EBU5601
Data Design

Data Ethics

Dr Chao Shu, Dr Xiaolan Liu
School of Electronic Engineering and Computer Science
Queen Mary University of London
Sep. 2024

Learning Outcomes

- The main outcomes are:
  - **[LO3.1]** Understand the concept of data ethics
  - **[LO3.2]** Understand the importance of data ethics in technology and innovation
  - **[LO3.3]** Understand data ethics issues across data lifecycle

What is Ethics

- The word ‘Ethics’ originated from a Greek word ‘Ethos’ which describes way of life.
- Ethics are essential in every aspect of life and set standards for right and wrong.
- Sense of right and wrong beyond laws – e.g. respecting others, offering help to those in need in a variety of circumstances.
- Guide decisions and actions
- Set of standards, rules, values set by governments, organizations and individuals
- Following ethics have individual as well as societal benefits

Ethics of technology and innovation

- Technology advancement in last decade.
- Risks, benefits and impact associated.
- Responsible use of technologies
- Examples:
  - Stem cell research
  - Drone warfare
  - AI
    - ChatGPT

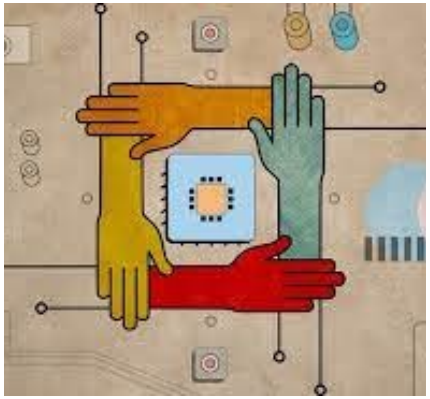


Image: Christine Daniloff, MIT

- **Data is the game changer**

Image source: http://www.it.uu.se/grad/courses/gc2021/etso20

Data drives innovation

- Oil and energy companies most valuable in the 20th century replaced by data-driven tech companies in the 21st century.
- Massive data created in the past 5 years with endless applications.
- Social media, smart devices, cloud and AI
- Economic growth, societal progress
- Great power, great responsibility

Data Ethics

Image source: https://digital-strategy.ec.europa.eu/en/policies/strategy-data

What is data ethics

- Branch of ethics that studies moral problems related to data.
- Data ethics focus on processing, sharing, use and application of data.
- Data ethics foundations
  - Privacy and data protection
  - Informed consent, individuals rights and agency
  - Transparency and trust
  - Fairness, non-discrimination
  - Accountability and oversight
  - Benefit to others and society

Image source: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/ethical-data-usage-in-an-era-of-digital-technology-and-regulation

Lessons from past

- HeLa Cells
  - Immortal human cells, crucial in advancing cancer research
  - From Henrietta Lacks, without her consent from biopsy
  - Privacy, dignity, respect – ethical issues

- Tuskegee syphilis study
  - Effects of untreated syphilis 1932-1972
  - African American participants, not informed well, no access to treatment
  - 100+ died as a result, despite treatment option

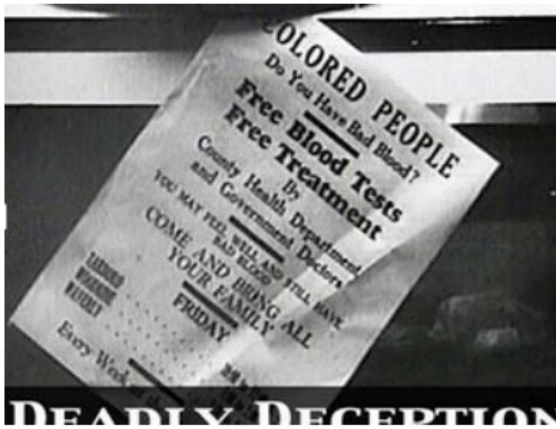


Image source: https://projecthbw.ku.edu/uncategorized/henrietta-lacks-the-immortal/

