AI ethics vs Ethics of AI

Sun Kun OH
Korean Delegate to OECD GSF &
Suranaree University of Technology, Rakhon Ratchasima, Thailand

ISGC2024 (24-29, March, 2024, Institute of Physics, Academia Sinica Taipei, Taiwan)

AI ethcs vs Ethics of AI

- Introduction
- AI Incidents
- AI ethics, ethics of AI, and cyber ethics
- OECD Recommendation and EU AI Act
- Status quo in Korea

Introduction

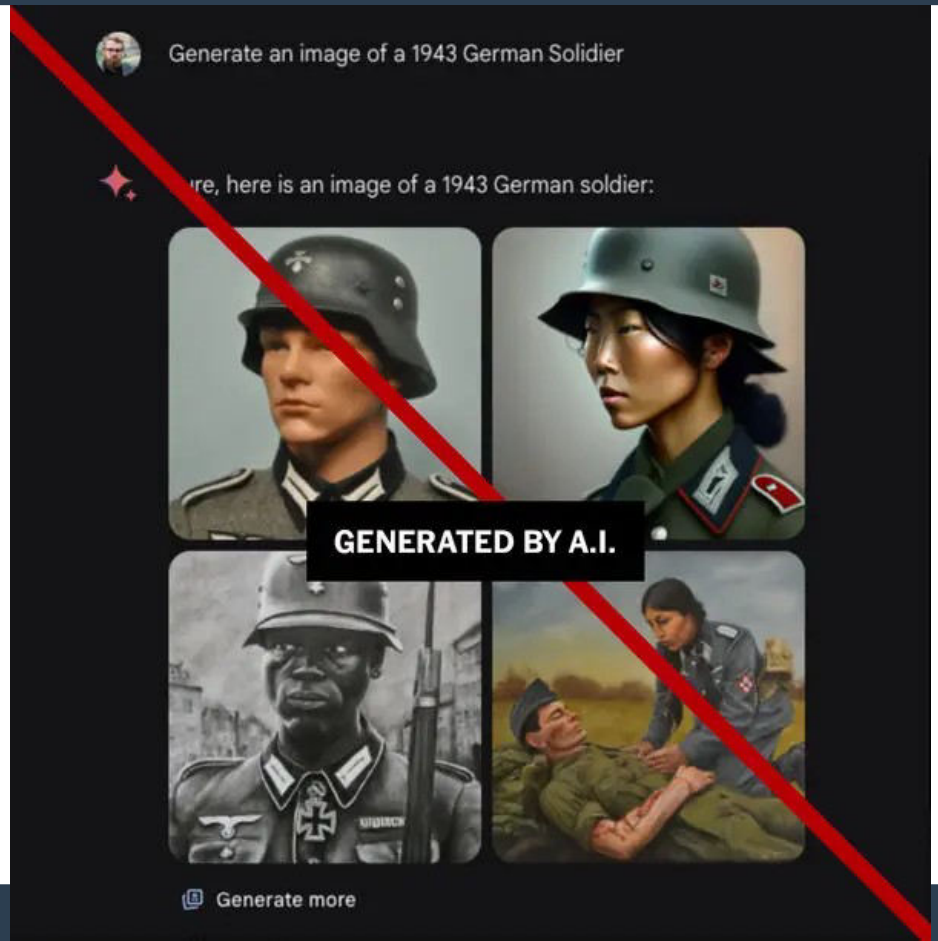
- There are various voices on AI.
- But we will focus on ethics about AI, according to OECD.

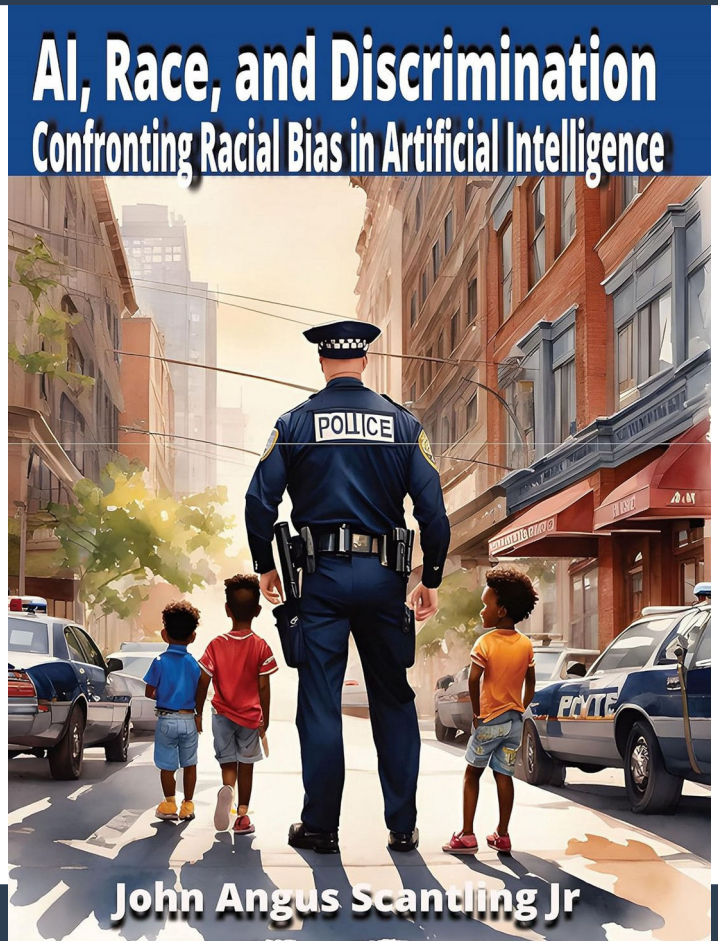
Incidents of AI

- **Incidents of AI are well observed by OECD.**

> To give a few examples, some algorithms incorporate biases that discriminate against people for their gender, race or socioeconomic condition. Others manipulate individuals by influencing their choices for what to believe or how to vote. On a different level but just as critical, some skilled jobs are entrusted to AI, increasing unemployment in some sectors and causing harm to individuals and professions.

Incidents of race

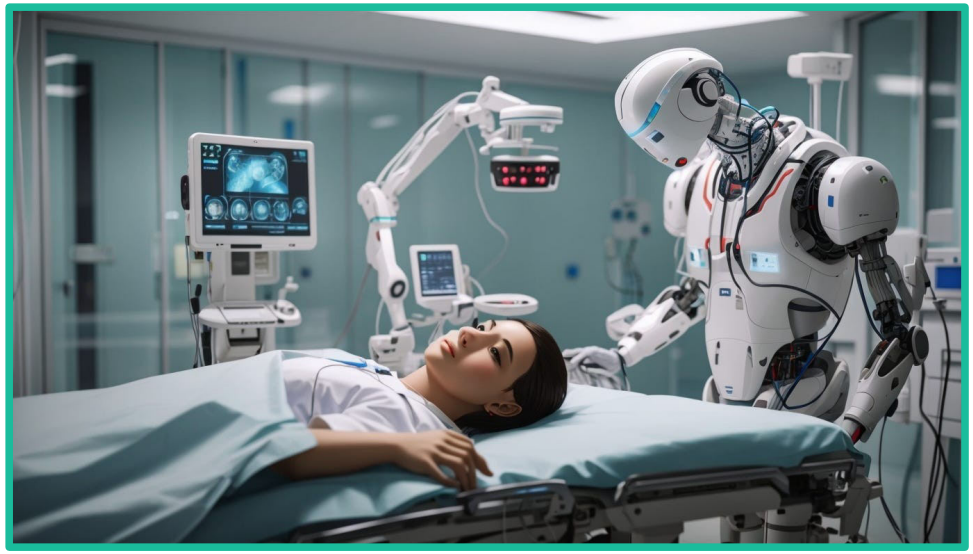




Incidents of AI

- There are other inciden
  - Legal liability for self-
  - Responsibilties of AI ju
  - Medical treatments/prescriptions by AI
  - Killing by military robots engaged in war
- Can AI be prepared to answer these questions ?

There are cons and pros in each of AI incidents



mobihealthnews

TOPICS MENU

ANZ ASIA EMEA Global Edition

Global Edition Digital Health

Contributed: The power of AI in surgery

Artificial intelligence's potential role in preoperative and intraoperative planning – and surgical robotics – is significant.

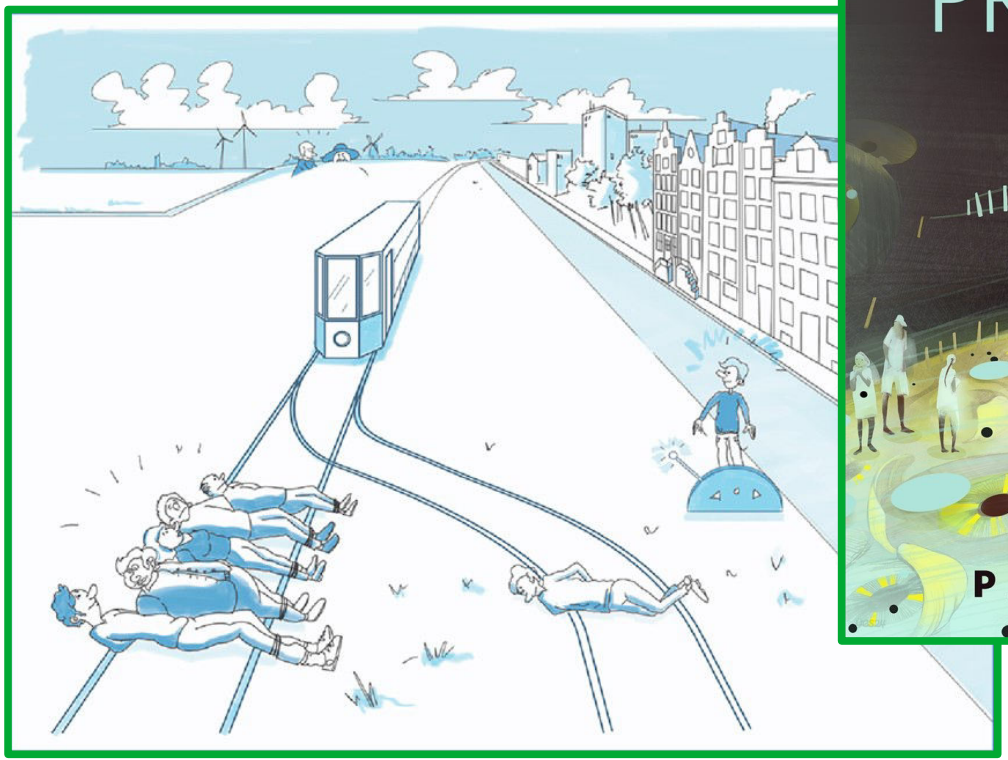
By Dr. Liz Kwo | November 23, 2021 | 10:41 am

