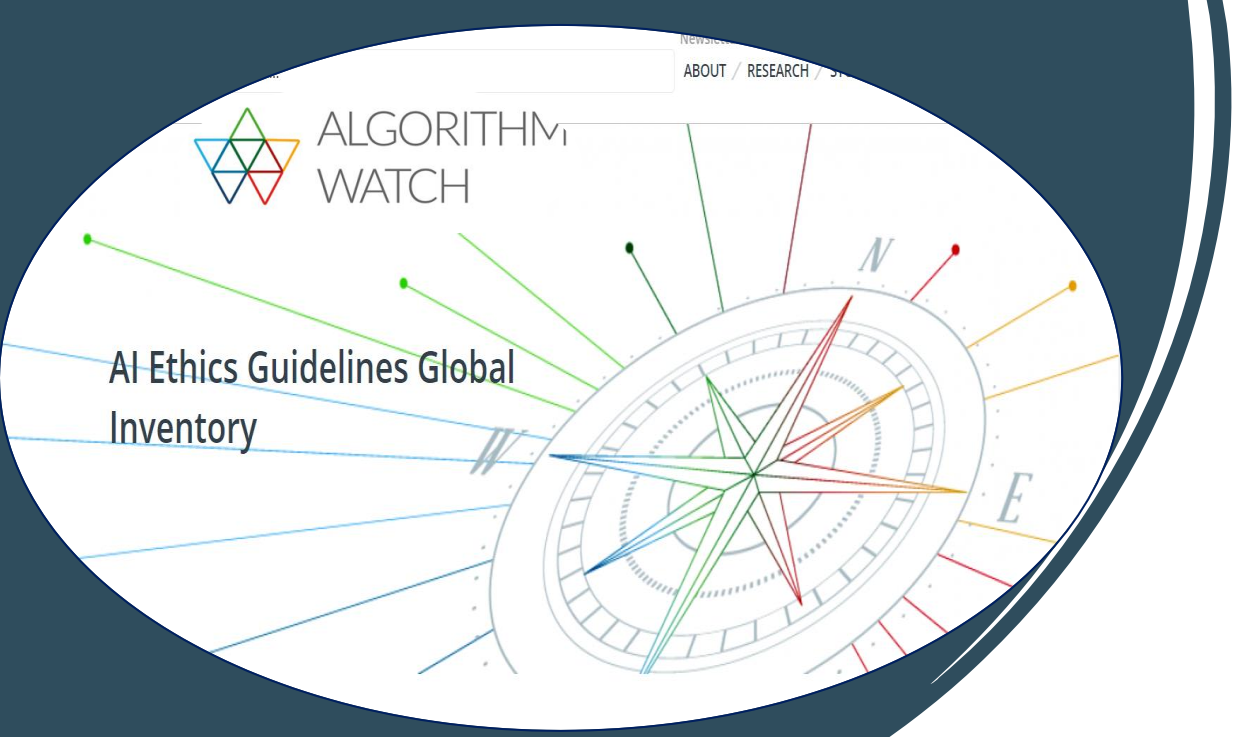
Law and innovation are not opposites…

The law is not able to develop as fast as technology and business models!

Disproved by the continued application of good, technology neutral law (both in Europe and the US…)

The law is not precise enough to regulate complex technology!

Law ≠ computer code; = result of democratic compromise; precisely its openness and contestability allows it to adapt to new circumstances…



It is time to move on to the crucial question in democracy, namely which of the challenges of AI can be safely and with good conscience left to ethics, and which challenges of AI need to be addressed by rules which are enforceable and based on democratic process, thus laws…

"principle of essentiality"

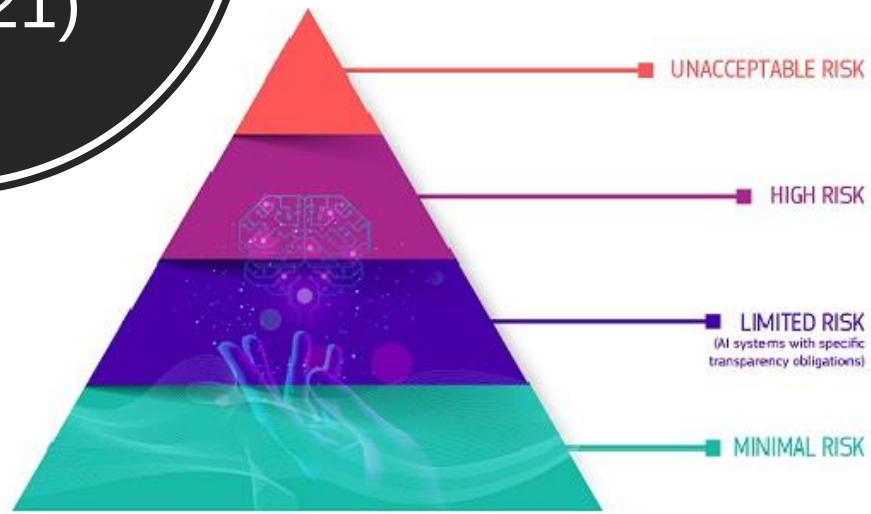
Who is liable?

Who is speaking: man or machine?

