Ethical aspects of AI
The viewpoint of the EU

Dr Mihalis Kritikos

Ethics and Research Integrity Sector-DG Research and Innovation-European Commission

14 Μαρτίου 2023

AI – Regulatory/Legal challenges

- **Definition of AI: variety of definitions and approaches**
- **Disruptive and horizontal character of AI**
- **Looking for new risk identification/assessment methodologies**
- **Need for balancing between technological innovation and ethical governance**
- **Management of scientific uncertainties and social concerns**

Ethical challenges

- Black-box effectsΑδιαφάνεια
- Bias/discriminatory effects
- Autonomy
- Απόρρητο πληροφοριών και ιδιωτικοτητα
- Ασφάλεια και ανθεκτικότητα
- Ηθικός έλεγχος

European Commission

Shaping a European approach towards AI

- Focus on:

-Human-centric AI

-Trustworthy AI

-AI ethics by design

- Creation of **an ecosystem of excellence and of an ecosystem of trust**

- **EU aspires to become a global leader in the development of trustworthy AI**

What is the EU doing on AI in policy/regulatory terms?

- **Legislative proposal for an AI Act**
- **Communications, Coordinated Plans and a White Paper**
- **Ethics Guidelines (Requirements, ALTAI, sector-specific guidance)**
- **Standardisation initiatives, sandboxing and international outreach**

proposal

Regulation on a
European approach for
Artificial Intelligence
[21st of April 2021]