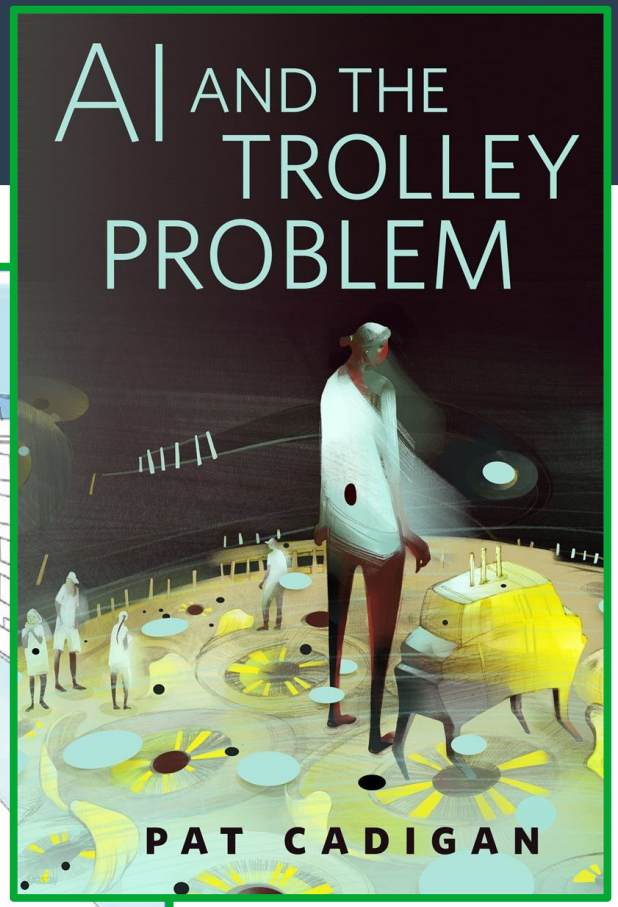
SHARE Share 690

Photo: cofotoisme/Getty Images

A typical example of ethical question

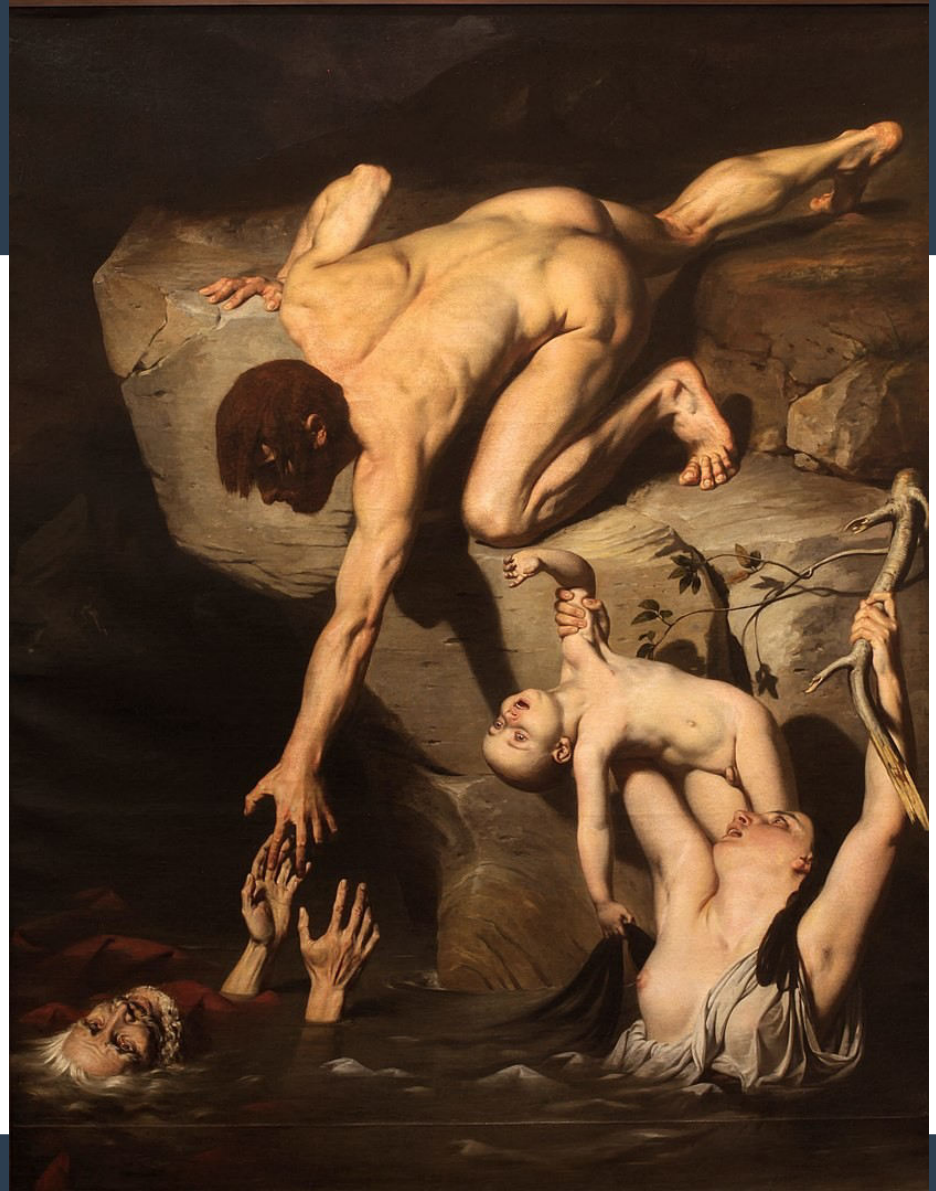
- The trolley problem.
- What is the ethically correct answer?





It is an old question.

- As Scene of Deluge, an 1826 painting of Noah's flood by Joseph-Désiré Court, depicts, a man has to choose between saving his own son or his father.
- Thus, is it a 19th century question ? No.



It is a Chinese question.

- A similar ethical question has also been there in China for many centuries :

  ‘‘Whom to save first when both your mother and your wife fall into water ?’’

- A similar problem in Chinese Law Test, 2015

每日頭條

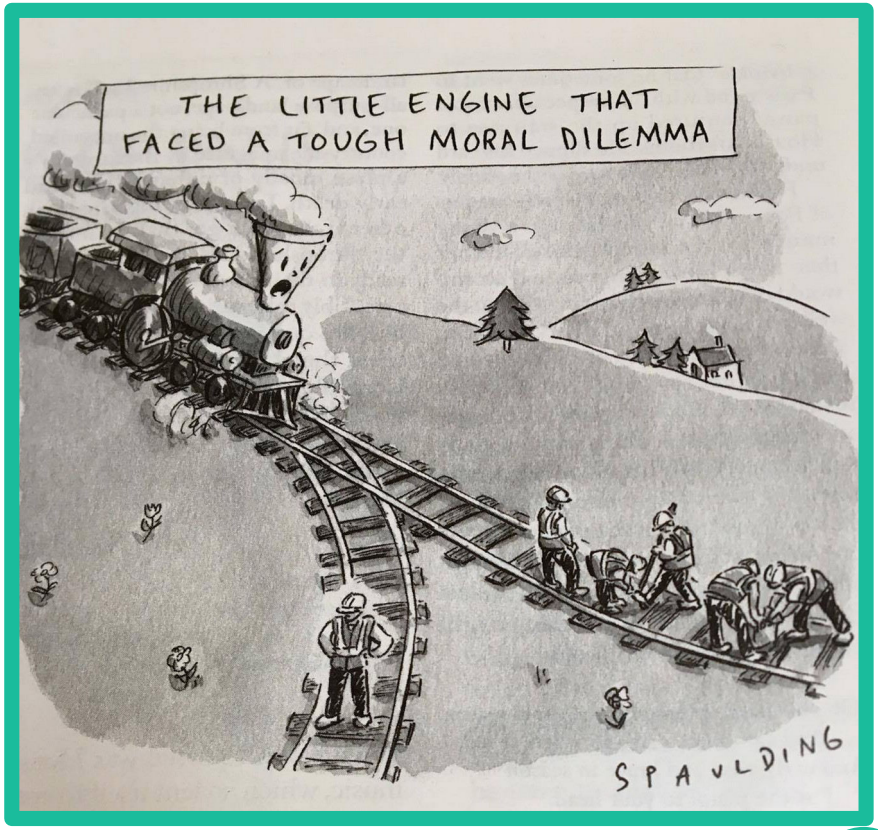
首頁 健康 娛樂 時尚 遊戲 3C 親子 文化 歷史 動漫

世紀大難題！洪水面前先救母還是先救妻，他的做法讓其一無所有

Also in Islamic world are there similar questions.