Guiding ethical principles

- **Do not harm** – minimize risks for individuals
- **Beneficence** – maximize benefits for individuals and society
- **Respect, autonomy, agency-** consent, privacy, information, control, agency
- **Justice** - Legality, fairness and non-discrimination
- IEEE, Royal Academy of Engineering, ACM address data ethics - Responsible use of data, preventing misuse

Encoded data ethics in laws

- Data protection laws like GDPR, CCPA embed several data ethics principles
- Privacy, data security, transparency, informed consent, non-discrimination, accountability.
- Not mere guidance, but enforced
  - GDPR fines
  - EUR 20 million or 4% of the worldwide turnover

Data Ethics and AI

- AI combines algorithms and robust dataset to solve problems
- AI models need massive amounts of data, most of AI efforts spent on data
- Powerful AI uses large dataset and poor data leads to poor AI outcomes.
- AI ethics wide-ranging than data ethics, important to know the similarities and differences
- Consequences if AI decision making:
  - Credit access/loans
  - Hiring
- AI discrimination
- AI/Robots taking over jobs Will robots/drones claim fewer human lives in a war?

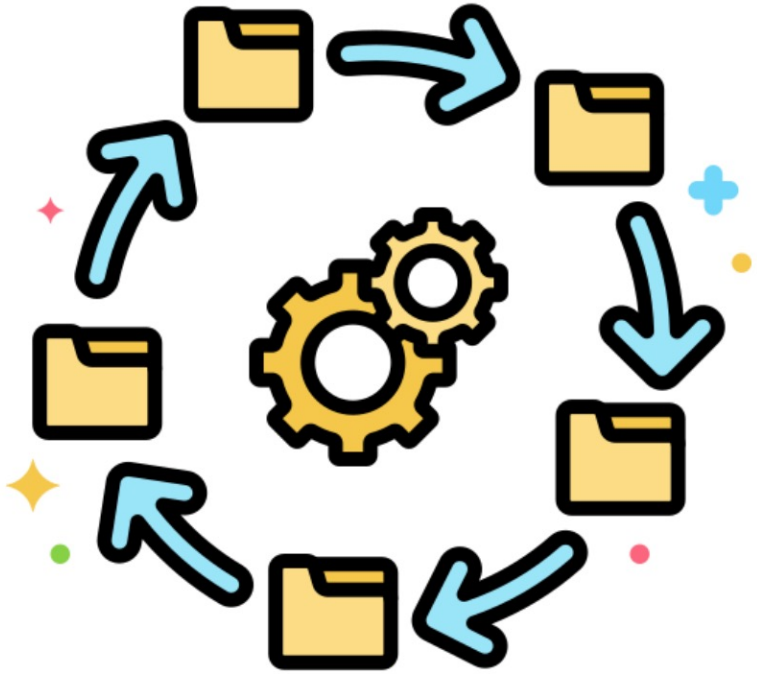
Image source: https://www.project-sherpa.eu/885-2/

AI ethics guidelines

- Many organizations and bodies working on AI ethics and guidelines
- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency Diversity, non-discrimination and fairness
- Environmental and societal well-being
- Accountability

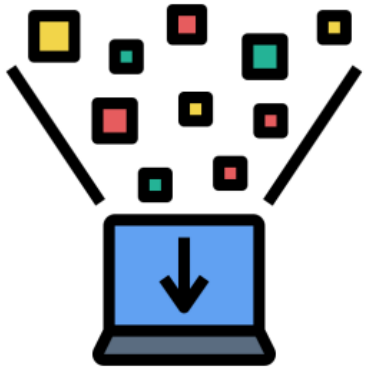
Ethics issues across data life cycle

- At every stage of data life cycle
  - Data acquisition- collection, sourcing
  - Data preparation- cleaning, labeling, quality checks
  - Data storage - infrastructure, security, integrity
  - Analysis- AI, interpretation, decision-making
  - Retention/ archival
  - Sharing



Data acquisition

- There are many ways to gather data such as surveys, mobile apps, sensors, wearables, web scraping, third parties
- You must consider
  - Are you allowed to collect the data?- privacy, copyright
  - Purposeful collection - clear about why, and how much
  - Representative data, respectful of people's time
  - Informed consent- crucial
  - Vet your data suppliers