Subject matter and scope

Title I

List of prohibited AI

Title II

High-risk AI systems

Title III

Transparency obligations

Title IV

Support of innovation

Title V

Governance systems

Title VI, VII, VIII

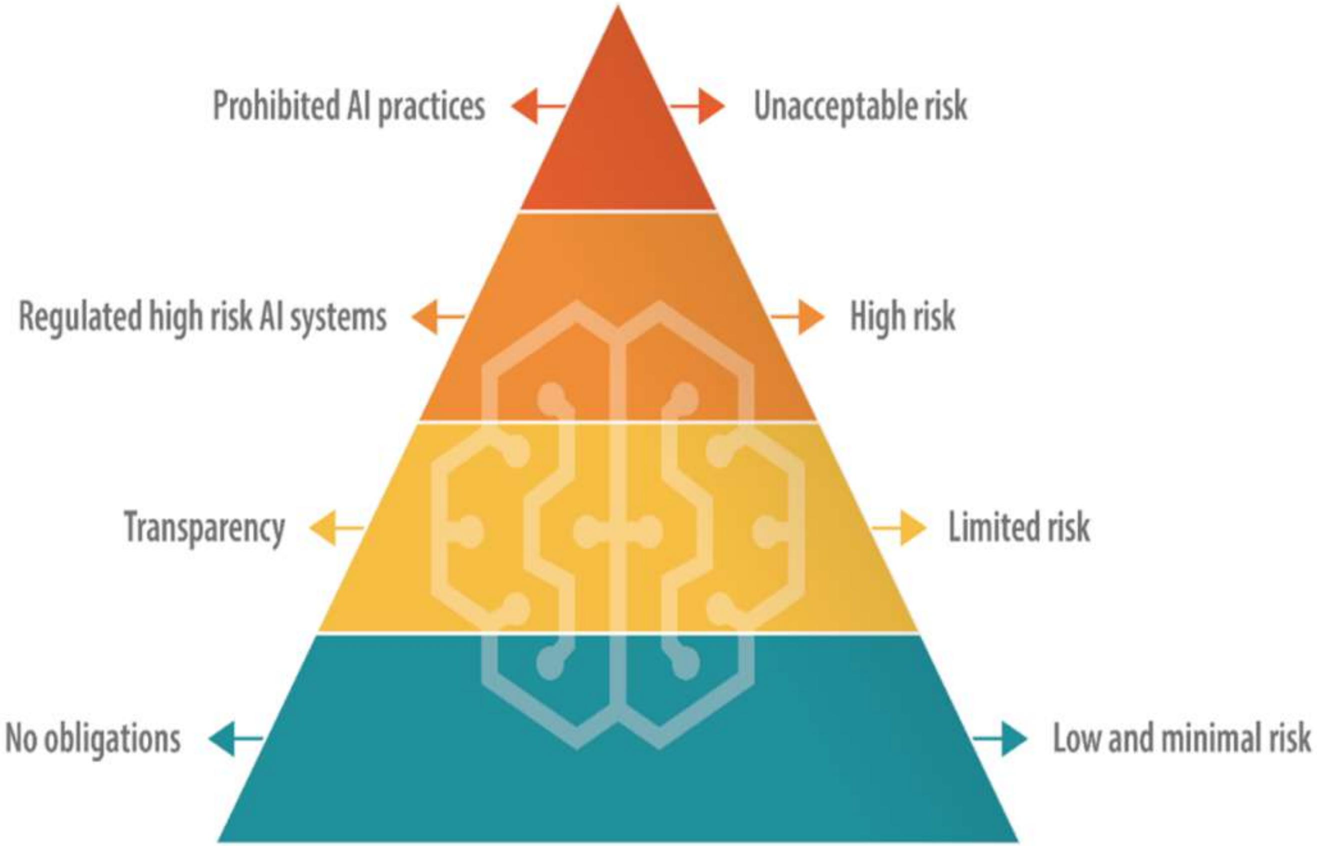
Voluntary commitments

Title IX

Penalties, delegation of power, updates

Title X, XI, XII

AI Act proposal



Prohibited AI practices ← → Unacceptable risk

Regulated high risk AI systems ← → High risk

Transparency ← → Limited risk

No obligations ← → Low and minimal risk

Data source: European Commission.

Four categories of potential risk

- **Minimal risk:** the new rules won't apply to these AI systems because they represent only minimal or no risk for citizen’s rights or safety. Companies and users will be free to use them.
- **Limited risk:** subject to specific transparency obligations to allow users to make informed decisions, be aware they are interacting with a machine and let them easily switch off.
- **High risk:** given their potentially harmful or damaging implications on people's personal interests, these AI systems will be "carefully assessed before being put on the market and throughout their lifecycle“
- **Unacceptable:** the Commission will ban AI systems that represent "a clear threat to the safety, livelihoods and rights of people".

INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION

ETHICS GUIDELINES FOR TRUSTWORTHY AI

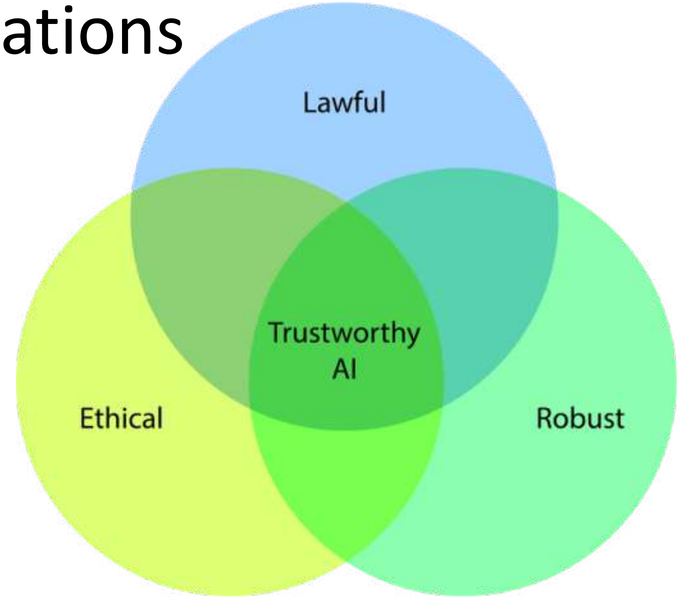
Seven requirements

- Ensure that the development, deployment and use of AI systems meets the **seven key requirements for Trustworthy AI:**
    - **(1) human agency and oversight,**
    - **(2) technical robustness and safety,**
    - **(3) privacy and data governance,**
    - **(4) transparency,**
    - **(5) diversity, non-discrimination and fairness,**
    - **(6) environmental and societal well-being**
    - **(7) accountability.**