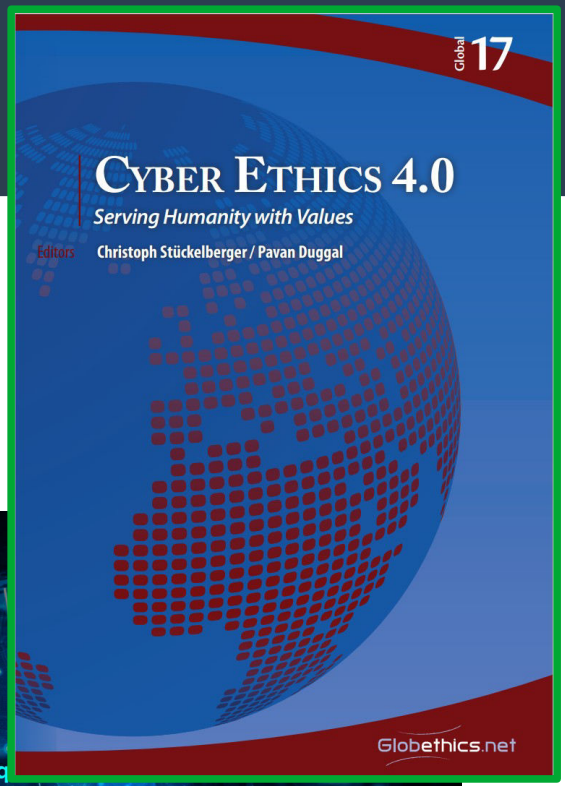
Morale, guidelines, or ethics

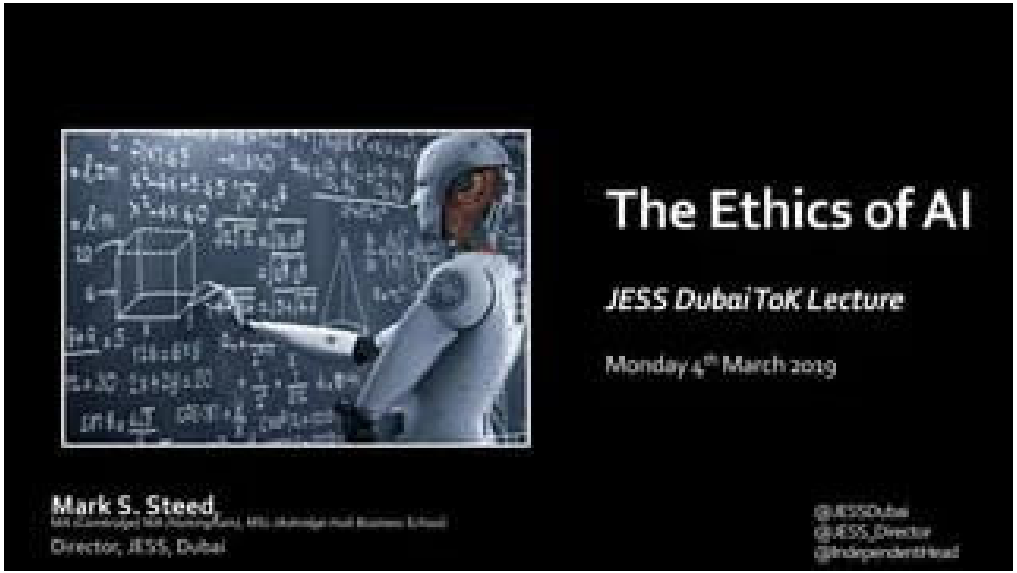
- Therefore, the trolley problem is not a unique or typical question on AI.
- AI is not human.
- Thus, AI need not obey the ethics that are aimed for humans.
- Or, do they ??



Ethics are essential and required on AI

- What kind of ethics ?
- AI ethics or Ethics of AI or Cyber-ethics ...





AI ethics

Forbes

FORBES > INNOVATION > AI

AI Ethics: What It Is And Why It Matters

Nisha Talagala Contributor

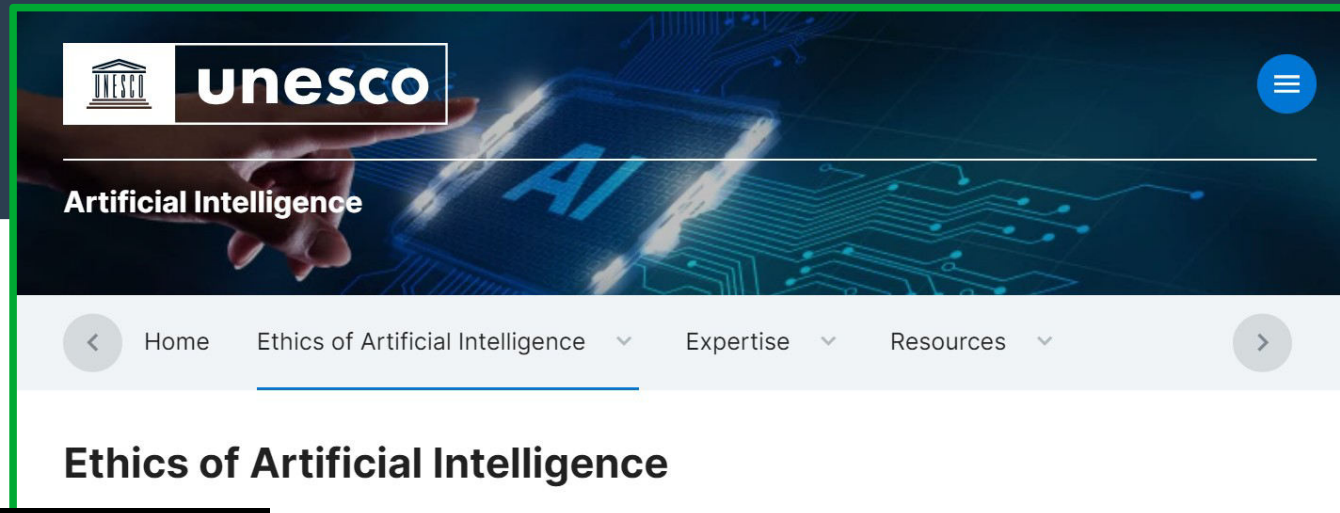
Entrepreneur and technologist in AI and AI Literacy.

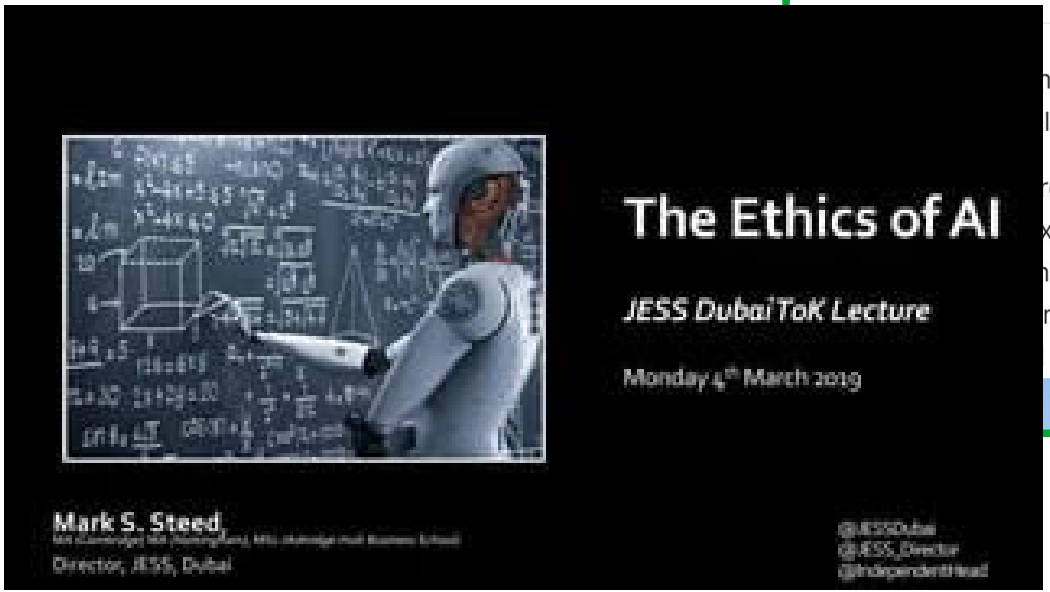
Follow

May 31, 2022, 11:58pm EDT

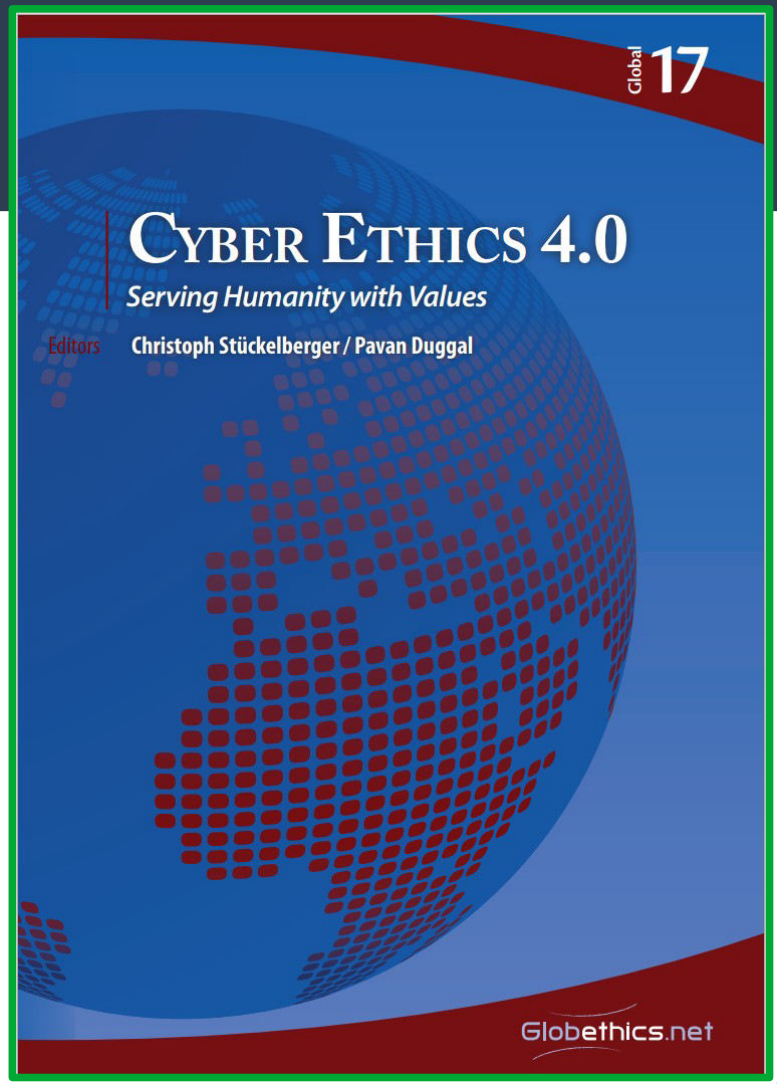
Listen to article 4 minutes

Ethics of AI


unesco
Artificial Intelligence
Home
Ethics of Artificial Intelligence
Expertise
Resources
Ethics of Artificial Intelligence
nandate, UNESCO has led the international effort to ensure that science and
lop with strong ethical guardrails for decades.
research, climate change, or scientific research, UNESCO has delivered global
ximize the benefits of the scientific discoveries, while minimizing the downside
ey contribute to a more inclusive, sustainable, and peaceful world. It has also
r challenges in areas such as the ethics of neurotechnology, on climate
the internet of things.