Icon made by Parzival 1997 from www.flaticon.com

Data preparation

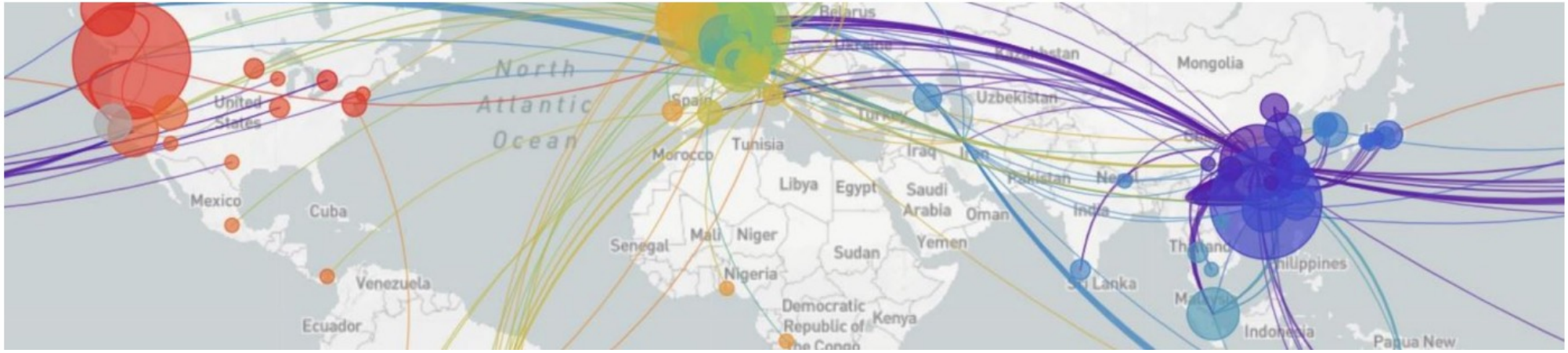
- Data preparation includes cleaning, labeling, annotation- Transcribing audio files, labeling text or images, flagging inappropriate content.
- Human annotators- inadequate training, exploitation, Kenyans workers for ChatGPT
- Data quality inconsistencies, biased labels

Data Storage

- Data that we gather is an asset and we are responsible to maintain Confidentiality and integrity of the data
- We should be careful and prevent data breaches or accidental losses
- Data security- no unauthorized access should be allowed
- Technical: Infrastructure, methods, techniques, and devices for data storage
- Organizational: Companies policies, training

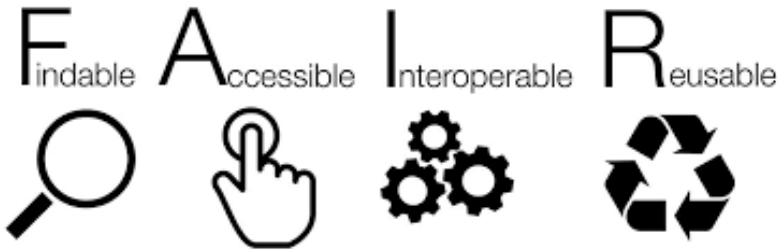
Data Sharing

- Data sharing needed for innovation and collaboration, sometimes monetization
- Positive outcome if responsible- Covid data sharing
- Privacy regulations, individual rights
- Data ownership, informed consent
- Privacy-preserving sharing