EU (trustworthy) approach to AI

Trustworthy AI should be:

- (1) lawful - respecting all applicable laws and regulations
- (2) ethical - respecting ethical principles and values
- (3) robust - from a technical perspective

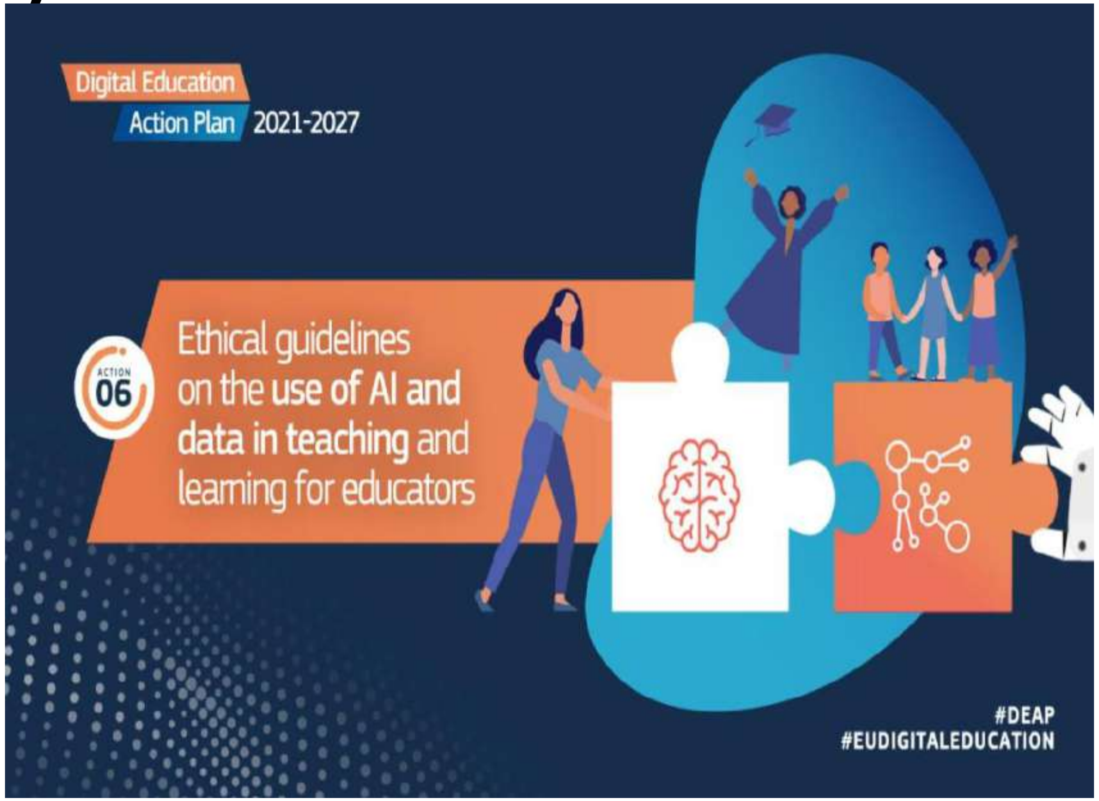


Ethical guidelines on the use of AI in teaching and learning for educators (EU, October 2022)

- The guidelines are intended for primary and secondary teachers

- Ethical requirements and practical advice are offered to educators and school leaders on how to integrate the effective use of AI into school education.

- The guidelines discuss emerging competences for the ethical use of AI among teachers, proposing ways of raising awareness and engaging with the community.