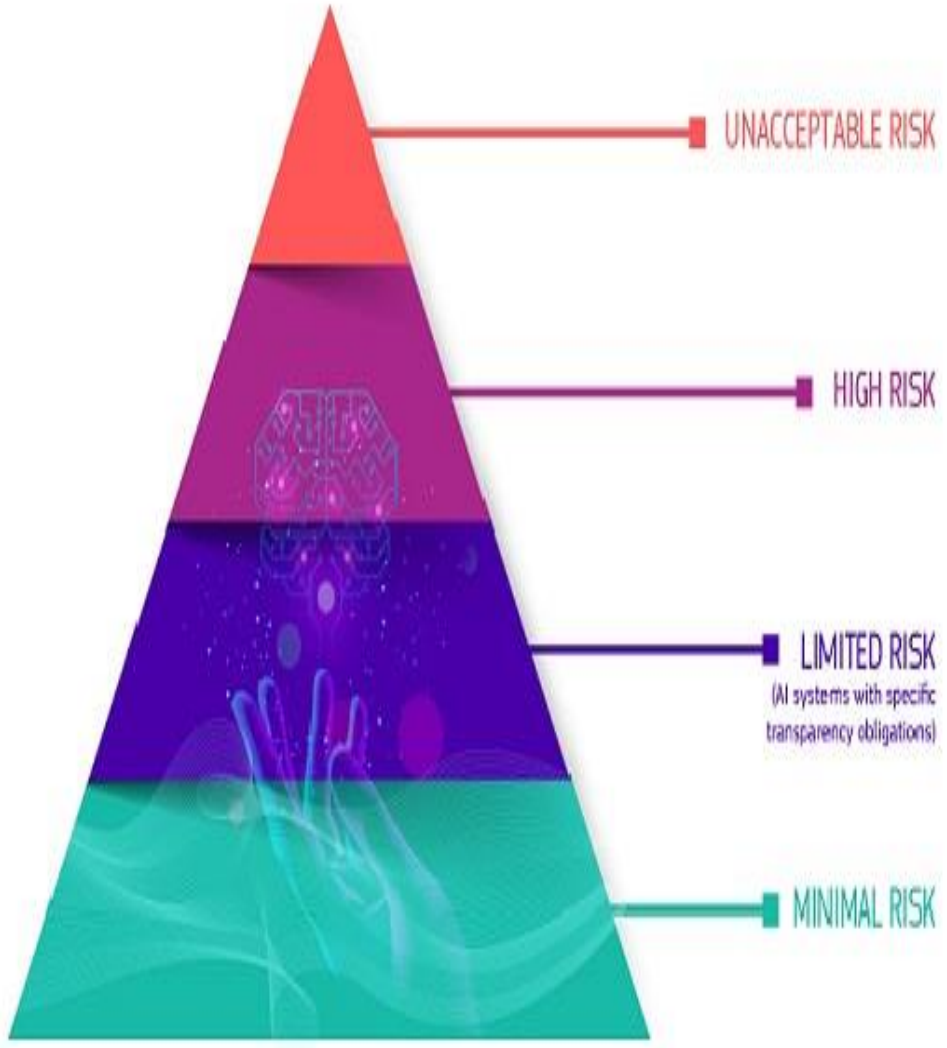
How to carry out *ex ante* impact assessment; *ex post* monitoring?

How to remain in control?

EU proposal AI Act (21.4.21)



Towards a European Convention on AI…?



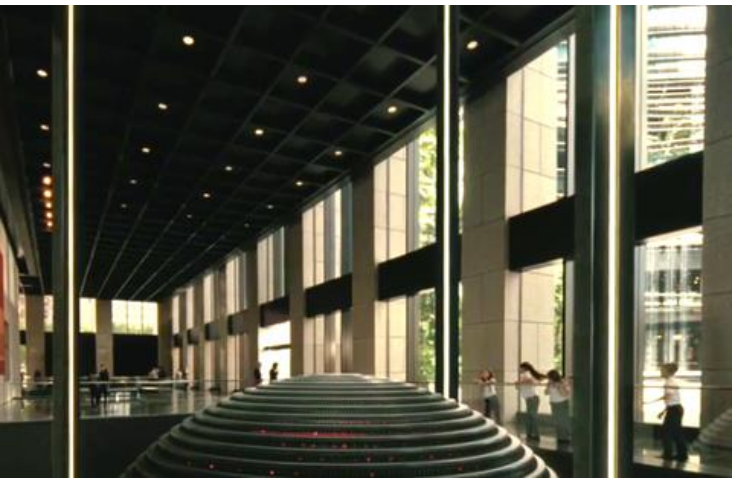
Unacceptable risk: All AI systems considered a clear threat to the safety, livelihoods and rights of people will be banned, from social scoring by governments to toys using voice assistance that encourages dangerous behaviour.

High-risk AI systems will be subject to strict obligations before they can be put on the market:

- Adequate risk assessment and mitigation systems;
- High quality of the datasets feeding the system to minimise risks and discriminatory outcomes;
- Logging of activity to ensure traceability of results
- Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance;
- Clear and adequate information to the user;
- Appropriate human oversight measures to minimise risk;
- High level of robustness, security and accuracy.

Limited risk, i.e. AI systems with specific transparency obligations: When using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.

Minimal risk: The proposal allows the free use of applications such as AI-enabled video games or spam filters. The vast majority of AI systems currently used in the EU fall into this category, where they represent minimal or no risk.



Thank you for your attention!