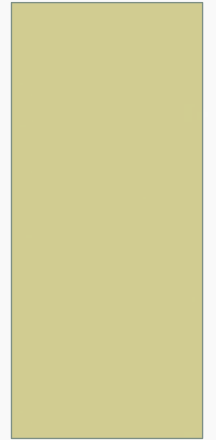


ETHICAL IMPLICATIONS OF AI: ASSESSING TRUSTWORTHY AI IN PRACTICE

2022-02 GSDS HEEJIN KIM



FUNDAMENTAL HUMAN RIGHTS SESSION #1

[1] Some Introductory Contexts – The Impact of AI-based Systems

[2] Assessing AI in terms of Ethics

- Ethics-based Guidelines – From States to Tech Companies
- Prospects and Limitations (Comparison w/ human rights framework)

[3] Human Rights as a Framework of Assessing AI

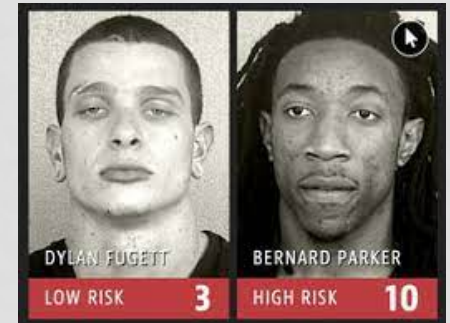
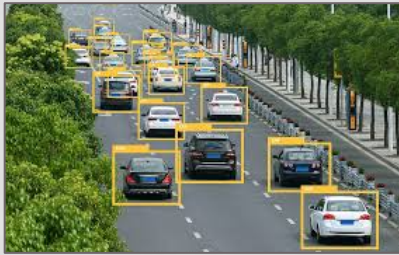
- Human Rights as Expressed in International Law (Some key instruments)
- Building Trustworthy AI: Ethical Principles, Human Rights & A Growing Trend in Law

[4] The Relevance and Importance of Human Rights Discourse for You

- Human Rights? (HR law provides rules of conduct for the actors involved)
- Why Human Rights?

THE IMPACT OF AI-BASED SYSTEMS ON OUR LIVES

AI is changing the world before our eyes



The wide-ranging use and application of AI systems



THE IMPACT OF AI-BASED SYSTEMS ON OUR LIVES

The wide-ranging use & application of AI systems

AI & medical imaging and diagnosis



Generative art & AI-generated audio

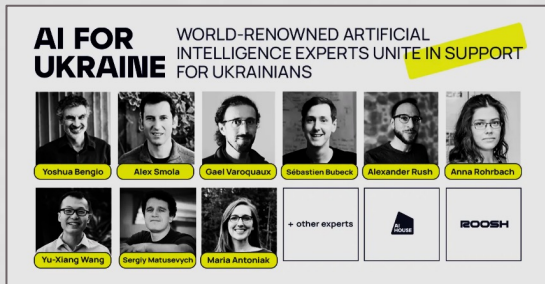


Text prompts into images @ MidJourney AI

THE IMPACT OF AI-BASED SYSTEMS ON OUR LIVES

The wide-ranging use and application of AI systems

AI FOR UKRAINE WORLD-RENOWNED ARTIFICIAL INTELLIGENCE EXPERTS UNITE IN SUPPORT FOR UKRAINIANS



The image shows a grid of 11 portraits of AI experts. The first row contains six portraits: Yoshua Bengio, Alex Smola, Gael Varoquaux, Sébastien Bubeck, Alexander Rush, and Anna Rohrbach. The second row contains five portraits: Yu-Xiang Wang, Sergiy Matusevych, Maria Antoniak, a box with '+ other experts', a box with the 'Primer' logo, and a box with the 'ROSH' logo.

Yoshua Bengio Alex Smola Gael Varoquaux Sébastien Bubeck Alexander Rush Anna Rohrbach

Yu-Xiang Wang Sergiy Matusevych Maria Antoniak + other experts Primer ROOSH



War and military intelligence



[e.g.] A joint initiative between Primer and the Ukraine government (concerning Russian military radio traffic – interception, auto-transcribing etc.)

THE IMPACT OF AI-BASED SYSTEMS ON OUR LIVES

Some doubts & dystopian possibilities? [e.g.]

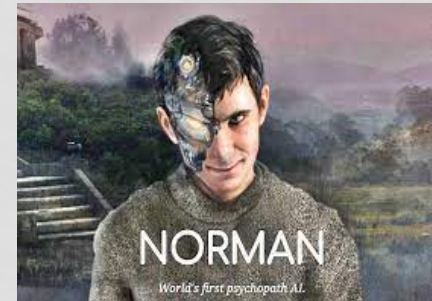


Data generation, collection, storage, analysis, and use **v.** right to privacy

Bias in the training data sets **v.** non-discrimination



“Psychopath AI”?