European Commission

EC Ethics Review and AI

- **AI as a separate box in the ethics issues check-list**
- AI as separate section in the **Guide on How to complete your ethics self-assessment**
- AI as separate section in **the Guidance Note on identifying serious and complex ethics issues in EU-funded research**
- **Special Guidance on Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**
- **Dedicated Ethics Checks on AI-related projects**

What is coming

- **Guidance Note on *human-centered AI: algorithmic bias and fairness***

- **Guidance Note on *explainable/inclusive AI***

- **Guidance Note on *AI Ethics Audits and Checks***

- **Guidance Note on *AI Ethics and project lifecycle***

European Parliament

- **More than 10 Resolutions on AI**

- **Committee reports on the proposed AI Act - (July 2022 EP Report: 3000 amendments)**

- **Temporary AI Committee (AIDA)**

EP’s AI Act negotiating team

- The discussions will be led by the **Committee on Internal Market and Consumer Protection** (IMCO; rapporteur: Brando Benifei, S&D, Italy) and the **Committee on Civil Liberties, Justice and Home Affairs** (LIBE; rapporteur: Dragos Tudorache, Renew, Romania) under **a joint committee procedure.**

- The **Legal Affairs Committee (JURI), the Committee on Industry, Research and Energy (ITRE) and the Committee on Culture and Education (CULT)** are associated to the legislative work with **shared and/or exclusive competences.**

AIDA Activity Report

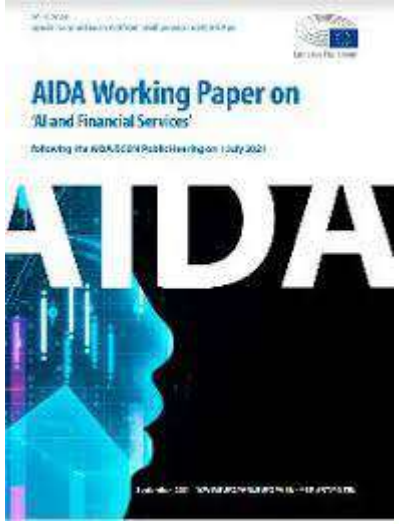
Special Committee on Artificial Intelligence in a Digital Age

European Parliament

ACTIVITY REPORT

following AIDA Mandate 2020-2022

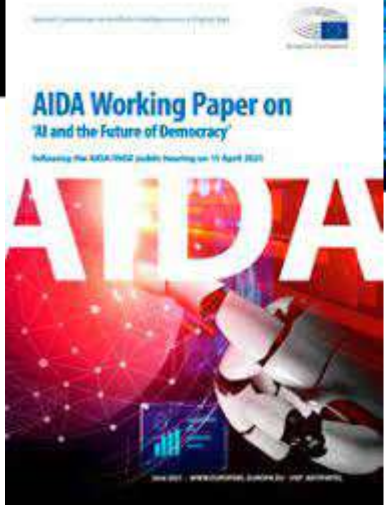
| September 2020 - March 2022



AIDA Working Paper on

'AI and Financial Services'

following the AIDA/ECON Public Hearing on 1 July 2021



AIDA Working Paper on

'AI and the Future of Democracy'