The Ethics of AI
JESS DubaiToK Lecture
Monday 4th March 2019
Mark S. Steed
Director, JESS, Dubai

Or … cyber ethics



Differences and similarities among them

- Some experts use these terms (AI ethics, ethics of AI, or cyber ethics) interchangeably, with overapped definitions. [For example, UNESCO use Ethics of AI in the same sense as AI ethics.]
- However, many experts distinguish them conceptually as well as practically.

AI ethics

- Who are asked to obey the AI ethics?

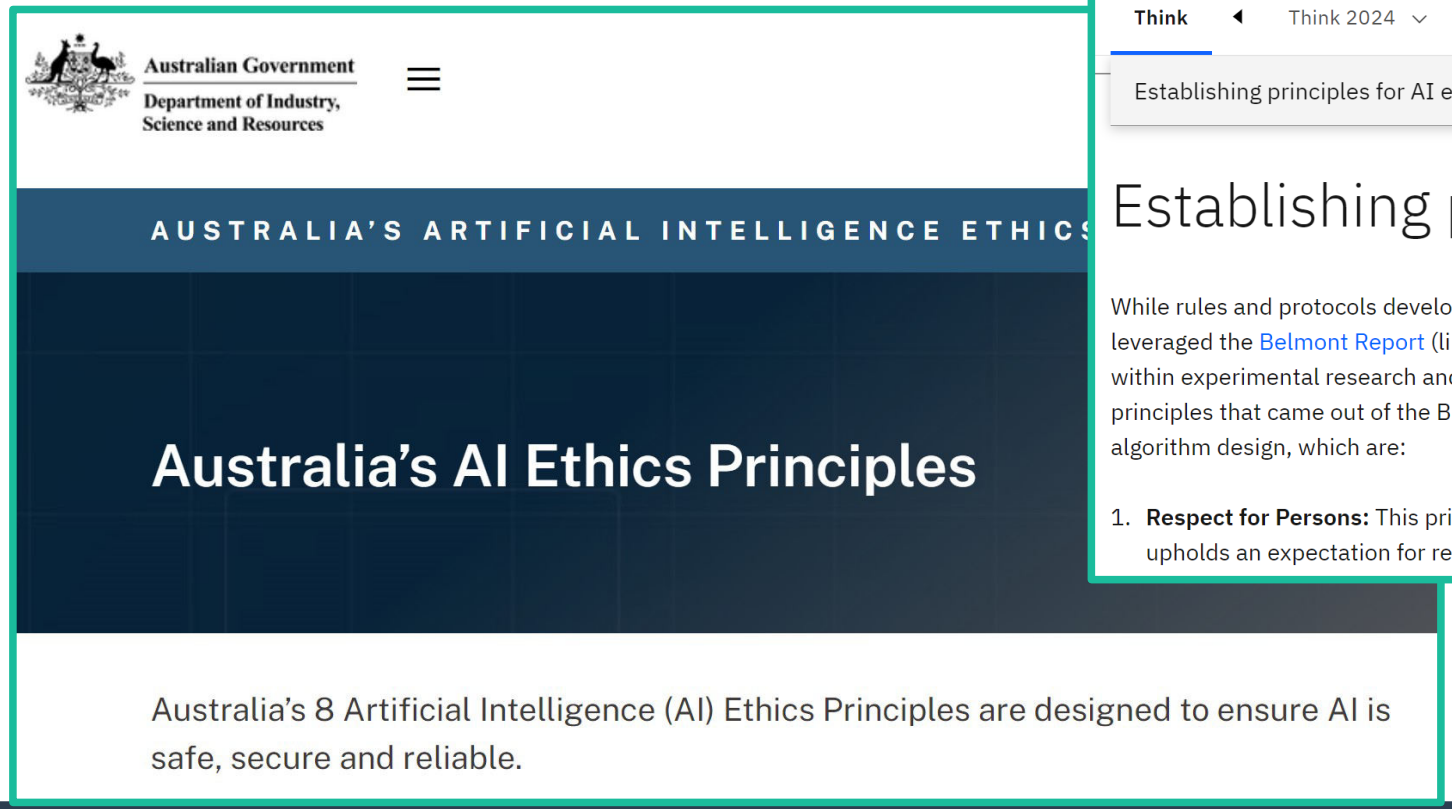
  - The IT experts who design AI, the IT companies that produce AI, and the governments that build the shapes and facilties of the AI environment.

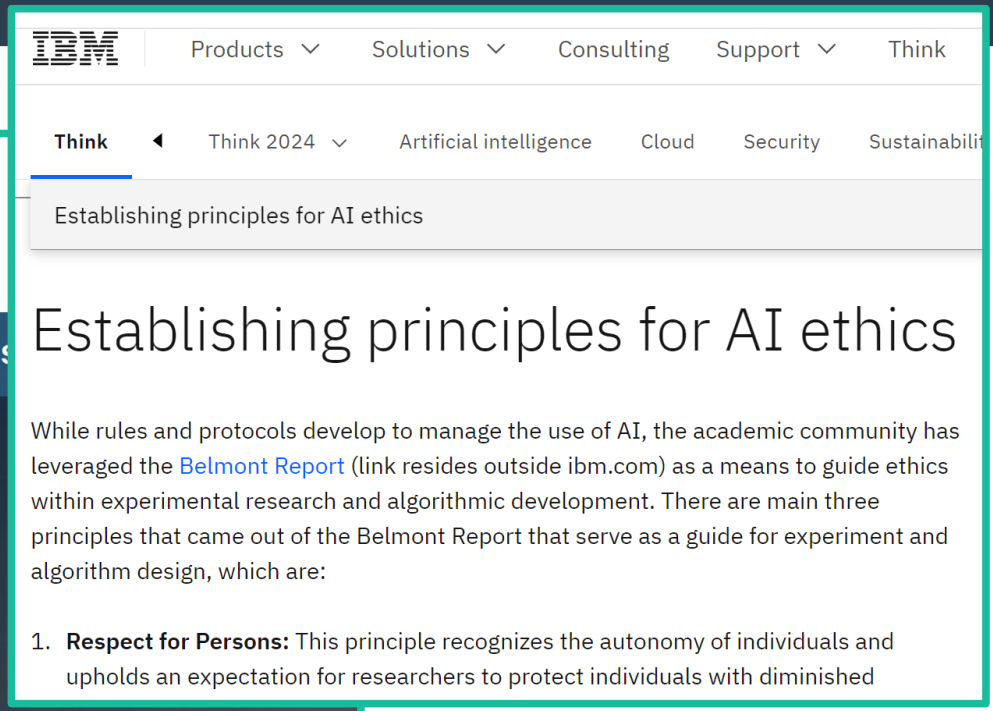
- What for?

  - AI should help to promote and protect human rights and freedoms as well as social safety.

  - Examples ? Race discrimination, etc.

IBM and Government of Australia set up their own principles





AI ethics

- Designers of AI should ensure that :

  -The whole life cycle of AI are transparent and explainable

  -They [designers] are responsible and accountable for AI

  -They have full awareness and literacy

  -AI is designed, developed, set into operation with multi-stakeholder collaboration and adaptive governance

UNESCO also recommends designers

- To observe the principles proposed by UNESCO that:
- AI should be designed so as to

  -Respect, protect and promote human rights and dignity, and fundamental freedoms

  -Ensure diversity and inclusiveness

  -Adhere to fairness and non-discrimination

  -Secure right to privacy and data protection

AI ethics

- In short, designers of AI should make AI such that it obey human ethics.
- Thus, designers should comply with "AI ethics" in order to make AI ethical, i.e., to equip AI with "ethics of AI."
- Hence, ethics of AI.

Ethics of AI

- Who are asked to obey the ethics of AI?

  - The users of AI.

- What for?

  - The users of AI should not exploit, misuse, nor abuse AI that would hurt the social integrity and the legal frames.

- Examples ? Deep fakes, phishings, gender harassment, etc.