→ Voices calling for more fairness, accountability, transparency, trustworthiness etc. [How?]

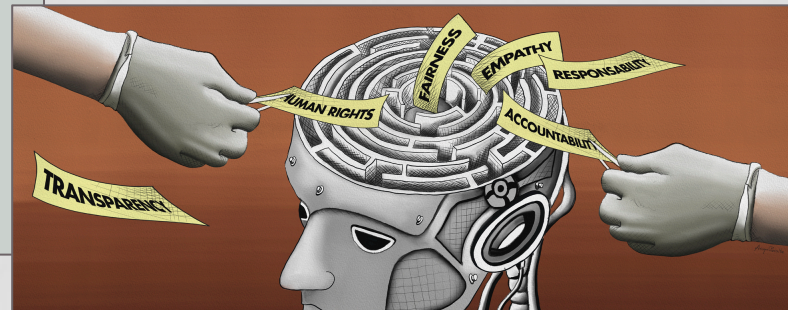
ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

What is ethics (moral philosophy)?

Ethics-based Guidelines – From States to Professional Associations & Tech Companies



- A growing trend to frame various social implications of emerging technologies including AI as *ethical* issues



ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

Ethics-based Guidelines – From States to Professional Associations & Tech Companies

[For instance]

- ❑ **IEEE's treatise, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems"** (latest ver. 2019) – concerning ethical design, development, and implementation of technologies

"The full benefit of these technologies will be attained only if they are aligned with our defined values and ethical principles."

ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

Ethics-based Guidelines – From States to Professional Associations & Tech Companies

[For instance]

EU – **Ethics Guidelines for Trustworthy AI** (Apr 2019)

Korea – **National AI Ethics Guideline** (Dec 2020, Ministry of Science and ICT)

- 3 fundamental principles (human dignity, common good for the society, purposiveness) AND 10 requirements (human rights protection, privacy, respect for diversity, data management, accountability, transparency etc.)

Australia – **AI Ethics Framework** (Nov 2019)

- 8 guiding principles (e.g. human-centered values, fairness, privacy protection and security, reliability and safety, contestability etc.)

ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

Ethics-based Guidelines – From States to Professional Associations & Tech Companies

[For instance]

Singapore – **AI Governance Framework** (Jan 2019; Jan 2020, Personal Data Protection Commission) – organized into four key areas

- Understanding how AI model reaches decision
- Safety and resilience of AI system
- Fairness and no unintended discrimination (fairness and data governance)
- Management and oversight of AI system (human accountability and control)

Japan – **Social Principles of Human-centric AI** (Mar 2019); **Governance Guidelines for Implementation of AI Principles** (Jan 2022)

US DoD [Sector-specific] – **Ethical Principles for Artificial Intelligence** (Feb 2020)

ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

Ethics-based Guidelines – From States to Professional Associations & Tech Companies

[For instance]

- Google, AI Principles; Responsible AI Practices
- Microsoft, Microsoft AI Principles; Microsoft Responsible AI Standard
- Kakao Corporation, AI Ethics Charter



Q: Limits of the Ethical Principles as a framework of assessing AI system and its impact?



ASSESSING AI IN TERMS OF ETHICAL PRINCIPLES

Ethical Principles as a framework of assessing AI system and its impact – **Prospects and Limits**

Cherry-picking

- Ethical guidelines are the value statement, which are not precisely defined (considerable room for interpretation)

Emphasis on self-regulation (operating on a voluntary basis)

- Reliance on the good-will of the relevant actors

Lack of formal enforcement and accountability mechanism

- Who bears the cost of an “unethical” use of AI? How do we monitor and enforce violations of the guidelines?