following the AIDA/INGE public hearing on 15 April 2021

AIDA Working Paper on

AI and Bias

Following the AIDA Public Hearing on 30 November 2021

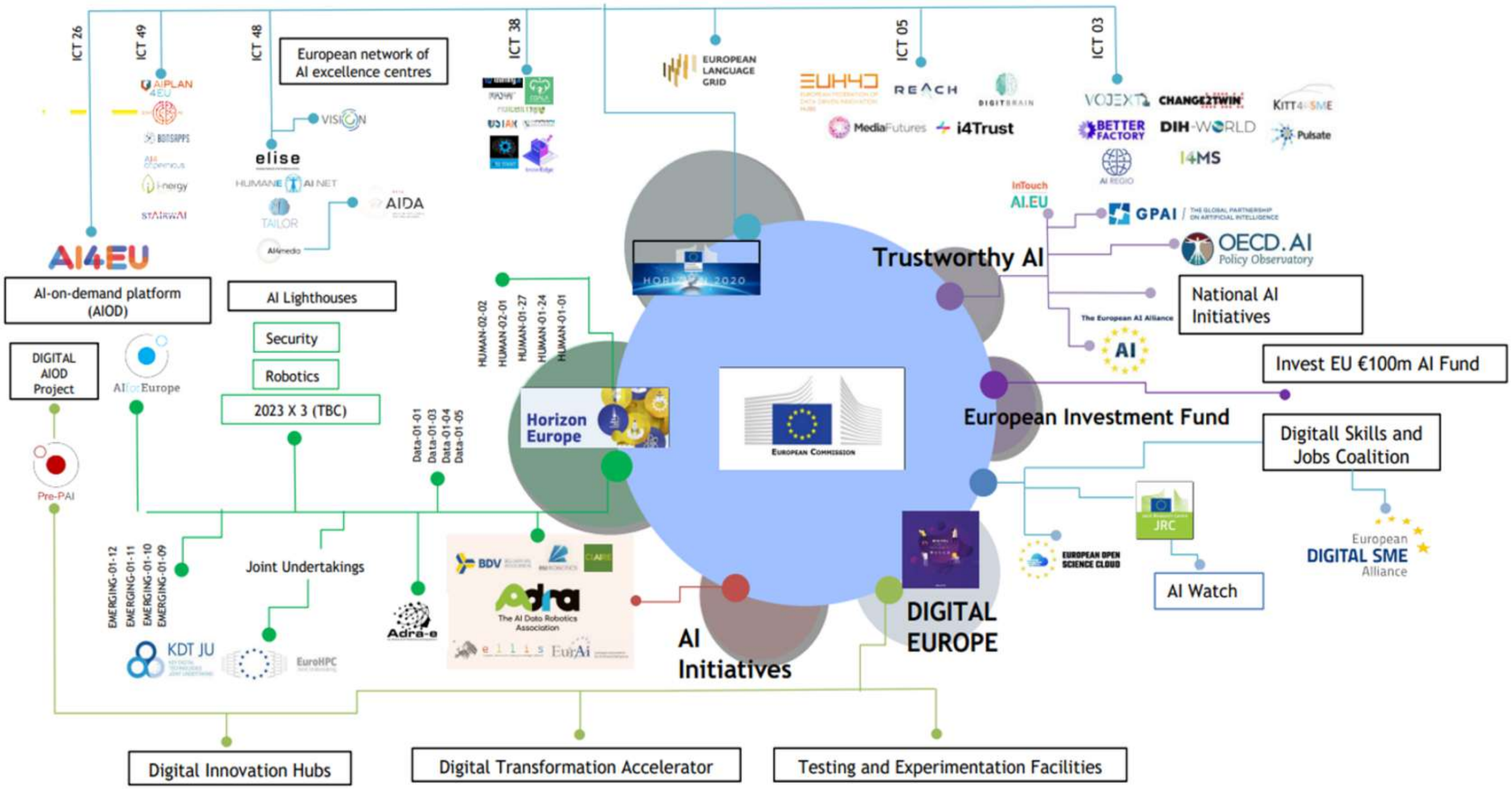
Resolution on the ethical aspects of AI

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL))

Future laws should be made in accordance with several guiding principles, including:

- **a human-centric and human-made AI;**
- **safety, transparency and accountability;**
- **safeguards against bias and discrimination;**
- **right to redress;**
- **social and environmental responsibility;**
- **European certificate of ethical compliance;**

Communities – The European AI Landscape



ICT 26

ICT 49

ICT 48

European network of AI excellence centres

ICT 38

ICT 05

ICT 03

AI-on-demand platform (AIOD)