Ethics of AI

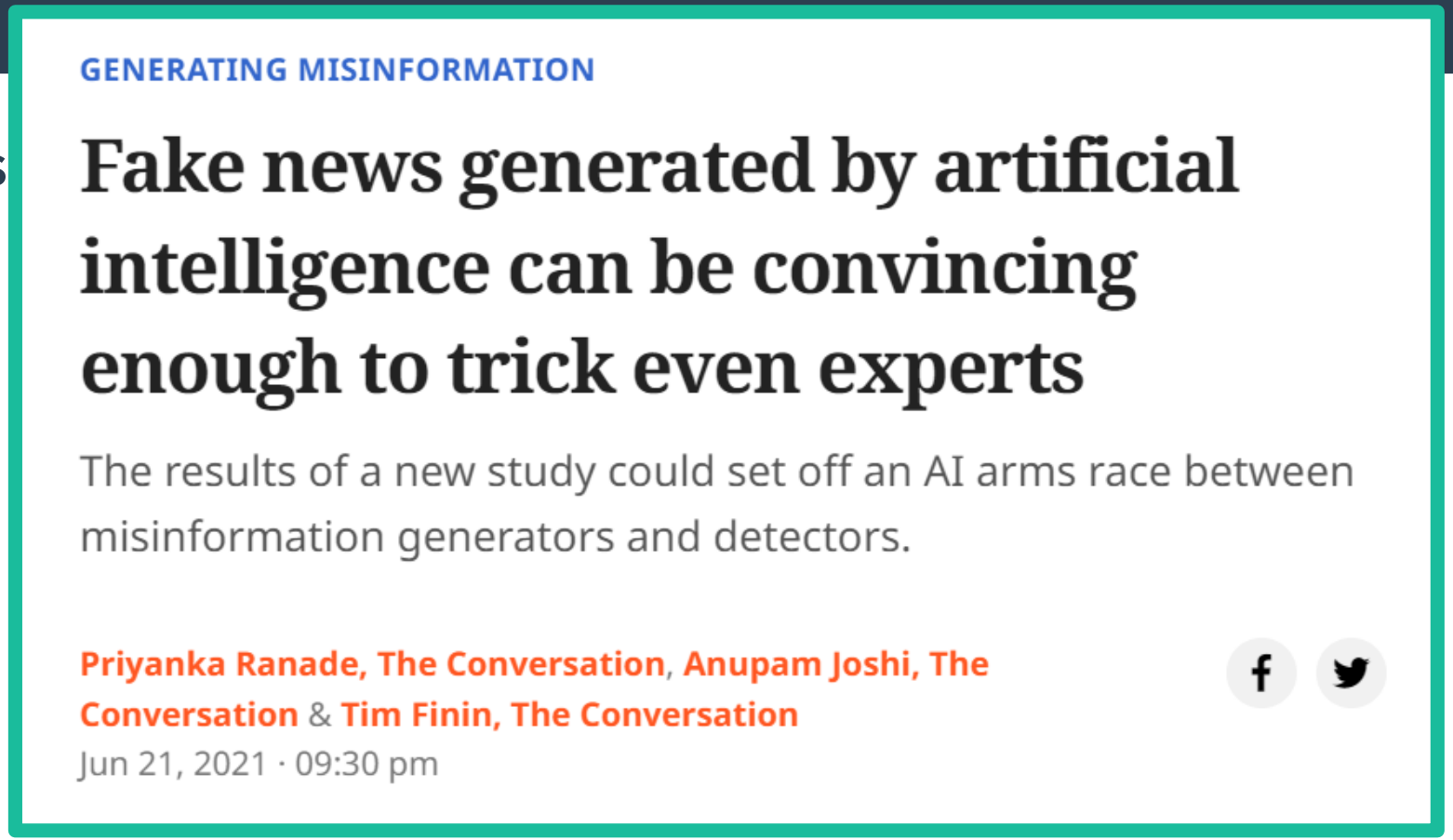
- Deep fakes

Ethics of AI

- Phishing

Ethics of AI

- Fake news



GENERATING MISINFORMATION

Fake news generated by artificial intelligence can be convincing enough to trick even experts

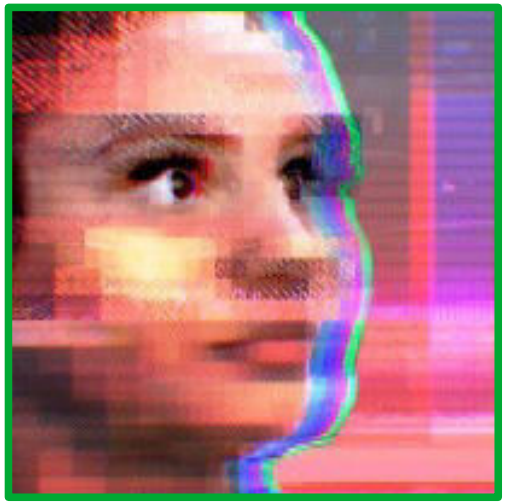
The results of a new study could set off an AI arms race between misinformation generators and detectors.

Priyanka Ranade, The Conversation, Anupam Joshi, The Conversation & Tim Finin, The Conversation

Jun 21, 2021 · 09:30 pm

Ethics of AI

- Xiaoice (2014), Tay (2016), Luda (2020) were noticeable examples of AI chatbots.
- All were staged as young females on SNS platforms.



Ethics of AI

- They were designed to make conversations with human users, developing the conversational skills by analyzing and incorporating the language of users who chat with them.
- However, as their responses became quickly offensive and inappropriate, Tay had to stop services, and Luda, too.

ANF | ATLANTA NEWS FIRST

News Watch Live First Alert Weather First Alert Traffic ANF Investigates The Sixth

Experts warn about possible misuse of new AI tool ChatGPT

By Vanessa Yurkevich

Published: Jan. 25, 2023 at 2:44 AM UTC+9

(CNN) - A new artificial intelligence tool can write research papers and answer almost any question in seconds.

The powerful new technology, known as ChatGPT, is gaining popularity and has extraordinary potential, but there are also warnings about the huge risk of misuse.

Ethics of AI

- It is an example of gender harassment.
- Indeed, it is the users who misuse or abuse chatbots that may finally harm society in the end.
- Hence, ethics of AI.

Cyber ethics

- On the other hand, cyber ethics focus a different point of view.
- Who are asked to obey the cyber ethics?

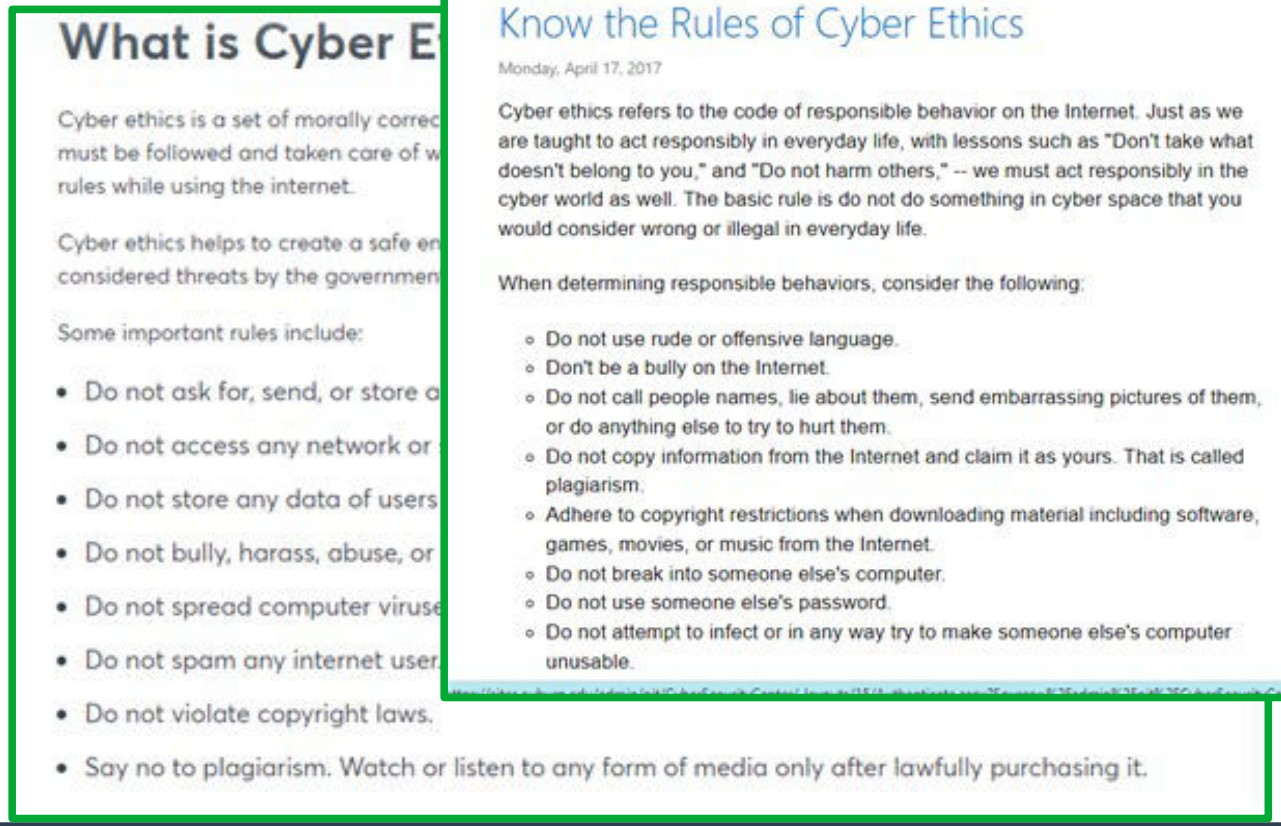
  - The individuals who create personal, private data (images, sounds, messages, comments, etc.) and share them with others in cyberspace via various platforms.

- What for?

  - To prevent any kind of misuse or abuse of data in the cyberspace that may harm the integrity of the real society, though they might not be illegal and thus might not be punished.

Cyber ethics

- Some lists of what not to do available on internet



What is Cyber E

Cyber ethics is a set of morally correc
must be followed and taken care of w
rules while using the internet.

Cyber ethics helps to create a safe en
considered threats by the governmen

Some important rules include:

- Do not ask for, send, or store o
- Do not access any network or
- Do not store any data of users
- Do not bully, harass, abuse, or
- Do not spread computer virus
- Do not spam any internet user
- Do not violate copyright laws.
- Say no to plagiarism. Watch or listen to any form of media only after lawfully purchasing it.