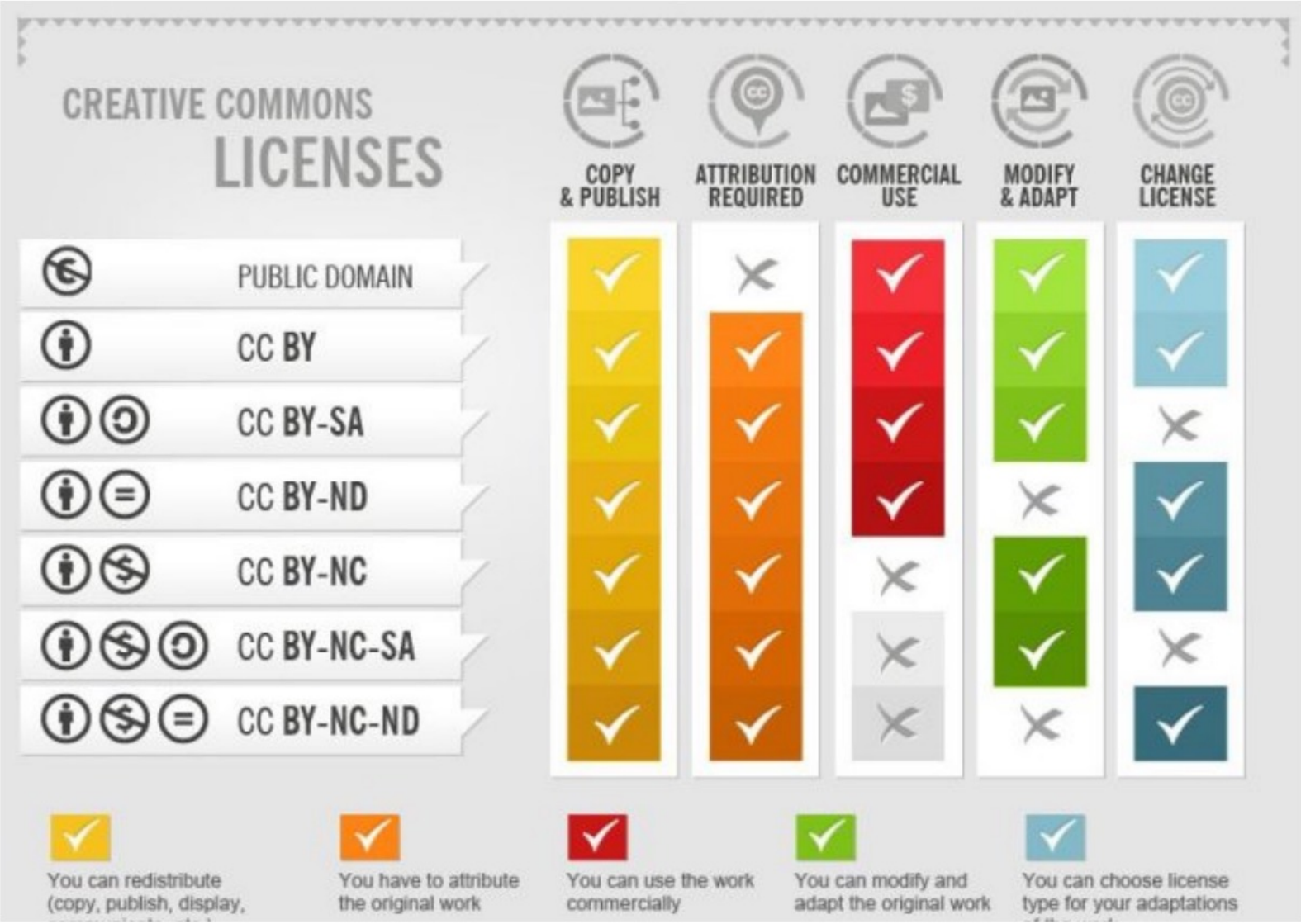


FAIR Data

- **FAIR**: Findable, Accessible, Interoperable, Usable
- It is important to have Fairness in data or unbiased data to have true representation of statistics
- But what are FAIR data principles
- Valorize data through efficient data management systems
- Reuse of data, better awareness, retrieval
- Making data ready for collaboration and sharing
- Scientific research, more data intensive industries
- FAIR data common in scientific research data management



Data licensing – Creative Commons



Open data for science and society

- Scientific advancement push for open data which can be reused in a beneficial way for the humanity.
- Open data reflects transparency and emphasize on reuse of data to enhance, business, creation and innovation
- Success Stories
  - Apply Board – Canadian Company
    - Online platform for international student recruitment
  - Global open data for Agriculture and Nutrition
    - Land use and productivity data
    - Pest and disease management

source: https://open-power-system-data.org

Questions

Use student forum on QM+

chao.shu@qmul.ac.uk
xiaolanliu@qmul.ac.uk