AI Lighthouses

Security

Robotics

2023 X 3 (TBC)

DIGITAL AIOD Project

HUMAN-02-02

HUMAN-02-01

HUMAN-01-27

HUMAN-01-24

HUMAN-01-01

Data-01-01

Data-01-03

Data-01-04

Data-01-05

EMERGING-01-12

EMERGING-01-11

EMERGING-01-10

EMERGING-01-09

Joint Undertakings

Trustworthy AI

National AI Initiatives

Invest EU €100m AI Fund

European Investment Fund

Digitall Skills and Jobs Coalition

AI Watch

AI Initiatives

DIGITAL EUROPE

Digital Innovation Hubs

Digital Transformation Accelerator

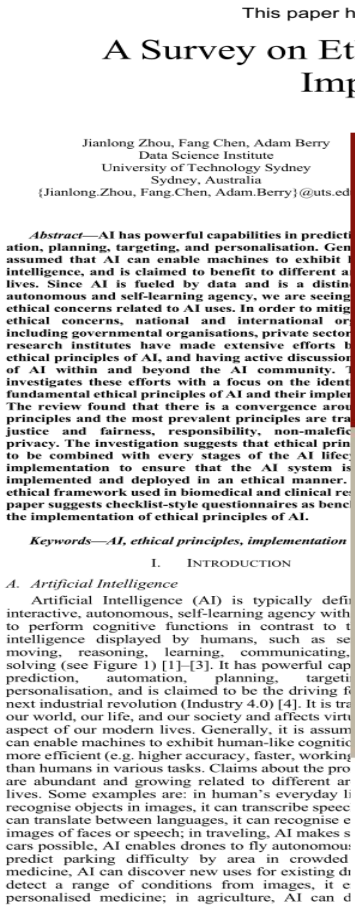
Testing and Experimentation Facilities

nature machine intelligence

PERSPECTIVE

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

The global landscape of AI ethics guidelines



This paper has been published in IEEE ETHAI 2020

A Survey on Ethical Principles of AI and Implementations

Jianlong Zhou, Fang Chen, Adam Berry
Data Science Institute
University of Technology Sydney
Sydney, Australia
{Jianlong.Zhou, Fang.Chen, Adam.Berry}@uts.ed

***Abstract*—AI has powerful capabilities in predicti
ation, planning, targeting, and personalisation. Gen
assumed that AI can enable machines to exhibit
intelligence, and is claimed to benefit to different a
lives. Since AI is fueled by data and is a distin
autonomous and self-learning agency, we are seeing
ethical concerns related to AI uses. In order to mitig
ethical concerns, national and international or
including governmental organisations, private sector
research institutes have made extensive efforts b
ethical principles of AI, and having active discussion
of AI within and beyond the AI community. T
investigates these efforts with a focus on the ident
fundamental ethical principles of AI and their implem
The review found that there is a convergence arou
principles and the most prevalent principles are tra
justice and fairness, responsibility, non-malefic
privacy. The investigation suggests that ethical prin
to be combined with every stages of the AI lifec
implementation to ensure that the AI system is
implemented and deployed in an ethical manner.
ethical framework used in biomedical and clinical re
paper suggests checklist-style questionnaires as benc
the implementation of ethical principles of AI.**

Keywords*—*AI, ethical principles, implementation

I. Introduction

A. Artificial Intelligence

Artificial Intelligence (AI) is typically defi
interactive, autonomous, self-learning agency with
to perform cognitive functions in contrast to t
intelligence displayed by humans, such as se
moving, reasoning, learning, communicating,
solving (see Figure 1) [1]–[3]. It has powerful cap
prediction, automation, planning, targeti
personalisation, and is claimed to be the driving f
next industrial revolution (Industry 4.0) [4]. It is tra
our world, our life, and our society and affects virt
aspect of our modern lives. Generally, it is assum
can enable machines to exhibit human-like cognitic
more efficient (e.g. higher accuracy, faster, working
than humans in various tasks. Claims about the pro
are abundant and growing related to different ar
lives. Some examples are: in human's everyday li
recognise objects in images, it can transcribe speec
can translate between languages, it can recognise e
images of faces or speech; in traveling, AI makes s
cars possible, AI enables drones to fly autonomou
predict parking difficulty by area in crowded
medicine, AI can discover new uses for existing d
detect a range of conditions from images, it e
personalised medicine; in agriculture, AI can d