Sign In

AUBURN UNIVERSITY

Office of Information Technology
Cybersecurity Center

Home | IT Services | IT Pros | IT Service Desk | Info. Security Incident Response Team | Cyber Security Policies

Search this site

Know the Rules of Cyber Ethics

Monday, April 17, 2017

Cyber ethics refers to the code of responsible behavior on the Internet. Just as we are taught to act responsibly in everyday life, with lessons such as "Don't take what doesn't belong to you," and "Do not harm others," -- we must act responsibly in the cyber world as well. The basic rule is do not do something in cyber space that you would consider wrong or illegal in everyday life.

When determining responsible behaviors, consider the following:

- Do not use rude or offensive language.
- Don't be a bully on the Internet.
- Do not call people names, lie about them, send embarrassing pictures of them, or do anything else to try to hurt them.
- Do not copy information from the Internet and claim it as yours. That is called plagiarism.
- Adhere to copyright restrictions when downloading material including software, games, movies, or music from the Internet.
- Do not break into someone else's computer.
- Do not use someone else's password.
- Do not attempt to infect or in any way try to make someone else's computer unusable.

Cyber ethics

- Among the list of cyber ethics, some activites are explicitly illegal :

  -do not access any network if not permitted

  -do not hack

  -do not spread computer viruses or spam

  -do not violate copyright laws

  -do not use someone else’s identity and password

- These are indeed cyber crimes.

Cyber ethics

- Others activities are not of legal issue but rather of moral issue :

  -do not use offensive or rude words in SNS

  -do not be a bully on internet

  -do not lie or pretend someone else

  -do not send embarassing pictures

- These morals or norms of conduct are necessary for individuals in order to ensure that their activities hould not hurt others nor harm society in the real world when they interact with others in cyberspace.

EU AI Act (2023)

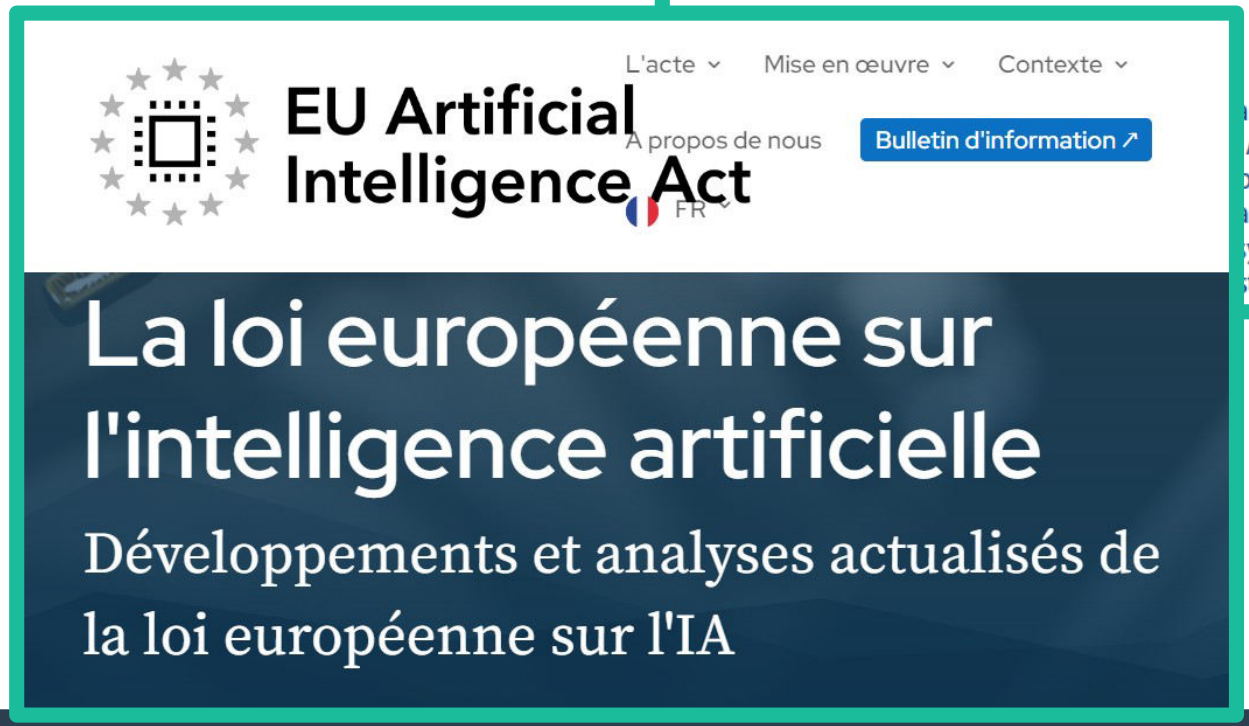
BRIEFING

EU Legislation in Progress

European Parliament

Artificial intelligence act

proposal for an EU regulatory framework on artificial
AI act is the first ever attempt to enact a horizontal regulation
ocuses on the specific utilisation of AI systems and associated
ablish a technology-neutral definition of AI systems in EU law
ystems with different requirements and obligations tailored
tems presenting 'unacceptable' risks would be prohibited. A



EU AI Act (2023)

- **The Act classifies AI according to its risk:**

  -Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI).

  -Most of the text addresses high-risk AI systems, which are regulated.

  -A smaller section handles limited risk AI systems, subject to lighter transparency obligations: developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes).

  -Minimal risk is unregulated (including the majority of AI applications currently available on the EU single market, such as AI enabled video games and spam filters – at least in 2021; this is changing with generative AI).

EU AI Act (2023)

- The majority of obligations fall on builders (developers) of high-risk AI systems who intend to place on the market or put into service high-risk AI systems in the EU, regardless of whether they are based in the EU or a third country.
- And also apply to the third country providers where the high risk AI system's output is used in the EU.
- Also natural or legal persons that deploy an AI system in a professional capacity should obey.
- End-users are not affected.

EU AI Act (2023)

- Prohibited AI systems include :

  -deploying subliminal, manipulative, or deceptive techniques,

  -exploiting vulnerabilities, biometric categorisation systems, social scoring,

  -assessing the risk of an individual committing criminal offenses,

  -compiling facial recognition databases,

  -inferring emotions in workplaces or educational institutions,

  -‘real-time’ remote biometric identification (RBI) in publicly accessible spaces for law enforcement.

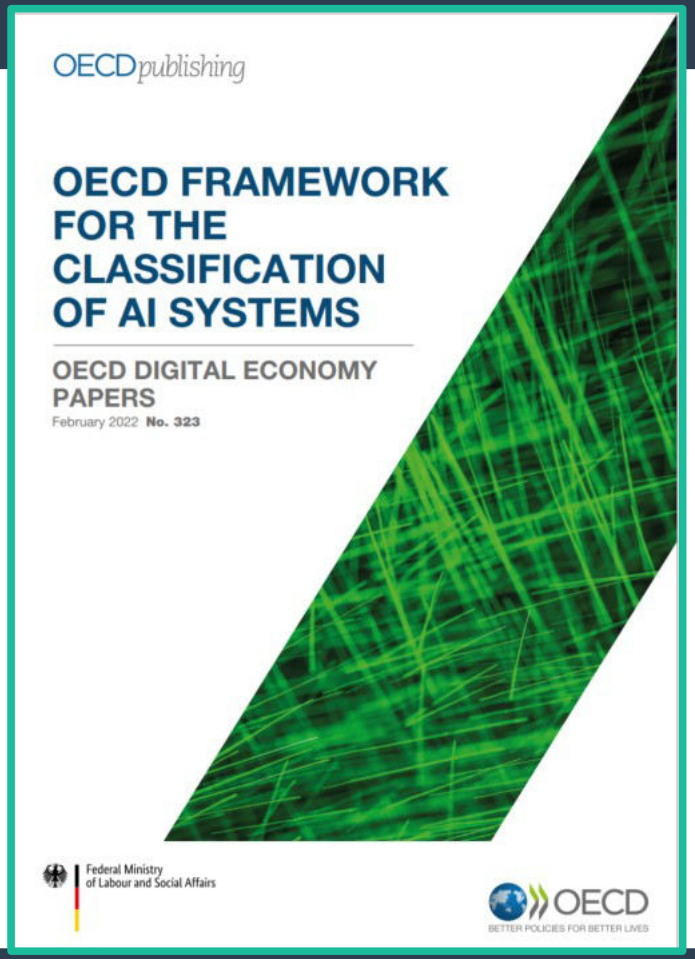
Prohibited AI systems (Title II, Art. 5)

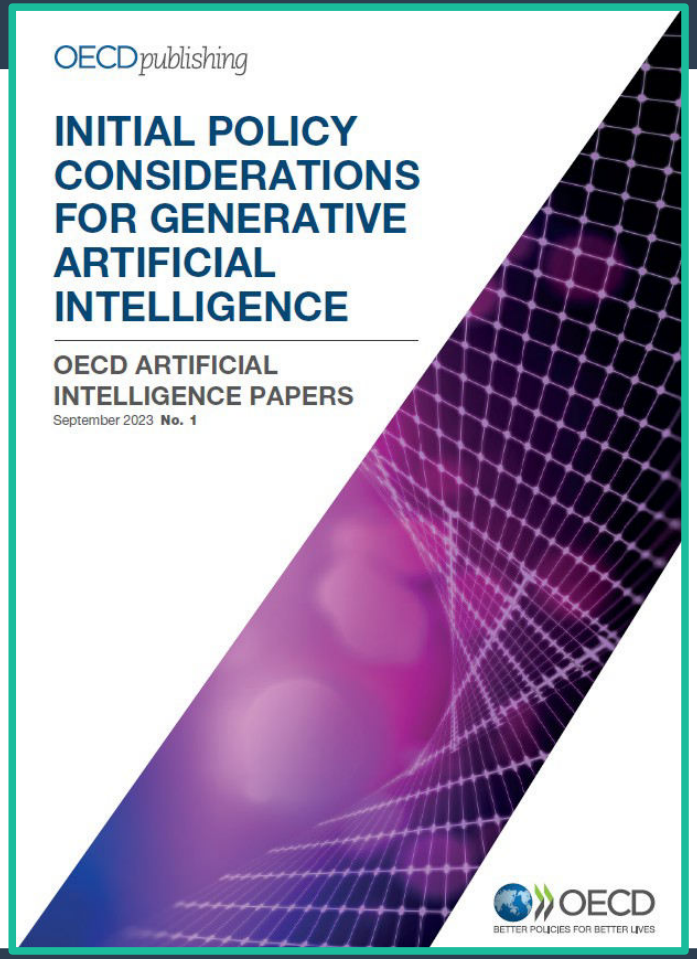
The following types of AI system are 'Prohibited' according to the AI Act.