HUMAN RIGHTS AS AN ASSESSMENT FRAMEWORK

Human rights as a framework to understand, assess and evaluate AI-based systems and their impact

[Our focus: Human rights as expressed and guaranteed by international human rights law]

- **International human rights law?**
 - Rules to promote and protect human rights in international law
 - Mechanisms to monitor and redress human rights violations



HUMAN RIGHTS AS AN ASSESSMENT FRAMEWORK

Human rights as a framework to understand, assess and evaluate AI-based systems and their impact

- **UN Human Rights Treaties** – constituting the main body of human rights law
 - UN has identified 9 core human rights treaties
 - Every UN member state (193 members in total) has joined at least one out of 9; and 80% of the states have joined 4 or more
- **The Three Key Human Rights Instruments in General**
 - Universal Declaration of Human Rights (UDHR)
 - International Covenant on Civil and Political Rights (**ICCPR**) and International Covenant on Economic, Social and Cultural Rights (**ICESCR**) --- legally binding treaties (joined by more than 170 states as of today)

HUMAN RIGHTS AS AN ASSESSMENT FRAMEWORK

- **AI & Human Rights @ UN and its specialized agencies**
 - HR = basis of any effective AI governance; setting the outer boundaries of AI governance
 - HR law = one of the frameworks for the AI design, development & deployment
- **Some of the Key HR Instruments relevant to AI & Human Rights**
 - International Covenant on Civil and Political Rights (**ICCPR**) and International Covenant on Economic, Social and Cultural Rights (**ICESCR**)
 - **UN Guiding Principles on Business and Human Rights**
 - [depending on the rights concerned] **Several core human rights treaties** concerning specific groups of people or specific human rights problems
 - (Will come back on Thurs)

International human rights law
(esp Core human rights treaties identified by the UN)
sets out **rules of conduct** for the actors involved

Rights

- **Q: Whose rights?**
- **Individuals** and (according to some treaties) **certain groups of people** (e.g. indigenous community)

Obligations

- **Q: Whose obligations?**
- **Primarily state actors** and (according to some treaties) companies and other relevant private actors

Will come back on Thurs

**Mutually-reinforcing for Assessing the AI
System and Its impacts**

**Human rights as
expressed in law**



Ethical principles

Human rights – Ethical principles – trustworthiness



“In civilized life, law floats in a sea of ethics.”
(1962, Chief Justice Earl Warren, US Supreme Court)

OTHER ASSESSMENT FRAMEWORKS

Development in the current & proposed legal frameworks regulating AI

For instance

- **EU** – AI Act (currently in draft; proposed amendments submitted by the European Parliament and the Council of the European Union)
 - Regulating a range of AI applications **through a risk-based approach** (= Applications sorted into categories of unacceptable risk, high risk, limited; minimal risk)
 - Addressing concerns about AI systems that affect **human rights** as **high-risk AI systems** (e.g. private sector AI application in hiring, access to education, credit scoring; AI used by governments for law enforcement, border control, utilities, judicial decision-making etc.)

OTHER ASSESSMENT FRAMEWORKS

Development in the current & proposed legal frameworks regulating AI

- **US** – currently a fragmented approach to AI regulation (~~nation-wide agreement~~)
 - Different laws at state-level; but recently, US Congress – National AI Initiative Act (2021), Algorithmic Accountability Act (2022); White House Office of Science and Technology Policy, Blueprint for an AI Bill of Rights (2022)
- **China**
 - [e.g.] Cyberspace Administration of China, Online Service Algorithmic Recommendation Management Regulation (2022) → extensive control on the use of AI in online recommendation systems
- **International trade law**
 - AI regulation @ Digital Trade Partnership Agreement & FTAs
- **Why important?**

THE RELEVANCE & IMPORTANCE OF HUMAN RIGHTS DISCOURSE

- **The notion of human rights** (as we have today) & **its universality** does not have a long history.
- **The origin of human rights**
 - Throughout the history, different countries, different religions & cultures have defined it in their own contexts. *e.g.*, the Code of Hammurabi, the Ten Commandments, Buddhist and Confucius teachings etc
- **Human Rights as expressed in law**
 - Relatively recent development

Human rights as the rights inherent to all human beings, regardless of race, gender, nationality, ethnicity, religion or any other status?



THE RELEVANCE & IMPORTANCE OF HUMAN RIGHTS DISCOURSE

- **Q:** Have you taken any human rights/ethics-related courses from the liberal arts program and/or school of social science (law, political science, sociology, literature, philosophy, public policy etc)?
- **Q:** [Your encounter with human rights]: in the course of your undergraduate; graduate; PhD studies and works, what was your first encounter with human rights issues?
 - (e.g. incidents where you realize human rights implications of the field of your specialization; discussion with your colleagues about the human rights issues concerning the project you were working on etc.)

THE RELEVANCE & IMPORTANCE OF HUMAN RIGHTS DISCOURSE

- [e.g.] Legal research in law and policy concerning how certain technologies are created, used, and applied.
 - Electronic signature regulation in emerging e-commerce markets in Southeast Asia
 - Controlling the transfer of cyber surveillance technology
 - Facial recognition tech in the airport and rights to privacy