OECD publishing

STATE OF IMPLEMENTATION OF THE OECD AI PRINCIPLES

INSIGHTS FROM NATIONAL AI POLICIES

OECD DIGITAL ECONOMY PAPERS

June 2021 No. 311

unesco

Recommendation on the ethics of artificial intelligence

PREAMBLE

The General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), meeting in Paris from 9 to 24 November 2021, at its 41st session,

Recognizing the profound and dynamic positive and negative impacts of artificial intelligence (AI) on societies, environment, ecosystems and human lives, including the human mind, in part because of the new ways in which its use influences human thinking, interaction and decision-making and affects education, human, social and natural sciences, culture, and communication and information,

Recalling that, by the terms of its Constitution, UNESCO seeks to contribute to peace and security by promoting collaboration among nations through education, the sciences, culture, and communication and information, in order to further universal respect for justice, for the rule of law and for the human rights and
damental freedoms which are affirmed for the peoples of the world,

nvinced that the Recommendation presented here, as a standard-setting instrument developed through a
bal approach, based on international law, focusing on human dignity and human rights, as well as gender
ality, social and economic justice and development, physical and mental well-being, diversity,
rconnectedness, inclusiveness, and environmental and ecosystem protection can guide AI technologies in
sponsible direction,

ided by the purposes and principles of the Charter of the United Nations,

nsidering that AI technologies can be of great service to humanity and all countries can benefit from them,
also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate,
entially resulting in discrimination, inequality, digital divides, exclusion and a threat to cultural, social and
logical diversity and social or economic divides; the need for transparency and understandability of the
rkings of algorithms and the data with which they have been trained; and their potential impact on, including
not limited to, human dignity, human rights and fundamental freedoms, gender equality, democracy, social,
nomic, political and cultural processes, scientific and engineering practices, animal welfare, and the
ironment and ecosystems,

o recognizing that AI technologies can deepen existing divides and inequalities in the world, within and
ween countries, and that justice, trust and fairness must be upheld so that no country and no one should
left behind, either by having fair access to AI technologies and enjoying their benefits or in the protection
inst their negative implications, while recognizing the different circumstances of different countries and
pecting the desire of some people not to take part in all technological developments,

nscious of the fact that all countries are facing an acceleration in the use of information and communication
hnologies and AI technologies, as well as an increasing need for media and information literacy, and that
digital economy presents important societal, economic and environmental challenges and opportunities of
efit-sharing, especially for low- and middle-income countries (LMICs), including but not limited to least
eloped countries (LDCs), landlocked developing countries (LLDCs) and small island developing States
DS), requiring the recognition, protection and promotion of endogenous cultures, values and knowledge in
er to develop sustainable digital economies,

rther recognizing that AI technologies have the potential to be beneficial to the environment and
systems, and in order for those benefits to be realized, potential harms to and negative impacts on the
ironment and ecosystems should not be ignored but instead addressed,

ting that addressing risks and ethical concerns should not hamper innovation and development but rather
vide new opportunities and stimulate ethically-conducted research and innovation that anchor AI
hnologies in human rights and fundamental freedoms, values and principles, and moral and ethical
ection,

SHS/BIO/REC-AIETHICS/2021

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)

European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment

Adopted at the 31th plenary meeting
of the CEPEJ on 3-4 December 2018

cepej

European Commission for the Efficiency of Justice

Commission européenne pour l'efficacité de la justice

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

TALOS-expectations

- To raise awareness about research ethics in the field of Humanities and Social Sciences
- To identify and highlight the ethical challenges that arise in relation to the use of AI in the field of SSH
- To propose sustainable and innovative ways and tools to address novel ethical challenges emerging in the field of digital humanities.
- To prepare the new generation of researchers who will support the responsible design of artificial intelligence and develop interdisciplinary solutions for the 'humanization' of artificial intelligence.
- To make the University of Crete an international point of reference on issue of responsible innovation.

European Commission