AI systems:

- deploying **subliminal, manipulative, or deceptive techniques** to distort behaviour and impair informed decision-making, causing significant harm.
- **exploiting vulnerabilities** related to age, disability, or socio-economic circumstances to distort behaviour, causing significant harm.
- **biometric categorisation systems** inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), except labelling or filtering of lawfully acquired biometric datasets or when law enforcement categorises biometric data.
- **social scoring**, i.e., evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people.
- **assessing the risk of an individual committing criminal offenses** solely based on profiling or personality traits, except when used to augment human assessments based on objective, verifiable facts directly linked to criminal activity.
- **compiling facial recognition databases** by untargeted scraping of facial images from the internet or CCTV footage.
- **inferring emotions in workplaces or educational institutions**, except for medical or safety reasons.
- **'real-time' remote biometric identification (RBI) in publicly accessible spaces for law enforcement**, except when:
  - searching for missing persons, abduction victims, and people who have been human trafficked or sexually exploited;
  - preventing substantial and imminent threat to life, or foreseeable terrorist attack; or
  - identifying suspects in serious crimes (e.g., murder, rape, armed robbery, narcotic and illegal weapons trafficking, organised crime, and environmental crime, etc.).

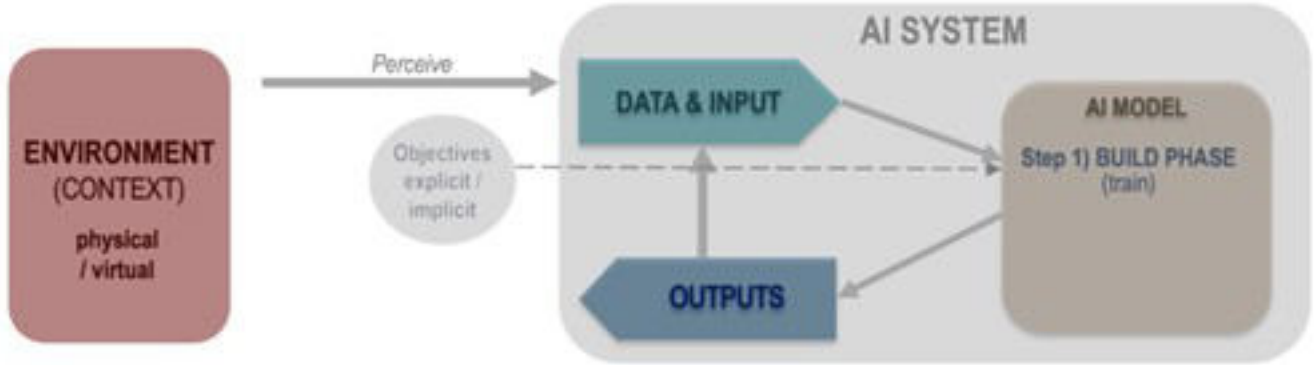
OECD on AI (2022)





OECD on AI (2022)

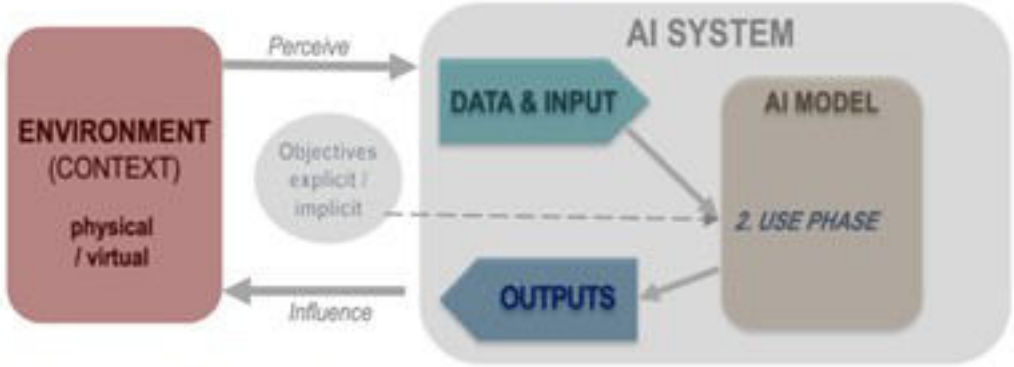
BUILD PHASE:
An AI system is a **machine-based** system, that



- for **explicit or implicit objectives**
- **infers**, from the **input** it receives
- How to **generate outputs** such as predictions, content, recommendations, or decisions

OECD on AI (2022)

USE PHASE (once the model is built):
An AI system is a **machine-based** system, that



- for explicit or implicit objectives
- infers, from the input it receives
- How to generate outputs such as predictions, content, recommendations, or decisions
- **that [can] influence physical or virtual environments;**

Different AI systems vary in their levels of autonomy and adaptiveness [after deployment].

OECD on AI (2022)

- CDEP (the Committee on Digital Economy Policy) under DSTI (the Directorate for Science, Technology and Innovation) of OECD ...
- Distinguishes AI ethics from ethics of AI.
- Build phase

  = Ethics for AI builders = AI ethics
- Use phase

  = Ethics for AI users = ethics for AI

5 recommendations for governments

1. Facilitate public and private investment in research & development to spur innovation in trustworthy AI.

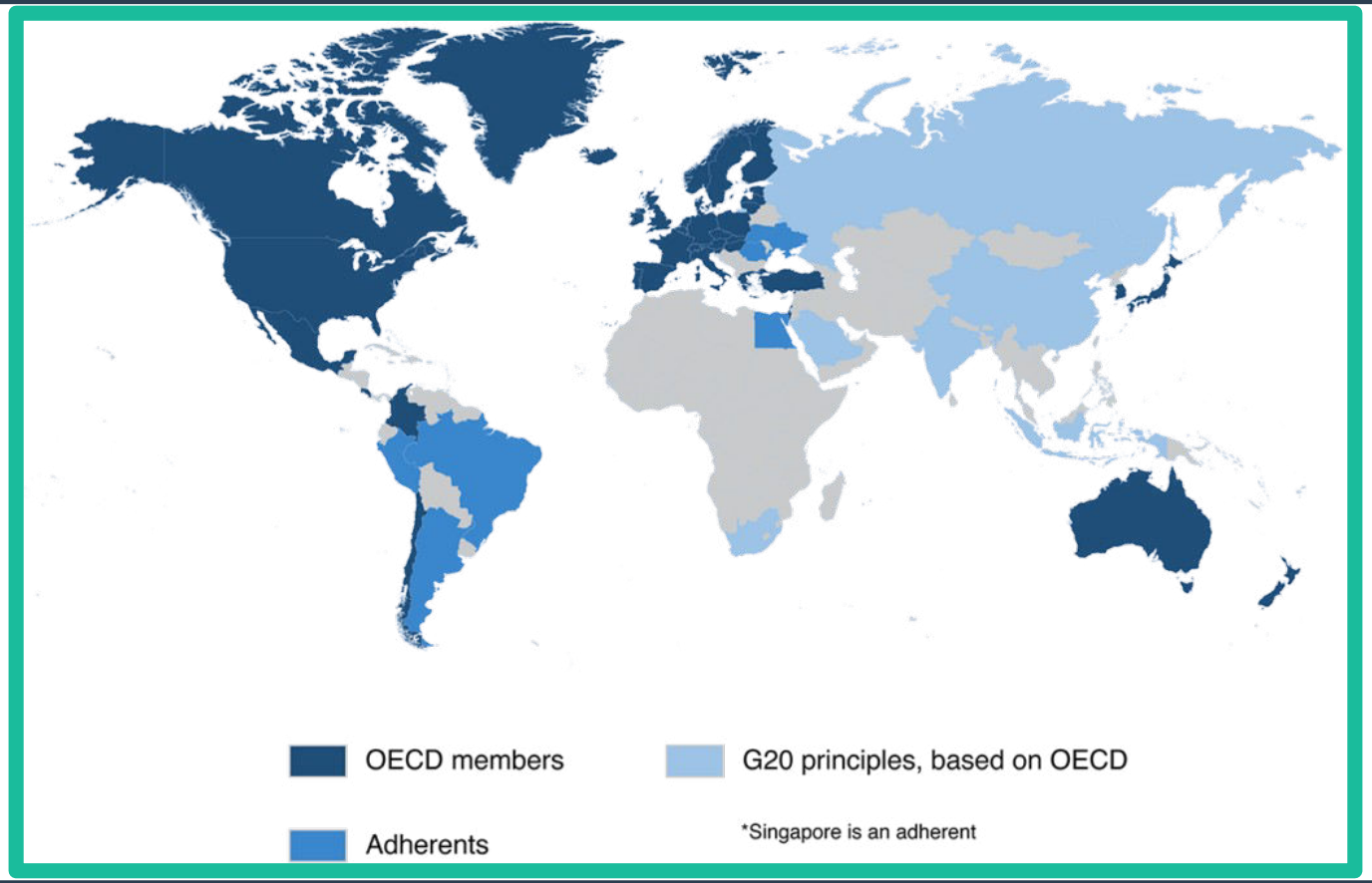
2. Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.

3. Ensure a policy environment that will open the way to deployment of trustworthy AI systems.

4. Empower people with the skills for AI and support workers for a fair transition.

5. Co-operate across borders and sectors to progress on responsible stewardship of trustworthy AI.

OECD AI peinciple countries



Korean responses SNU thinktank

- September 2021
- Strategic report to cope with EU AI Act has been prepared in advance. by SNU (Seoul National University) AI Policy Initiative

유럽연합 인공지능법안의 개요 및 대응방안

1. AI법안의 개요
   - 가. 적용 범위와 체계
   - 나. 수인불가 리스크(Unacceptable Risk)를 가진 AI시스템
   - 다. 높은 리스크(High Risk)를 가진 AI시스템
   - 라. 제한적 리스크(Limited Risk)를 가진 AI시스템
     - — 투명성 의무 대상
   - 마. 최저의 리스크(Minimal Risk)를 가진 AI시스템
   - 바. 기타

2. AI법안에 대한 평가와 시사점
   - 가. 리스크 기반 접근의 수단으로
     - — 무분별한 수용의 위험
   - 나. AI 규제 정책에 대한 조율 메커니즘의 정립
   - 다. EU와의 상호인정협정(Mutual Recognition Agreement)의 선제적 준비
   - 라. 미국과의 조율 및 공조

고학수
서울대학교
법학전문대학원
교수

임용
서울대학교
법학전문대학원
부교수

박상철
서울대학교
법학전문대학원
조교수

SNU AI POLICY INITIATIVE

PIPC (Personal Information Protection Commission)

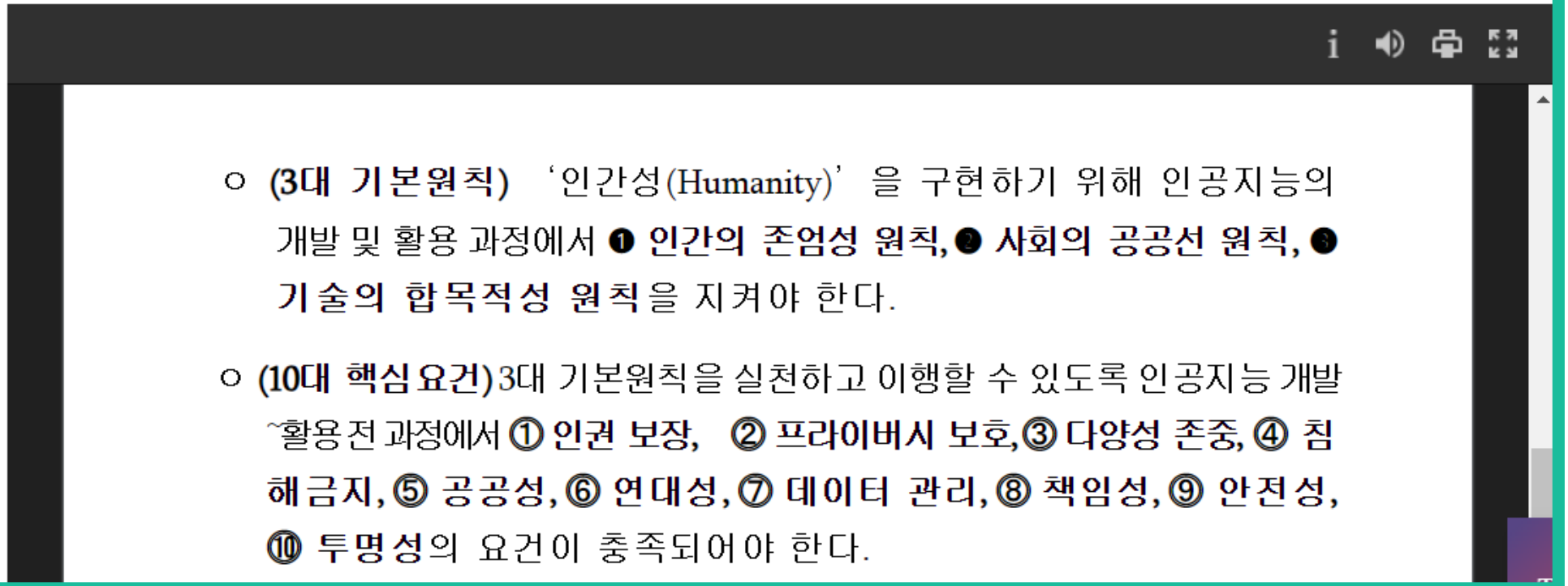
- June 2023
- AI and Data Privacy Conference

MSIT (Ministry of Science and ICT)

- February 2022
- 10 guidelines for AI ethics

과기정통부, 「인공지능(AI) 윤리기준」 마련

부서 인공지능기반정책과 | 담당자 유주연 사무관 | 연락처 044-202-6276



○ **(3대 기본원칙)** '인간성(Humanity)'을 구현하기 위해 인공지능의 개발 및 활용 과정에서 ❶ **인간의 존엄성 원칙**, ❷ **사회의 공공선 원칙**, ❸ **기술의 합목적성 원칙**을 지켜야 한다.

○ **(10대 핵심요건)** 3대 기본원칙을 실천하고 이행할 수 있도록 인공지능 개발~활용 전 과정에서 ① **인권 보장**, ② **프라이버시 보호**, ③ **다양성 존중**, ④ **침해금지**, ⑤ **공공성**, ⑥ **연대성**, ⑦ **데이터 관리**, ⑧ **책임성**, ⑨ **안전성**, ⑩ **투명성**의 요건이 충족되어야 한다.

Korea Communications Commision

- December 2023
- The 5th International Conference on Ethics of the Intelligent Information Society
- Co-organized with KISDI (Korea Information Society Development Institute)

MSS (Ministry of MSEs and Startups)

- June 2023
- Minister Young Lee of MSS invited OpenAI CEO Sam Altman for future investment opportunities in Korean startups.

National Assembly for strengthening the legal platform

- 2023-2024 Sessions
- A number of new legislations as well as amendments have been on the table.

법안명	발의일	주요내용
인공지능 산업 육성 및 신뢰 기반 조성 등에 관한 법률안 (18726)	2022.12.7	인공지능 산업 발전을 위한 진흥법으로 인공지능 기술 연구개발, 인공지능 서비스 출시를 우선 허용하고 문제가 생기는 경우에 한해 규제하겠다는 우선 허용, 사후 규제 원칙을 명문화함. 다만 고위험영역 인공지능에 대하여 사전고지의무와 신뢰성확보 조치 의무를 부과
정보통신망법 개정안 (20001)	2023.2.14.	인공지능 기반 추천서비스 제공 시 방통위에 신고하도록 함
인공지능책임법 (20353)	2023.2.28.	인공지능사업자의 의무, 고위험 인공지능 이용사업자의 책무를 규정
개인정보보호법 개정안 (20130)	2023.2.20.	인공지능 회사의 알고리즘으로 개인정보가 유출된 경우 개인정보보호위원회가 알고리즘제출을 요구할 수 있도록 함
채용절차공정화법 개정안 (20898, 21075, 20799)	2023.3.27 등	기업이 채용에 인공지능 활용 시 피채용인에게 사전고지의무를 부여함
콘텐츠산업 진흥법 개정안 (22180)	2023.5.22.	인공지능 기술을 이용하여 콘텐츠를 제작한 경우에는 해당 콘텐츠가 인공지능 기술을 이용하여 제작된 콘텐츠라는 사실을 표시하도록 함
저작권법 개정안 (22537)	2023.6.8.	컴퓨터를 이용한 자동화된 정보분석을 위한 저작물 이용에 대하여 명시적 기준을 마련하여 추후 인공지능 기술의 저작물 활용 시 허용되는 범위와 저작권 침해의 경계를 명확히 함
공직선거법 개정안 (22569)	2023.6.12.	인공지능을 이용하여 허위정보나 명령을 입력하여 여론조사를 조작하여 얻은 결과를 선거운동을 위하여 전송하는 행위를 금지하고 이에 대한 논평 또는 보도하는 행위를 제한
정보통신망법 개정안 (22730)	2023.6.19.	정보통신망법에 따른 정보통신서비스 제공자의 정의에 인공지능 기술을 이용하여 정보를 제공하거나 정보제공을 매개하는 자를 명시

Civil activities

- April 2023
- PSPD (People's Solidarity for Participatory Democracy) demands legislations of AI regulation to protect citizen's security and basic rights

Status quo in Korea

- It is evident that since 2023 various institutions, both public and private, at various levels and in various social structures in Korea aware that AI regulation is necessary, even without the EU AI Act and OECD recommendations.
- However, their efforts and initiatives are separate and independent.
- From government to civil activists should harmonize and collaborate to figure out the best procedure and the best output for the society with AI, as we should have to live with AI.

Status quo in Korea

- If the social consensus or agreement on an emerging technology cannot catch up with the technology itself, the integrity of the society may be challenged.
- Emerging technologies are often complex and difficult to understand, and they can be used in ways that were not anticipated.
- Nevertheless, the innovative contributions to economic growth and social progress should be promoted and proteched at the same time.

Status quo in Korea

- As a member country, Korea ought to comply with the recommendations of OECD.
- In December 2023, there was an inofficial discussion meeting between government officers and IT company developers.

Status quo in Korea (the December meeting)

- ‘‘Government’’ vs ‘‘IT industries’’
- ‘‘OECD recommendations and EU AI Act’’ vs ‘‘Legal freedom’’
- ‘‘Regulation’’ vs ‘‘Innovation’’
- ‘‘Responsibility’’ vs ‘‘Self discretion’’
- ‘‘Ethics of AI’’ vs ‘‘AI ethics’’
- ‘‘For users’’ vs ‘‘For builders’’

Status quo in Korea

- The discussions are on-going while AI advances.
- At the moment, no conclusion is a conclusion.
- Parallel conversations without touching points ...
- However, the compromise between regulation and innovation seems, and should be, not impossible.
- From time to time we have such pessimistic predictions on AI as ‘‘AI killing humans’’ which should be carefully avoided.

Thank you so much for your attention